International Journal of Engineering Sciences & Emerging Technologies, Dec. 2013.
ISSN: 2231 – 6604 Volume 6, Issue 3, pp: 344-351 ©IJESET

VOICE RECOGNITION BASED SECURE ANDROID MODEL FOR INPUTTING SMEAR TEST RESULTS

Teenu Therese Paul[1], Shiju George[2]

[1]PG Scholar, Department of Computer Science and Engineering,
MG University, Kerala, India
teenutheresepaul@gmail.com

[2]Faculty, Department of Information Technology, MG University, Kerala, India
shijugeorge@amaljyothi.ac.in

ABSTRACT

Voice recognition technology is the process of identifying and understanding the voice signals of a user, which is converted into text or commands for a program. In this work the voice recognition technology is applied into a laboratory information system for identifying each technician's voice. i.e. By using the user's voice sample, a secure authentication system is developed where the unique features of the user's voice are extracted and stored at the time of registration. Afterwards during the login stage, unique features of the user's new voice sample are extracted. Then compare the features with all the stored features rather than the just previous one. For this, a unique username is set to all the users. The comparison operation is performed with all voice samples under that particular user name. The voice feature comparison process is done by using Fast Fourier Transform techniques. After a successful login the user can enter the results of smear test through his voice rather than typing into the system. The system mainly consists of two parts - a client system and a server system. The client system is developed using Android and the server system is implemented in Java.

KEYWORDS: *Voice Recognition, Speaker Identification, Speech Recognition, Smear Test, Fast Fourier Transform.*

I. INTRODUCTION

Voice recognition technology is the process of identifying, understanding and converting voice signals into text or commands. There are different types of authentication mechanisms available today like alphanumeric passwords, graphical passwords etc. Along with these, biometric authentication mechanisms like fingerprint recognition system, voice recognition system, iris recognition system etc. add more security for data. One of the important areas of research is Voice recognition technology. Research in voice recognition involves studies in physiology, psychology, linguistics, computer science, signal processing, and many other fields. In this research the focus is on a voice based authentication system. This voice recognition technology consists of two different technologies such as speaker recognition and speech recognition, which are both considered to be emerging areas of research.

Speaker recognition is the process of identifying the exact user who is speaking. It involves two systems – speaker's voice identification and speaker verification. Speaker identification involves identification of the unique speaker's voice from a set of other voices. This is done by inputting a user's voice into the recognition system. This recognition system stores a set of all known user's voices. From the input voice, the system needs to identify who is the speaker from the available list of voices. Thus the speaker identification system works within a closed set of data. In speaker verification system, a user's given voice sample is verified to check whether the user is valid or not. This is done by comparing the user's new voice features with the stored voice features. This is carried out for the purpose of authenticating the user. [1]

The speaker recognition system falls into two categories, namely text- dependent and text-independent. In text-dependent speaker recognition system the text given to all users at the time of enrollment and verification are same or unique. In text independent speaker recognition system, the

analysis is not based on the uniqueness of the text said, but it is judged using the unique voice features like accent, frequency etc. [4] Speech recognition is the process of translating the words spoken into corresponding text format. This is applied usually for hands-free operations. In essence speech recognition technology recognizes words. In this research both speech and speaker recognition technologies are studied and implemented in Android platform. With the help of these technologies a secure voice based authentication system for inputting the laboratory test results is developed. The details regarding this research are explained in subsequent sections.

This paper is organized as follows: the following section depicts the smear test procedure in NIRT. The third section contains the proposed system architecture model. The fourth section explains the implementation procedure of the system. Fifth section shows the results and final section gives the conclusion.

