Voice Recognition for User Authentication at Online Examinations

A Dissertation Submitted for the Degree of Master of Computer Science

W.D.C.P.K.Wijithapala

University of Colombo School of Computing

2021

UCSC

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis. This thesis has also not been submitted for any degree in any university previously.

Student Name: W.D.C.P.K. Wijithapala

Registration Number: MCS/2018/102

Index Number:18441024

30.11.2021

Signature of the Student & Date

This is to certify that this thesis is based on the work of Mr. /Ms. W.D.C.P.K. Wijithapala under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by,

Supervisor Name: Dr. (Mrs) H.N.D Thilini

30/11/2021

Signature of the Supervisor & Date

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisor Dr. (Mrs.) H.N.D. Thilini, Senior Lecturer of University of Colombo School of Computing for the continuous support for my study, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis.

Besides my supervisor, I would like to thank the rest of lectures and staff members of University of Colombo School of Computing for the knowledge they shared throughout the academic period.

Specially I would like to thank Language Technology Research Laboratory of University of Colombo School of Computing for providing the required datasets.

Last but not the least, I would like to thank university colleagues and my family for the continuous encouragement and support they gave me throughout the project. Finally, I would like to thank all who gave support in every way to finish my research successfully

ABSTRACT

Digitization has made a huge impact on education sector and many institutes all over the world had already transitioned to online from traditional face-to-face lectures. While learning is being practiced online, most of the examinations were conducted in general approaches within the institute's premises. But the sudden outbreak of COVID-19 pandemic forced all the educators to adapt online platforms within a short period, who were being reluctant to shift earlier. Therefore, the necessity of digital assessment platforms with secure testing environments have arisen not only for educational sector but also in recruitment procedure of employees.

In implementing an online examination system, the prime challenge is to maintain the credibility and the transparency with participants authentication. Therefore, introducing a better approach of user authentication is crucial in every examination platform. Considering the economical and the educational background in Sri Lanka, we are unable to expect the students to have accessibility of specific instruments and high-speed internet. Therefore, this study introduces a voice-based user authentication approach for online examinations which can be acquired with limited facilities.

Voice based user authentication is to identify a user by analyzing the unique features of his /her voice sample. Recent studies show that user authentication with GMM (Gaussian Mixture Models) have efficiently used in speaker recognition. The key features of text independent voice signals are obtained using MFCC (Mel Frequency Cepstral Coefficients) and a unique model for the all the speakers who enrolled to the system is generated using Gaussian Mixture Models. The maximum likelihood algorithms are used to match the users voice samples against the speakers who are already enrolled.

The dataset for the study is obtained from the UCSC/LTRL Speech corpus which contains 60 users with speech utterances of Sinhala language. A Web based application has been developed to implement the user authentication approach using python http (Hypertext transfer protocol) service and PHP. The accuracy of over 90% on correct identifications is obtained by models with voice samples relatively higher duration with GMM and MFCC where 20 trials are tested against the whole set of 60 trained models.

Key Words: Speaker Recognition, GMM, MFCC, User Authentication

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Overview

The evolution of the Internet and the World Wide Web have influenced our lives in many ways including economy, communication, professional networks, news, information, knowledge sharing and learning. The expansion of the facilities and the accessibility of technology has led online learning or e-learning to grow in significance as an educational tool. With addition to that, the growth of emerging platforms and software like wikis, blogs, social networks, and even some gaming technologies are being incorporated with Online learning. These online learning environments challenge notions of teaching and learning by avoiding the place where you are located, barriers of time by delivering equal access to education for the students all over the globe.

Though online learning was practiced only in some institutions over the past few decades, the recent outbreak of the Coronavirus pandemic has made the people to rethink on adapting to distant learning or e-learning. It has considerably changed the lifestyle of humans being advised to stay at home in isolation and work and learn from home. Not only education, but it has also affected all sort of human activities. Among those, the education sector remains one of the worst-hit by Coronavirus outbreak.

Exams and assessments play a major role in every education system in measuring the knowledge and abilities gained through what the students have learned and what they have been taught. Not only the education institutions, but some of the companies also conduct exams in recruiting their employees.

Along with this situation, online education systems are widely being adopted by many institutions these days to continue with their academic work and with those online examinations also have come to forth. The main advantage of adapting to online examinations is that they can be fully automated with the use of technology. One major challenge in the online examinations is the difficulty of providing true user authentication. Since online examinations are done without face-to-face interactions, not supervised, or invigilated and happens in uncontrollable remote locations it is very difficult to authenticate the person who is attempting the examination.

Most common authentication method that has been used in an online examination is user-password authentication. Other than that, there are many biometric authentication methods that have been used for online examination like finger, face recognition and speech/voice recognition that varies according to the environment and the facilities available.

In Sri Lankan education system, where online platforms are frequently being adapted in educational institutes, there is a need for an online examination system with a suitable user authentication methodology that can be used in the prevailing pandemic situation. To overcome the issues in user authentication, voice-based user authentication approach is introduced to the online examination platform. With that we can ensure the genuine interaction of the student with the system by increasing the security.

1.2 Motivation

To obtain the scheduled learning outcomes and maintain the same curriculum, the education sector had to quickly shift to online education with the available resources at the moment where the education institutions were closed to slow the spread of COVID-19. From nursery to PhD students around the world, are experiencing these altering effects of this transition as classrooms and exams are transformed to online education.

Being a developing country, even Sri Lanka has been able to implement and provide appreciable online education while higher education institutions being closed. According to Hayashi and others, a survey done on university students reveals that around 90% of students who responded the survey were able to engage in online learning accordingly with in the pandemic period. (Hayashi, et al., 2020), For the continuous teaching-learning progress , the online education systems are needed to be provided with online examination approach.