II. RELATED WORKS

The most common approaches to voice recognition can be divided into two classes: "template matching" and "feature analysis". Template matching is the simplest technique and has the highest accuracy when used properly, but it also suffers from the most limitations. As with any approach to voice recognition, the first step is for the user to speak a word or phrase into a microphone. The electrical signal from the microphone is digitized by an "analog-to-digital (A/D) converter", and is stored in memory. To determine the meaning of this voice input, the computer attempts to match the input with a digitized voice sample or template that has a known meaning. This technique is a close analogy to the traditional command inputs from a keyboard. The program contains the input template, and attempts to match this template with the actual input using a simple conditional statement. [4]

Speech recognition system is essentially a kind of pattern recognition system, including three basic units such as feature extraction, pattern matching, and reference model library. The unknown speeches is converted into electric signals through microphone, attached to the input of identification system, preprocessed first, then establish the model according to the characteristics of human speech sounds, analyze the input voice signal, and extract the desired characteristics. The speech recognition templates we need are acquired based on it. In the process of recognition, the input speech signal characteristics were compared with the voice template stored in computer according to the speech recognition model, finding out a series of optimal template matching with input phonetic by a certain search and matching strategies. Then, computer provides recognition results by checking the table according to the template definition. Obviously, the best results lie on the feature selection, the quality of speech model, and the accuracy of template. [8]

Speaker recognition systems generally consist of three major units as shown in Figure 2.1. The input to the first stage or the front end processing system is the speech signal. Here the speech is digitized and subsequently the feature extraction takes place. There are no exclusive features that convey the speaker's identity in the speech signal, however it is known from the source filter theory of speech production that the speech spectrum shape encodes in it the information about speaker's vocal tract shape via formants and glottal source via pitch harmonics. Therefore some form or the other of the spectral based features is used in most of the speaker recognition systems. The final process in the front end processing stage is some form of channel compensation. Different input devices impose different spectral characteristics on the speech signal, such as band limiting and shaping. Therefore channel compensation is done for removal of these unwanted effects. Most commonly some form of linear channel compensation, such as long and short-term cepstral mean subtraction are applied to features. The basic fundamental of spectral subtraction is that the power spectrum of speech signal corrupted by additive noise is equal to the sum of the signal power spectrum and noise power spectrum. Power spectrum of the noisy signal is computed and from this spectrum an estimate of noise power spectrum is subtracted. [7]