Due to the anonymity that exists within the examination, the users try to increase their marks by having another person doing the online examination for them. To avoid that user authentication approaches are introduced. Most of the authentication approaches that are currently available, needs high quality internet connectivity and specific instruments which are not generally available with every person. At some scenarios, the flow of education process is collapsed due to unavailability of secure online examination systems.

Therefore, the Sri Lankan education system is currently in a point where it should be provided secure and reliable online examination system with proper user authentication.

1.3 Objective

When compared to taking an exam in a classroom environment, online examination is a totally different situation where students' knowledge is evaluated using modern computer technology. Students should understand the guidelines and parameters of the test taking procedures and should have the relevant facilities to involve in the exams. When the exam system is fully automated, there are many benefits gathered from having the exam online like reducing the cost, time and effort required for the whole process. It assists the invigilators with reducing work, checking answer sheets, and producing results too.

One of the most important phases in online examination is user authentication. The prerequisites of existing user authentication methodologies do not support the social and economic background in Sri Lankan society. Therefore, this project aims to introduce a more reliable, easily accessible user authentication approach to online examinations conducted in general educational environment of Sri Lanka.

According to Sharma & Bansal's review of speaker recognition approaches, voice is one of the best biometric features which easily captured through a telephone or the internet and verified from remote users (Sharma & Bansal, 2018). During the prevailing situation and considering the facilities that can be found at household in a low-income country like Sri Lanka, voice authentication will be more advanced than traditional methods and will be easy to practice.

When considering the implementation, the voice authentication approach will support Sinhala language usage, the native language of Sri Lanka and will be presented as an online approach where it can be accessed remotely.

1.4 Scope

This research mainly focusses on one of the major obstacles in online examinations on users end: user authentication. Though there are many methods of authentication that are being practiced in technical environments and online platforms, most of them are not applicable for a student who is home bound and engaging distant learning.

Considering the minimal resources that are accessible to a person who is already engaged in online education, this research proposed a voice-based authentication method that is a step ahead of the long-established user-password authentication. Using a voice/speech wave file obtained from a person, voice will be recognized and analyzed to extract the relevant features. Authentication is carried out by comparing voice print of the individual that was previously stored with the current voice print. The user must register the system using the voice as an identification method other than the traditional username/password method. Then at the beginning and throughout the exam, the user's voice will be verified with respect to the previously stored voice data.

This research will lead to a new dimension of user authentication within an online examination held for participants engaged in distant learning and as well for any type of online examination held where the participant identification is a crucial part.

1.5 Thesis Outline

This thesis opens with the introduction of online examinations and user authentication about them, by describing its background and recently emerged increased usage of online education and examinations due to pandemic situation. In addition, it brings out how the online education came into practice, the technologies that are related with them and the user authentication methods that have been practiced for a long time.

Also, it describes the how voice is used as an authentication biometric in existing online examination platforms and how the voice analyzing is done. In Chapter 3 and Chapter 4, it is continuing the discussion on experimental set up, the theories that they are based on and how they are applied to the online examination platforms. Last section concludes the thesis with the discussion, conclusion, and the possible future work.

CHAPTER 2
LITERATURE REVIEW

As previously discussed, the study is associated with the existing solutions of online examinations and user authentication that have been practiced in online education systems. To obtain a fundamental understanding of the background and the technical solutions, the recent research that have been conducted in the same fields of study should be studied. This chapter includes the summary of related recent research that have been studied, which would be significant in defining the implementation of the system and analyzing its applicability to the environment it is deployed.

2.1 Online Education

Today's version of distance education is online education, which uses computers and the Internet as the main delivery mechanism with at least 80% of the course content delivered online (Allen & Seaman, 2008). It utilizes a blend of numerous technologies and resources like Internet, World Wide Web, Audio Visual Aids and Specially designed computer programs. Most of all the educational institutions are recommending Web-based education and are rapidly moving into online classes to satisfy student needs worldwide efficiently.

As per the records, Distance education has begun in the United States in the 1800's where some teachers and students tried to communicate through correspondence programs who located in different areas(Gunawardena & McIsaac, 1996). With the emerging computer related technologies, lately the online education stared to expand dramatically. A remarkable milestone of online education was the origin of World Wide Web in 1991 and was the factor that made the online education become what it is today. Since then, online education has grown synchronously with technological enhancement and is being widely practiced in education institutes all over the world.

Online education is vastly appreciated by both students and teachers due to many reasons. Being able to access it from anywhere at any time, having many interesting ways to deliver the lecturers, being more affordable and less costly and the availability of learning materials are some of them.

While there are so many beneficial aspects common in online learning, there are some drawbacks that have being identified like availability of technology or internet facilities, unlimited screen time and one of the key challenges is academic dishonesty.

When it comes to education, exams play a title role in it where we can assess what the students have learned with regards to subjects, and it is very same for online education. Since the online examinations are held in a different environment than the face-to-face classroom examinations, the procedure, assumptions, and techniques should be changed to maintain the expected quality. However, online education is frequently being used all over the world in despite of the challenges that are faced in implementing online education systems. They have come handy in various situations where traditional education cannot be practiced due to external causes.

More than 1.2 billion children in 186 countries have been affected by school closures because of the unforeseen spread of Covid-19 (Hayashi, et al., 2020). With the unexpected movement away from courses in most parts of the world, online education has emerged as one of the most viable options for continuing education. Accepting the modern innovations that supports online education/ examinations can only help the educators to survive with the pandemic and the concern is about how the institutes can adopt online learning in such a massive manner (Carey, 2020) .

Some of the challenges that have been aroused when adapting to online education are providing hardware and internet facilities for online examination and ensuring equal access. But with the time the governments and the educational institutes have been able to introduce many facilities with low and affordable costs for students. But introduction of a credible online examination system is still not completely successful. One of the main reasons for this is the problems in user authentication in remote examinations. Understanding and addressing new concerns and obstacles in online education is critical. Because online education will continue to be used in the future, it is essential to design a well-defined online education system.

2.2 User Authentication in Online examinations

With the transition from in-class education to online or distant education, there is a requirement of upgrading all the teaching-learning activities, so that they can be accessed by the students and the teachers effectively and efficiently. Examination being a key activity in student learning is also implemented within online education systems with parallel to them. Learning strengthened by online technologies follows the basic principle of "anywhere" and "anytime". Accordingly, online examinations are held in virtual environment for the students in remote locations with online / web-based platforms, user authentication also changed to a whole new dimension.

In traditional examinations where assessments take place within the institute, always they have identified the students by checking the student card, driving license, NIC or a valid person identification issued within the country. When the paper and pen based of examinations had to migrate to online / web-based platforms, user authentication also changed to a whole new dimension.

Most of the time, assessments in online education are submitted remotely without any face-to-face interactions and user authentication has become a continuing challenge there. There is a risk of malpractices occurring due to a lack of human invigilation in order to artificially improve their grades in online examinations. A study done shows that 73.6% of students have thought that it is easier to cheat on online examinations other than traditional examination environment (King & Guyette, 2009). Therefore, online education should be powered with effective and reliable security mechanisms to provide user authentication.

User authentication plays a key role in an online examination scenario, where it attempts to identify the user who they claim to be. Reliable authentication guarantees the legitimate interaction between the user and the examination and lead to maintain the expected quality of education gained through it. Basically, there are three main user authentication techniques as discussed by previous researchers (Ramu & Arivoli, 2013) and (Salameh & Shukur, 2015). They are discussed below.

1) *Knowledge Based Authentication*

In knowledge-based authentication, user identification is based on some facts that user already knows or was pre-shared with the user. The most used User-Password authentication method, challenge questions, security questions are some examples for this. This cannot be considered as a reliable authentication method when it comes to online examinations as it can be shared with anyone.

2) *Object Based Authentication*

Physical Objects like electronic cards with chips, digital keys and magnetic cards are used in object-based authentication and user who possess them are identified by reading them using various types of devices. They are not much applicable for an online examination where students are participating from remote locations.

3) *Biometrics Based Authentication*

Identifying a user using physiological and behavioral characteristics is biometric based Authentication. Face, Hand (Palm Print/ Fingerprint), Eye and DNA are some frequently used physiological biometrics and voice, key stroke, mouse movement are some behavioral characteristics. A person should register the system first providing the required biometrics and where they are saved and matched with the biometric provide at the point of verification.

With the evolving technologies, various user authentication methods have been introduced for various platforms over the past years. Below given in Table 1, is a summary of some of the studies that have done on user authentication at online examinations

Table 2.1 Summary of Authentication Methods

Research	Authentication Method	Description
(Ullah, et al., 2012)	**Challenge Questions**	A profile-based authentication framework (PBAF) together with a user-id and password and challenge questions
(Ramu & Arivoli, 2013)	**Keystroke and Knowledge based authentication using user -password technique**	Proposed a two-layer authentication with keystroke and user password techniques
(Awojide, et al., 2018)	**Biometric Fingerprint**	Developed a secure fingerprint Biometrics system for authentication using pattern recognition
(Fayyoumi & Zarrad, 2014)	**Face Recognition**	User authentication using face recognition
(Shdaifat, et al., 2020)	**Iris Recognition and User-Password**	Model using both username password method and iris recognition system
(Purnama, et al., 2020)	**Mouse tracking**	Proposed an affordable authentication facility with using mouse tracking
(Kydyrbekova, 2020)	**Voice Recognition**	Proposed an authentication approach using DNN features and i-vector for voice recognition
(Azeta, et al., 2018)	**Voice Recognition**	A voice based online examination framework for visually impaired students using VoiceXML

2.3 Voice Recognition as a Technique of User Authentication

In daily life of humans, voice/speech is the most used form of communication in interacting with each other. A voice biometric or "voice print," is a numerical model of the sound, pattern, and rhythm of an individual's voice (Kim & Stern, 2010) and unique to an individual as a finger or palm print (Rudrapal, et al., 2012). There are both mathematical and computational approaches to analyze voice of an individual for verification. Basically, voice recognitions mean speaking to your computer and the capability of the computer to correctly identify you using different characteristics.

Unlike password or token-based authentication, biometrics use unique biological characteristics in user authentication. Among that voice recognition comes handy in most of the situations other than biometrics like face recognition, fingerprint / palm print recognition, retina, iris recognition due to many reasons. Because it does not require specialized hardware integration from the user's end, voice authentication is a versatile and cost-effective method of authentication in various platforms. When we consider the accessibility of the internet, online resources or the World Wide Web, voice is one of the best ways of communication for the people with minimal hardware resources. There are many approaches used in previous research to identify and authenticate human voice.

Most of the currently existing online examinations are using user-password authentication. Academic dishonesty has often been a big concern, despite the supposed benefits of online learning. Though there are few systems previously proposed in the educational industry they lack strong authentication methods. An Online examination system should have reliable identity management facilities with an accurate authentication method.