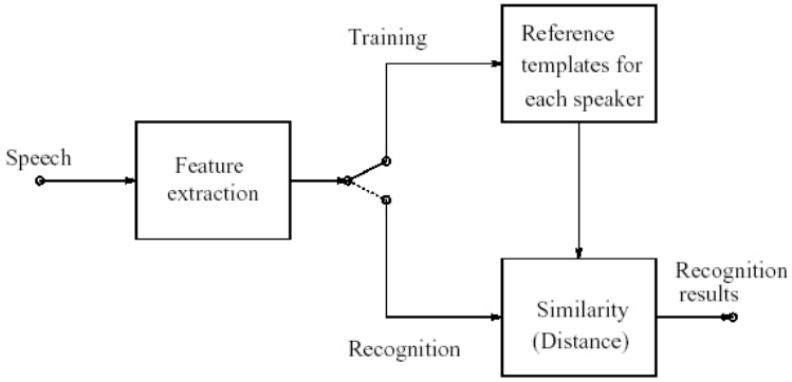

Fig.2.1 Structure of Speaker Recognition System

The applications of speaker recognition technology are quite varied and continually growing. Below is an outline of some broad areas where speaker recognition technology has been or is currently used.
Access Control: Originally for physical facilities, more recent applications are for controlling access to computer networks (add biometric factor to usual password and/or token) or websites (thwart password sharing for access to subscription sites). Also used for automated password reset services.
Transaction Authentication: For telephone banking, in addition to account access control, higher levels of verification can be used for more sensitive transactions. More recent applications are in user verification for remote electronic and mobile purchases (e- and m-commerce).
Law Enforcement: Some applications are home-parole monitoring (call parolees at random times to verify they are at home) and prison call monitoring (validate inmate prior to outbound call). There has also been discussion of using automatic systems to corroborate aural/spectral inspections of voice samples for forensic analysis. Speech Data Management: In voice mail browsing or intelligent answering machines, use speaker recognition to label incoming voice mail with speaker name for browsing and/or action (personal reply). For speech skimming or audio mining applications, annotate recorded meetings or video with speaker labels for quick indexing and filing.
Personalization: In voice-web or device customization, store and retrieve personal setting/preferences based on user verification for multi-user site or device (car climate and radio settings). There is also interest in using recognition techniques for directed advertisement or services, where, for example, repeat users could be recognized or advertisements focused based on recognition of broad speaker characteristics (e.g. gender or age). [2][3]
Set of features of speaker speech production like semantic, phonologic, phonetic and acoustic, speaker-specific information. The semantic level deals with transformation caused on the speech signal according to the communicative intent and dialog interaction of the speaker. For example, the vocabulary choice and the sentence formulation can be used to identify the socio-economic status and/or education background of the speaker. The phonological level deals with the phonetic representation such as, duration and selection of phonemes, intonation of the sentence can be used to identify the native language and regional information. It is deals with the vibration of the vocal cords and the movements of articulators (lips, jaw, tongue, and velum) of the vocal tract. For example, speaker can use a different set of articulator movements to produce the same phoneme. The acoustic level deals with the spectral properties of the speech signal. For example, the dimensions of the vocal tract, or length and mass of vocal folds will define in some sense the fundamental and resonant frequencies, respectively. The pattern matching is responsible for comparing the features to speaker models. There are various types of pattern matching methods. Some of the methods include Hidden Markov Models (HMM), Dynamic Time Warping (DTW), and Vector Quantization (VQ). In open-set applications (speaker verification and open-set speaker identification), the estimated features can also be compared to a model that represents the unknown speakers. [1]

III. SMEAR TEST PROCEDURE IN NIRT

Smear test is a laboratory test used for finding the presence of Mycobacterium tuberculosis. The technicians in the Bacteriology lab of National Institute for Research in Tuberculosis prepare the

smear for each sample. In smearing only one slide is prepared. For the smear, Auramine Phenol Staining is used. With Auramine staining, the bacilli appear as slender golden yellow fluorescent rods, standing out clearly against a dark background. A counter-stain is employed to highlight the stained organisms for easier recognition. [5]

For preparing the smear select the thick viscous portion of the sputum. With the help of a loop spread it evenly on the glass slide. Place the slides on a staining rack, with the smeared side facing up; the slides do not touch each other. Flood the slides with freshly filtered 0.3% Auramine-phenol. Let them stand for 7-10 minutes. Then wash well with running tap water, taking care to control the flow of water so as to prevent washing away of the smear. Drain the water from the slides. Then decolorize by covering completely with 1% acid-alcohol for 1-2 minutes. Wash well with running tap water. Drain the water from the slides. Counter stain with 0.1% potassium permanganate for 30 - 45 seconds. Wash well with tap water and allow the slides in slanting position to dry in hot plate maintained at 75°C - 80°C. Then bring the smear slides to the smear reading room. [10]

On receiving a batch of smeared slides the technician in the reading room starts reading it to know the presence of Mycobacterium tuberculosis. Smear reading is done by two technicians first by the reader and second by the checker. After reading each slide, they grade them as either positive or negative and mark the result in the register.

IV. PROPOSED SYSTEM ARCHITECTURE MODEL

The model is based on the basis of voice recognition. The participants of this model include the general users, a third party voice service provider and a server system. First, the server system receives the customer's voice messages and then transfers the message to the third party to deal with. The model includes voice information storage, voice recognition, and voice features update. [11]

4.1. Voice Information Storage

During the transactions, the server system first transforms the user's voice information into digital signals and stores the digital signals in specialized voice database. Then the server will send new voice to a voice recognition system which belongs to the third party voice service provider, where the voice will be denoised and the voice features will be extracted. After the third party voice service provider obtaining the features of voice successfully, the features information will be automatically sent to the server and stored in the server's voice features database for subsequent voice recognition. Both the server system and third party voice service provider are set up as separate nodes in a single PC.

4.2. Voice Recognition

Voice recognition mainly consists of four steps: receiving the user voice signal, using normalization to denoise, extracting feature and comparing the voice features. Specifically, the user's voice signal denoising and feature extraction of voice are completed by the third party voice service provider. Next, the third party voice service provider sends the new voice features to the server. The server system finds the latest voice features in its database for the user and compares them with the voice features which send from the third party service provider. If it's found that the voice features consistent with each other, it also indicates that the voice recognition is passed by the system and the user is legal.

4.3. Voice Features Update

The server system should not only accept the input voice information, but also update the voice features timely. We collect enough historical information and summarize new law, and establish a system for regularly view of the user's voice features to update the voice features timely.

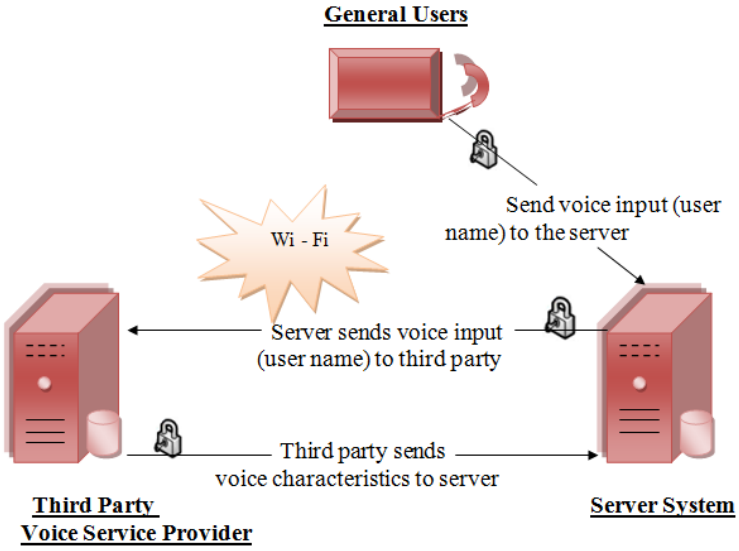

Fig.4.1 Voice Authentication Model

V. IMPLEMENTATION

Implementation is the stage of the research when the theoretical design is turned into a working system. The implementation stage is a system project in its own right. It includes careful planning, investigation of current system and its constraints on implementation, design of methods to achieve the changeover, training of the staff in change over procedure and evaluation of changeover method.

5.1. Speaker Identification

Speaker Identification is the process of identifying and authenticating who is speaking. For implementing this there are two parts needed to be set up, first a client system and second a server system. The client system is implemented in Android and the server system is implemented in java. Tablet PC or Smartphone act as client system and the Windows PC or Laptop act as server system. The user interface is set up in the client side for authenticating the user's voice samples and thereby inputting the smear test results. In the initial stage each user need to register for entering into the system. At the registration phase each user submits his personnel details such as first name, last name and designation along with a user name and his voice sample. A text displayed on the screen at the time of registration. This text is read by the user and the system captures the user's voice. There are 5 different texts stored in the system. Each time a text will be displayed in random.

The details entered into the client system are assigned into a serializable interface and then send to the server system through a Wi-Fi network and stored over there. From there the user name and corresponding voice sample is send to the third party voice service provider. The feature of this voice sample is extracted over there. For extracting the voice features it is first convert into the digital form. The main voice features extracted here are frequency, pitch contours and co-articulation. Then the extracted voice characteristics are sending to the server system and stored there.

After a successful registration the user need to login to the system. For that the user first enters his user name and then the voice sample as explained before. The server compares the features of the new voice sample with the stored set of samples under that particular user name. If it matches the user can successfully login into the system. The main advantage of this voice based authentication system is, it compares the new voice sample of a user with all his previously recorded voice samples rather than the just previous sample or the first sample.

All training and testing samples were recorded through an external sound recording program (MS Sound Recorder) using a standard microphone. Each sample was saved as a WAV file with the above properties and stored in the appropriate folders where they would be loaded from within the main application. The PCM audio format (which stands for Pulse Code Modulation) refers to the digital

encoding of the audio sample contained in the file and is the format used for WAV files. In a PCM encoding, an analog signal is represented as a sequence of amplitude values. The range of the amplitude value is given by the audio sample size which represents the number of bits that a PCM value consists of. [12]

Since not all voices will be recorded at exactly the same level, it is important to normalize the amplitude of each sample in order to ensure that features will be comparable. Since all samples are to be loaded as floating point values in the range [-1.0, 1.0], it should be ensured that every sample actually does cover this entire range. The procedure is relatively simple: find the maximum amplitude in the sample, and then scale the sample by dividing each point by this maximum. [12]

The FFT filter is used to modify the frequency domain of the input sample in order to better measure the distinct frequencies we are interested in. Two filters are useful to speech analysis: high frequency boost, and low-pass filter.