Not only in online examinations but in all the systems that allow remote users to access the system, the authentication must carefully be designed. Below Table 2 discusses few approaches used in various voice authentication platforms

Table 2.2 Summary of Voice Recognition Methods

Research	Used Approach & Dataset	Designed Platform/ System	Final Evaluation	Description
(Abdul-Hassan & Hadi, 2019)	MFCC & Fuzzy Classifier / English language speech database for speaker recognition (ELSDSR) dataset	Central intelligent biometric authentication	95.45% accuracy in offline user authentication	The fuzzy classifier is developed as an inelegant verification method that depends on a predefined threshold for accepting the voice sample.
(Zhang, et al., 2018)	MFCC & GMM	Android smart phone platform	Success rate of authentication is 89% -96%	A random shuffling algorithm to create a voice cipher, MFCC is used to extract features and speaker models are modeled using GMM
(Herrera-Camacho, et al., 2019)	MFCC, GMM & Maximum Likelihood Estimation (MLE) Recordings from speakers of the Spanish language dialect used in Central Mexico	Forensic Speaker Recognition	Accuracy of 93% in user identification in any condition	MFCC and GMM used for parametrization and MLE for classification

Table 2 Continued.

Research	**Used Approach Dataset**	**Designed Platform/ System**	**Final Evaluation**	**Description**
(Mauryaa, et al., 2018)	Vector Quantizati- on & MFCC with GMM 17 Speech utterances from each 10 male and 5 female	Implement speaker recognition for Hindi		Both text independent and dependent approaches using Hindi speech samples is considered and the accuracy is high in text dependent approach
(Kumari & Jayanna, 2018	i-vector / NIST 2003 data set			The verification system uses MFCC and LPCC in feature extraction at the foundation level and i-vector classification with limited data using fusion techniques.
(Godfrey & Nichie, 2013)	Combined framework of ANN & GMM 30 different speech utterances from 20 males and 10 females	Voice recognition system	Average recognition rate of 77% for 5-word utterances	The speaker's voice is identified by continues wave form distribution using combined approach of ANN and GMM that is used for classification and the result is obtained to match features.

Table 2.2 Continued.

Research	Used Approach Dataset	Designed Platform/ System	Final Evaluation	Description
(Mahola, et al., n.d.)	HMM (Hidden Markov Models) with merging techniques	Speaker Identification System	9.8 % testing improvements was obtained with linear merging techniques when compared to nonlinear merging GMM and SVM techniques	A sub-band-based speaker identification system is presented to enhance the live performance while comparing linear and non-linear merging techniques like GMM and SVM. The results show that more accurate performance is observed with linear merging techniques.
(Kumar, 2016).	Decimated wavelet (DW) and Relative Spectra Algorithm / 50 voice signals from 10 individuals	Human voice recognition system	Accuracy rate of nearly 90% is observed	The combination of DW and Relative Spectra Algorithm is used for feature extraction. For identifying the training and testing feature vectors Euclidian distance is used, If the calculated distance between them is 0, they are matched.

Table 2.2 Concluded.

2.4 Framework for Voice Recognition System

In biometric recognition systems fingerprint, face, voice and body measurements are commonly used features for authentication. There are two key reasons that have made voice a convincing biometric in terms of adoption and deployment. First one is that speech is a natural signal produced and no extra effort should be in providing a speech sample for authentication and the second one is it can be obtained from communication lines such as telephone networks. When the speech utterance is unable to be obtained from a telephone conversation, the devices microphones, sound cards are easily accessible and can be acquired with a low cost.

Voice recognition systems can be mainly categorized in to two categories:

- Speaker Verification (determining whether an unknown voice is from a particular enrolled speaker)
- Speaker Identification (associating an unknown voice with one from a set of enrolled speakers).

When it comes to user authentication or access control most of the developed systems are based on the scenarios where speaker verification is implemented as addressed in the previous researches (Azeta, et al., 2018) , (Reedy, et al., 2021) (E.Chandra, et al., 2014) (Zhang, et al., 2018). Basically, voice authentication or speaker verification systems have the same fundamental architecture as shown in Figure 1.1 .

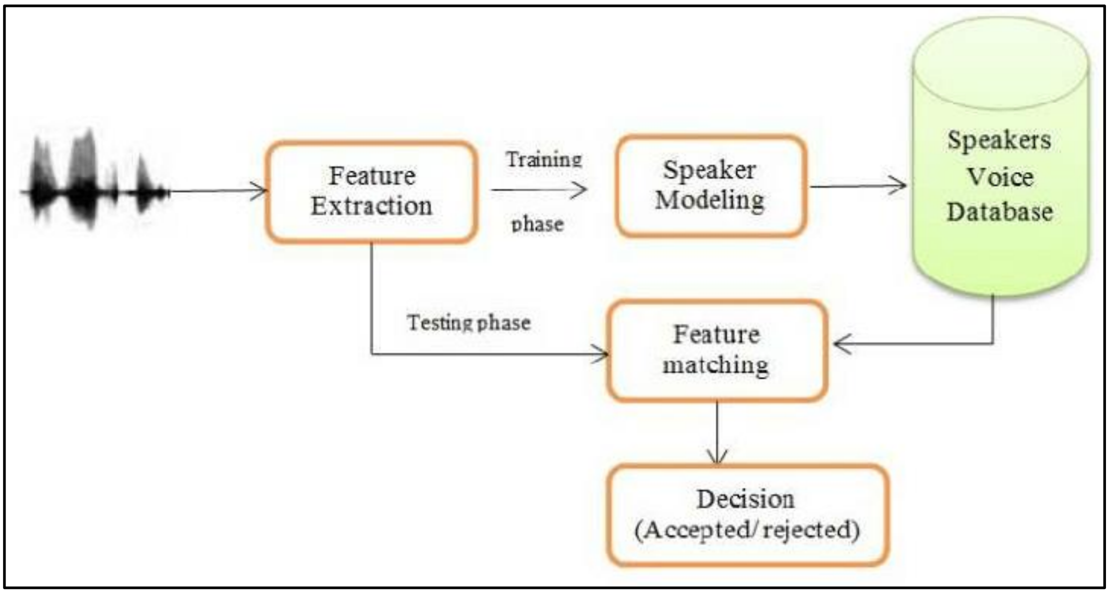

Figure 2.1. Fundamental Architecture of Speaker Verification System

The steps presented in Figure 2.1 can be described as below:

1) ***Speech Acquisition:*** Gathering the audio data/ speech signal for processing
2) ***Pre-Processing:*** Converting a high-quality speech signal to low-quality speech signal by keeping the voice characteristics to itself.
3) ***Feature Extraction:*** extracting feature vectors from the voice signal using different techniques
4) ***Speaker Modeling:*** Building speaker model by obtained feature vectors
5) ***Feature Matching / Decision:*** Involves in computing a match score in speaker recognition system.