Speech tends to fall off at a rate of 6 dB per octave, and therefore the high frequencies can be boosted to introduce more precision in their analysis. Speech, after all, is still characteristic of the speaker at high frequencies, even though they have lower amplitude. Ideally this boost should be performed via Compression, which automatically boosts the quieter sounds while maintaining the amplitude of the louder sounds. However, we have simply done this using a positive value for the filter's frequency response. The low-pass filter is used as a simplified noise reducer, simply cutting off all frequencies above a certain point. The human voice does not generate sounds all the way up to 4000 Hz, which is the maximum frequency of our test samples, and therefore since this range will only be filled with noise, it may be better just to cut it out. [6][9]

Essentially the Fast Fourier Transform filter is an implementation of the Overlap-Add method of FIR filter design. The process is a simple way to perform fast convolution, by converting the input to the frequency domain, manipulating the frequencies according to the desired frequency response, and then using an Inverse- FFT to convert back to the time domain.

The code applies the square root of the hamming window to the input windows (which are overlapped by half-windows), applies the FFT, multiplies the results by the desired frequency response, applies the Inverse-FFT, and applies the square root of the hamming window again, to produce an undistorted output. Another similar filter could be used for noise reduction, subtracting the noise characteristics from the frequency response instead of multiplying, thereby remove the room noise from the input sample. [12]

The Fast Fourier Transform (FFT) algorithm is used both for feature extraction and as the basis for the filter algorithm used in preprocessing. Essentially the FFT is an optimized version of the Discrete Fourier Transform. It takes a window of size 2k and returns a complex array of coefficients for the corresponding frequency curve. For feature extraction, only the magnitudes of the complex values are used, while the FFT filter operates directly on the complex results.

The implementation involves two steps: First, shuffling the input positions by a binary reversion process, and then combining the results via a "buttery" decimation in time to produce the final frequency coefficients. The first step corresponds to breaking down the time-domain sample of size n into n frequency-domain samples of size 1. The second step re-combines the n samples of size 1 into 1 n-sized frequency-domain sample.

The frequency-domain view of a window of a time-domain sample gives us the frequency characteristics of that window. In feature identification, the frequency characteristics of a voice can be considered as a list of "features" for that voice. If we combine all windows of a vocal sample by taking the average between them, we can get the average frequency characteristics of the sample. Subsequently, if we average the frequency characteristics for samples from the same speaker, we are essentially finding the center of the cluster for the speaker's samples.

Since we are dealing with speech, greater accuracy should be attainable by comparing corresponding phonemes with each other. That is, "th" in "the" should bear greater similarity to "th" in "this" than will "the" and "this" when compared as a whole.

The only characteristic of the FFT to worry about is the window used as input. Using a normal rectangular window can result in glitches in the frequency analysis because a sudden cutoff of a high frequency may distort the results. Therefore it is necessary to apply a Hamming window to the input

sample, and to overlap the windows by half. Since the Hamming window adds up to a constant when overlapped, no distortion is introduced.
When comparing phonemes, a window size of about 2 or 3 ms is appropriate, but when comparing whole words, a window size of about 20 ms is more likely to be useful. A larger window size produces a higher resolution in the frequency analysis. [12]

5.2. Speech Recognition

Speech recognition is the process of converting voice signal into corresponding text or commands. In this research this kind of voice to text conversion is done at the time of smear test result entry. The technician speaks the result of the smear test of each patient through the microphone and the converted text is inserted in the corresponding result field and stored in the database. The speech recognition application in smart phones is incorporated to implement this speech to text conversion operation.