2.4.1 Feature Extraction

It's a method for determining a speaker's personal characteristics unique to voice signal. These feature vectors carry characteristic information about the signal, which can identify the speaker (Chen, 2009). In the literature, there are several common feature extractors such as Linear Predictive Coefficients (LPC), Linear Predictive Cepstral Coefficients (LPCC), Mel Cepstral Coefficients (MFCC) and Perceptual Linear Predictive Cestrum Coefficients (PLPCC).

A research done using Hindi speech signa using MFCC as a feature extractor and Gaussian Mixture Models (GMM) have shown an accuracy of 85.49% speaker identification by using text-dependent approach consisting 10 male and 5 female speakers (Mauryaa, et al., 2018). Abel Herrera-Camacho and others have designed and tested Corpus for Forensic Speaker Recognition using the same approaches MFCC, GMM and Maximum Likelihood Estimation approach for classification showing accuracy of more than 93% identification of the speaker at any condition, proving a good recognition model (Herrera-Camacho, et al., 2019).

C. Sunitha & E. Chandra have accomplished speaker recognition using Mel Frequency Cepstral Coefficient (MFCC) and Weighted Vector Quantization (VQ) algorithm with a higher accuracy even when the speech sample produced by the speaker is short (Sunitha & Chandra, 2015). Also using the same techniques Nimesh V Bhimani have proposed a method of speaker recognition system that can recognize the speakers using MFCC and VQ algorithm (Bhimani, 2014)

Sithara and others have proposed a speaker biometric application comparing two feature extraction methods MFCC and Inner Hair Cell Coefficient (IHC). They have used these methods with Gaussian Mixture Model - Universal background model (GMM - UBM) and i-vector speaker modeling approaches. The findings of the study mentions that MFCC supas the IHC feature extraction in the two modeling techniques used, GMM and i-vector (Sithara, et al., 2018).

Chelali & Djeradi has presented a speaker recognition application which is based on Algerian Berber language and there they have used Linear Predictive Coefficients (LPC) And MFCC feature for feature extraction. Both dependent and independent approached have been analyzed in the study using one dataset with 8 isolated words and another one with continuous speech. They have used Discrete Wavelet Transformation, MFCC and LPC together and the usage of LPC have improved the efficiency of the application (Chelali & Djeradi, 2017).

When referring the recent research that have been done on the field of voice verification MFCC and LPC are the mainly used approaches in feature extraction and out of these two it can be observed that MFCC have been frequently used. Reynolds have done an experimental study on the feature extraction methods used in speaker recognition and his study show that MFCC has been commonly used in many scenarios due to its ability of elimination noise and spectral elimination in diverse conditions in voice recording (Renolds, 1994).

2.4.2 Speaker Modeling

A speaker model is created using the features that is extracted through the feature extraction process. According to Dharma & Bansal, A speaker model's characteristics are (Sharma & Bansal, 2018) :

- a theoretical underpinning to understand model behavior and mathematically approach extensions and improvements
- generalizable to new data
- parsimonious representation in both size and computation

The characteristics of the voice sample that is been used, prediction efficiency, memory usage, the effectiveness of the training and testing effects on modeling the features extracted by speaker signal. The following discussed is a summary of popular modeling techniques that have been used in recent studies.

In a speaker recognition modeling phase, the Gaussian Mixture model is used frequently. It's a type of density model made up of several component functions. Using MFCC as the feature extraction technique the studies done by Mauryaa & others and Herrera & others (Mauryaa, et al., 2018) and (Herrera-Camacho, et al., 2019) have shown a remarkable level of accuracy of speaker recognition systems that they have implemented. Aaron Niche and Godfrey A. Mills (Godfrey & Nichie, 2013) used a combination of artificial neural networks (ANN) and statistical Gaussian mixture (GMM) to implement voice recognition system and used 30 different utterances from 20 male ad 10 female and have obtained higher rates of efficiency in performance.

Another technique “Hidden Markov Models (HMM)” are used to present statistical variation of the features and how the sounds are generated statistically. In this HMMs are used to model the by encoding the evolution of characteristics. This technique had been mostly used with text-dependent speaker recognition systems where the speech utterance used is a finite text for all the speaker who are enrolling in the system. Mahola and others have presented a model for speaker identification using HMMs where they were able to improve the efficiency of the speaker identification over the GMM and Support Vector Machines with improvement of 9.78% (Mahola, et al., n.d.).

I-vector subspace modeling is a novel technology that has recently become popular in this field. It is a more basic concept-based model introduced in (Dehak, et al., 2011) which eliminates the distinction between speaker and channel variability subspaces and models a single constrained low-dimensional space dubbed referred to as the “total variability space”. A novel application of a combination of i-vector and GMM is introduced by Ibrahima and Ramli that were able to surpass the performance by 9.11% than other techniques used in speaker identification (Ibrahima & Ramli, 2018). De Gruyter has been able to overcome a very common issue in biometric applications: limited data, using an i-vector approach on a limited voice data set and have been able to achieve a significant reduction in error rates using fusion techniques along with i-vectors.

2.5 Speaker Recognition

The human ear is capable of many functions like identification of people by their voices, identifying positions of objects that produce sound objects other than functioning for communication by receiving and decoding spoken languages. In many areas like Access control, Forensics, Transaction Authentication there are many applications that are being used to identify people from their voices. Voice recognition can be as addressed as speaker recognition in implementing such systems because the voice biometric is considered in identifying a speaker based on the characteristics of their voice.

Basically, determining the person who is speaking is considered to be speaker recognition which relies on many modern computer science fields. There are many classifications in these techniques according to the data and the approaches that are used. Below described are few of the approaches are shown in figure 02. According to the identified requirements in problem analysis and the outcome expected this research would be a speaker verification, text dependent and based on a closed set.

.

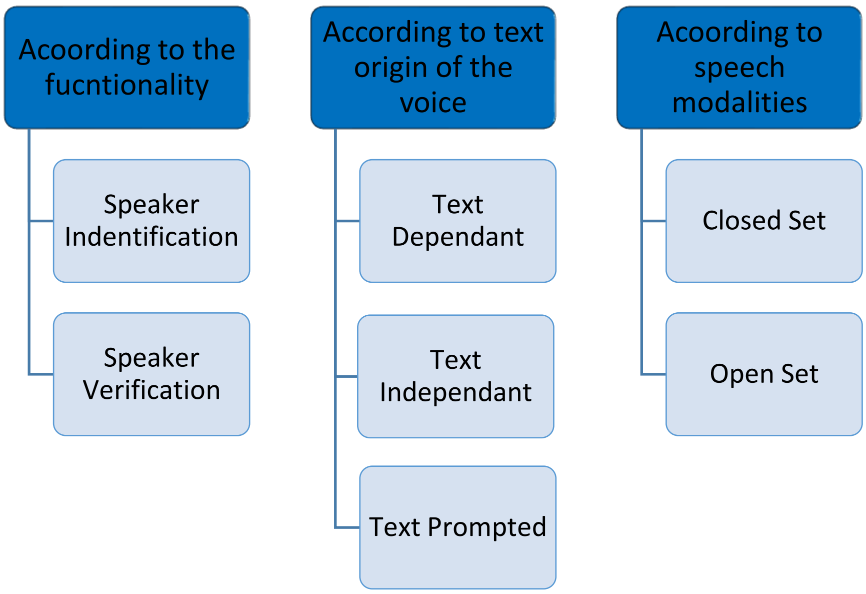

Figure 2.2 Different approaches of voice recognition approaches

- **Speaker Identification:** Identifying an unidentified speaker from a set of formerly known individuals by comparing a voice sample with a saved voice samples set.
- **Speaker Verification:** Speaker Verification is verifying a speaker is based on the provided test speech utterance and previously provided speakers train speech utterance (accepted or rejected)
- **Text Dependent:** Previously known words /sentences used for training and recognition
- **Text Independent:** Do not depend on the text or words used are not known previously
- **Closed Set:** The voice that should be identified belong to a set of registered speakers
- **Open Set:** The voice that should be identified belong unregistered speakers

An overview of a speaker recognition framework is discussed in brief within this chapter, which includes a variety of feature extraction and modeling methods. These methods bring up various levels of performance according to the environment implemented, input data and may still have a variety of drawbacks.

CHAPTER 3

METHODOLOGY

The methodology chapter walks you through steps in which theoretical knowledge gained in literature review is applied in order to achieve the expected outcome. As mentioned in the previous sections many mathematical and computational approaches have been followed in various voice authentication platforms, where voice biometric is used as an authentication method and final solution for the identified problems is to implement a voice-based user authentication n mechanism for online examinations.

This chapter provide an overview of the approach that is followed in design and implementation steps which have been carried out so far and planned further during the project to make it successful. The main part in the voice recognition system is the speaker identification phase. Here we present the algorithms and the techniques considered within the speaker recognition system: Front-end Processing, Feature Extraction, Speaker Modeling, Pattern Matching or Logical decision.

3.1 Problem Analysis

In today's world, teaching and learning have accommodated the new ways of using internet and technology, whereas the exams are continued in general paper pen-based ways. Online education has been practiced by some higher education institutes and along with that online examination systems were developed. But these were not that popular in the developing countries due to inability of acquiring the relevant resources and internet facilities.

Because of the COVID-19 outbreak many educational institutes were closed and as a result, the teaching and learning activities had to face an emergency transition to online learning and assessment. (Reedy, et al., 2021). Due to this online education and examinations has emerged as a possible technological method to education at the present and researchers state that education will never return to the stage where it was before after this (Rahme, et al., 2021).

Examination is a basic requirement in education. As a result, educators are concerned about establishing a secure means of conducting exams using e-learning strategies as they adjust to online education. These online examinations should be available to every student while ensuring the credibility and transparency and should be able to be accessed with minimal hardware requirements.

In this process one of the main challenges that they came across was user authentication. Though there were secure authentication methods, there were multiple challenges in implementing them due to lack of required hardware facilities and they had to move on with traditional username/password authentication methods allowing the students to take an exam from home.

In Sri Lanka also, most of the higher education instituted has adapted to online education approach in order to continue teaching learning activities during the prevailing pandemic situation on the country. Along with that they have started conducting inline examination for remote learners using e-learning platforms.

In order to obtain a basic idea on the online education and examination an open online survey was conducted targeting students of higher education students (Appendix 01). This was conducted as an open survey shared among target group of undergraduates who are currently engaged in higher education in private, government and semi- government institutes. The responses were collected via a Google form shared among randomly selected participants using an open questionnaire as it is a more reliable source of collecting data though internet. The participants were directed to a set of structured questions generally based on Education Status, Institute Type, Online Education and Exams. Responses from 150 participants were collected and analyzed. The collected data was analyzed and presented as pie charts to view the general distribution of data. As this was to gain a primary understanding of the existing background of online education and exams, this was rather extended to a qualitative analysis. Below given charts indicate the results obtained by the survey.