VI. RESULTS AND DISCUSSIONS

The research "Voice Recognition Based Secure Android Model for Inputting Smear Test Results" is successfully implemented and different kinds of test criteria are applied. In this the text to be read by the user is given to him by the system. If the text size is small i.e. the voice input is of small duration the identification procedure is little bit difficult. But in the case of large text the system is correctly recognizing the user. In this system each time different text is displayed to the user for his authentication purpose, so the accent, frequency, style etc. of each user can successfully identified. If the speaker uses the system in a noisy environment then the authentication procedure is difficult. So this system will successfully work in non-noisy environment.
In this system two stages of authentication procedure is performed. In the first case the user should give a username for his entry into the system which must be unique. In the second stage he or she should submit their voice sample to the system for successful authentication. The submitted voice stored under the particular username. These two stages created a successful authentication system based on voice.

VII. CONCLUSION AND FUTURE WORKS

Human voice is an important biometric and can be used for authentication purposes. In this research human voice is set up as an authentication key. A speaker identification based secure android model is developed in this research under the voice recognition technology. The voice recognition is done based on Fast Fourier Transform. Thus the user can successfully log into the system by authenticating his or her voice. Also the user is entering the laboratory smear test results through his or her voice. So the technician's voice can be stored for later error correction by online analysis/responsibility fixing. Thus the system can act as an efficient authentication mechanism for laboratory purposes. This application can be used in Android equipment.
The enhanced version of this system can be successfully used in the authentication phase of E-commerce transactions. This system can be used to identify each person's voice samples stored in Adhar database. Also the enhanced version of this system can be used in different kinds of games and also in various authorized speaker controlled systems.

REFERENCES

[1] Bansod N.S., Seema Kawathekar and Dabhade S.B., "Review of Different Techniques for Speaker Recognition System" Advances in Computational Research 2012
[2] Douglas A. Reynolds, "An Overview of Automatic Speaker Recognition Technology" IEEE Conference 2002
[3] Hamdy K. Elminir, Mohamed Abu ElSoud, L. M. Abou El-Maged, "Evaluation of Different Feature Extraction Techniques for Continuous Speech Recognition" 2012 International Journal of Science and Technology
[4] Jim Baumann, "Voice Recognition" Human Interface Technology Laboratory, Washington

[5] Jisha Babu, Neema Babu, Teenu Therese Paul, Shiju George, Jayakrishna V., Dr. Gomathi Sekar, "Business Process Reengineering of Bacteriology Laboratory Using Tablet PC"
[6] Marco Grimaldi and Fred Cummins, "Speaker Identification Using Instantaneous Frequencies" IEEE Transactions on audio, speech and language processing, vol. 16, no. 6, August 2008
[7] S. K. Singh, "Features and Techniques for Speaker Recognition" M.Tech. Credit Seminar Report, Electronic Systems Group, EE Dept, IIT Bombay submitted Nov 03
[8] Youhao Yu, "Research on Speech Recognition Technology and Its Application" 2012 International Conference on Computer Science and Electronics Engineering
[9] Alan V. Oppenheim, Ronald W. Schafer, John R. Buck, "Discrete Time Signal Processing"
[10] Standared Operating Protocol for Mycobacteriology Laboratory, NIRT, Chennai
[11] Yang Wujian1, Wu Yangkai 2 and Chen Guanlin, "Application of Voice Recognition for Mobile E-commerce Security" IEEE Conference 2011
[12] Serguei A. Mokhov, "Experimental Results and Statistics in the Implementation of the Modular Audio Recognition Frameworks API for Text-Independent Speaker Identification"

AUTHORS

Teenu Therese Paul completed her engineering graduation in Computer Science from Mahathma Gandhi University in 2011. Currently she is pursuing her post-graduation in Computer Science from Mahatma Gandhi University, Kottayam. Active areas of interest includes wireless sensor networks and network security.

Shiju George is currently working in Amal Jyothi College of Engineering, Kanjirapally, Kottayam, as Asst. Prof. in the department of IT. He had obtained his Bachelor degree in Computer Science Engineering from Institute of Technology, GGDU, Bilaspur Chhattisgarh in 2001, there after MS in Software Systems from BITS, Pilani, Rajasthan. Area of interest includes Intelligent Systems, Database, and Networking