As in the graph illustrated in Figure 3.1, Participants of the survey were distributed among government, semi government and private institutes.

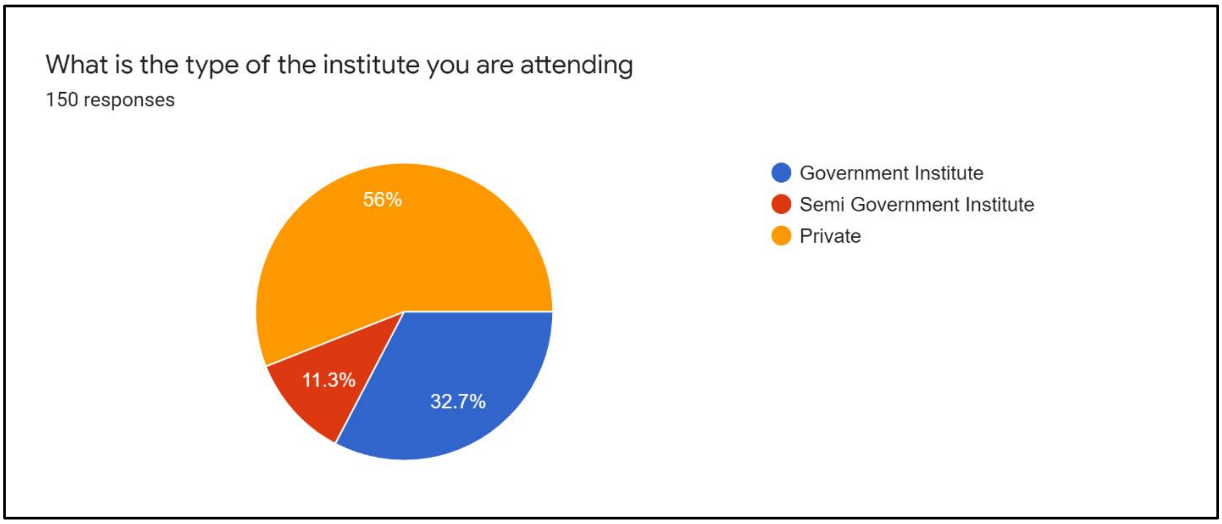

Figure 3.1 Types of the institutes of the participants

In the below, Figure 3.2 we can observe that around 2/3 portion on the students participate in online examinations. That reveals that the Sri Lankan education institutes have moved into online platforms and conducting online examinations.

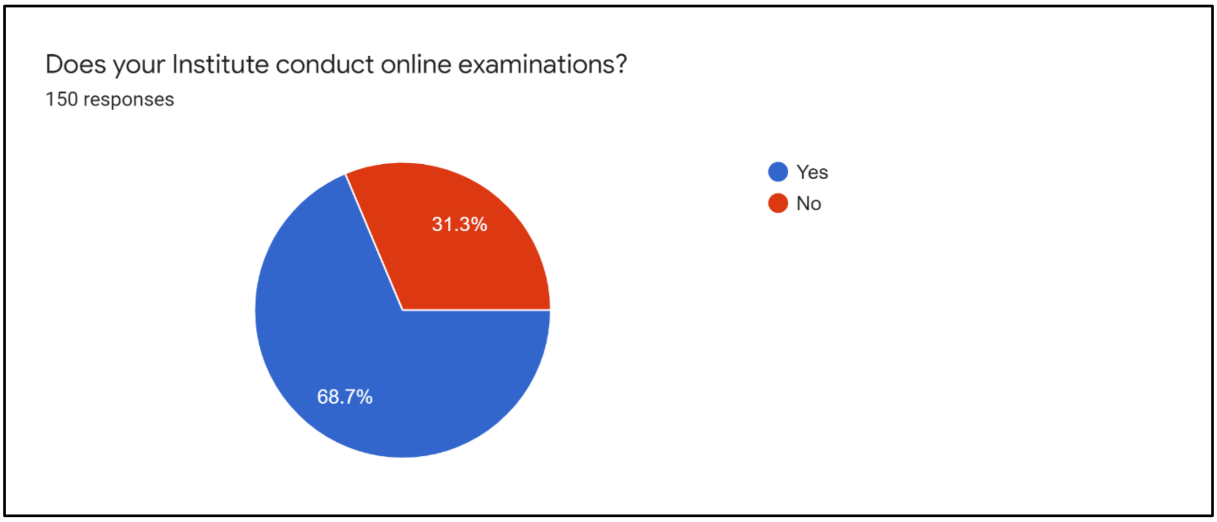

Figure 3.2 Availability of online examinations at institutes

The survey was included with a question asking about the authentication methos that are being used in the institutes where online examinations are held. According to the outcomes as in the graph included in Figure 3.3 of the survey done, we can observe that more than 2/3 of the institutes use traditional username/ password for user authentication. But this cannot be considered as a suitable authentication methodology because the credentials can easily be shared with another person. Therefore, we are going to extend the verification process in to two steps: Username/Password Authentication and Voice based authentication to ensure the security to a higher level.

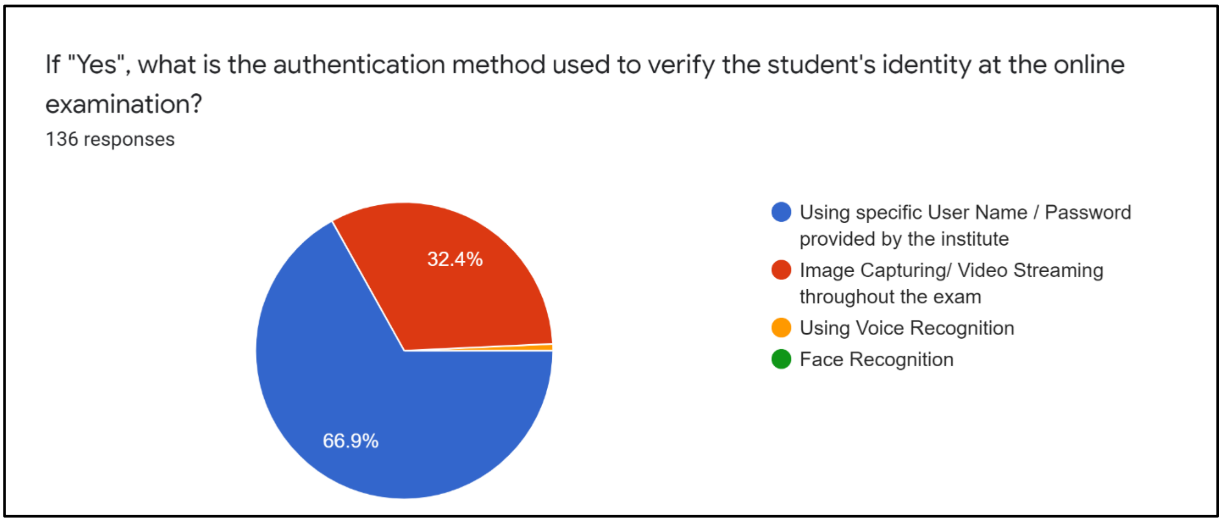

Figure 3.3 User authentication methods used in online examinations

Since we are going to implement a voice-based technique in authentication, question was added to check whether the participants possess required hardware facilities: headphone/mic/headset and as presented in the graph in Figure 3.4, more than 87% of the participants stated that they already have required facilities. If someone don't, it needs a very low cost to achieve them, and they are available fairly for everyone.

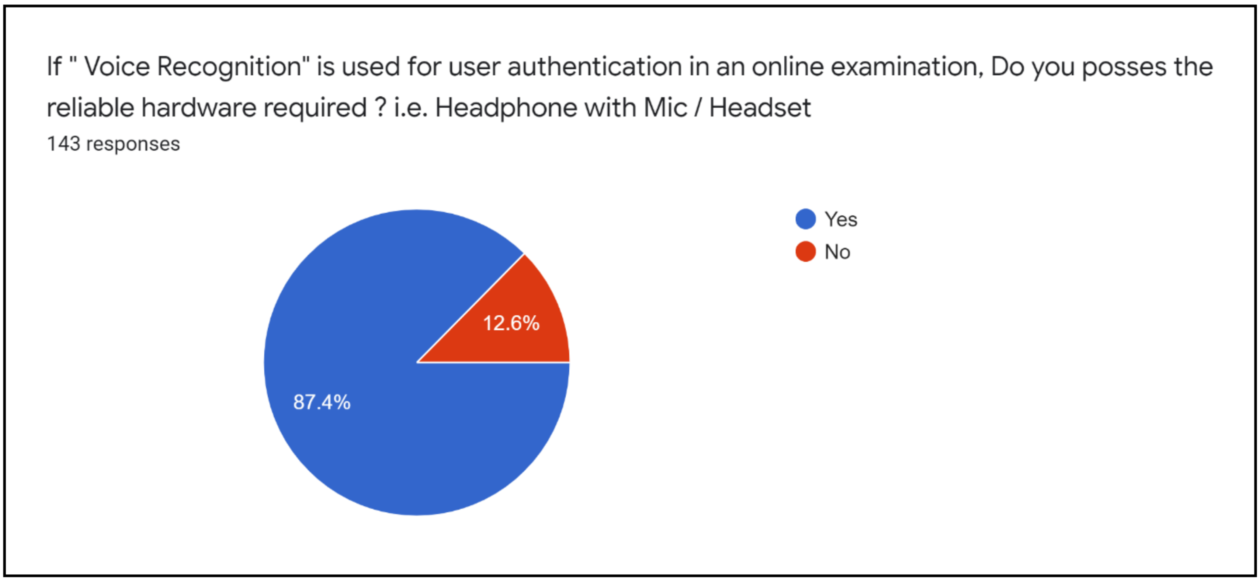

Figure 3.4 Availability of required hardware for voice authentication implementation

Therefore, we can observe that the basic requirements for implementing a voice-based user authentication are full filled by most of the students. So, we can obtain more accurate user authentication in online examination by adding a voice verification step in exam enrollment and reduce because it is very important to follow secure authorization process to verify that the correct candidate is taking the exam in a safe environment free of malpractices.

3.2 Proposed Design

As identified in the problem analysis, and through the basic ideas gathered through the online survey done we can observe that most of the institutes do not follow secure procedure in user authentication in online examinations. Most of current systems depend on username/password authentication which can allow students to engage in various malpractices at the exam due to unavailability of proper authentication process.

The proposed system will affect in increasing the accuracy of the user authentication process with 2-step authentication based on username/password method and voice authentication. Therefore, we can implement this proposed system in online examinations environment to achieve successful results.

Strengths of proposed system

- **Collectability:** Obtaining voice samples are inexpensive and no expensive equipment's are required.
- **Portability:** Speaker verification is simple to use and requires minimum computing power but provide high accuracy.
- **Acceptability:** Speaking is a normal activity that does not necessitate any uncommon behavior.
- **Accuracy:** Higher accuracy can be obtained from since there are 2-step verification of user with two approaches username/password and voice

The Figure 3.5 shows the high-level diagram of the system designed in the user's perspective.

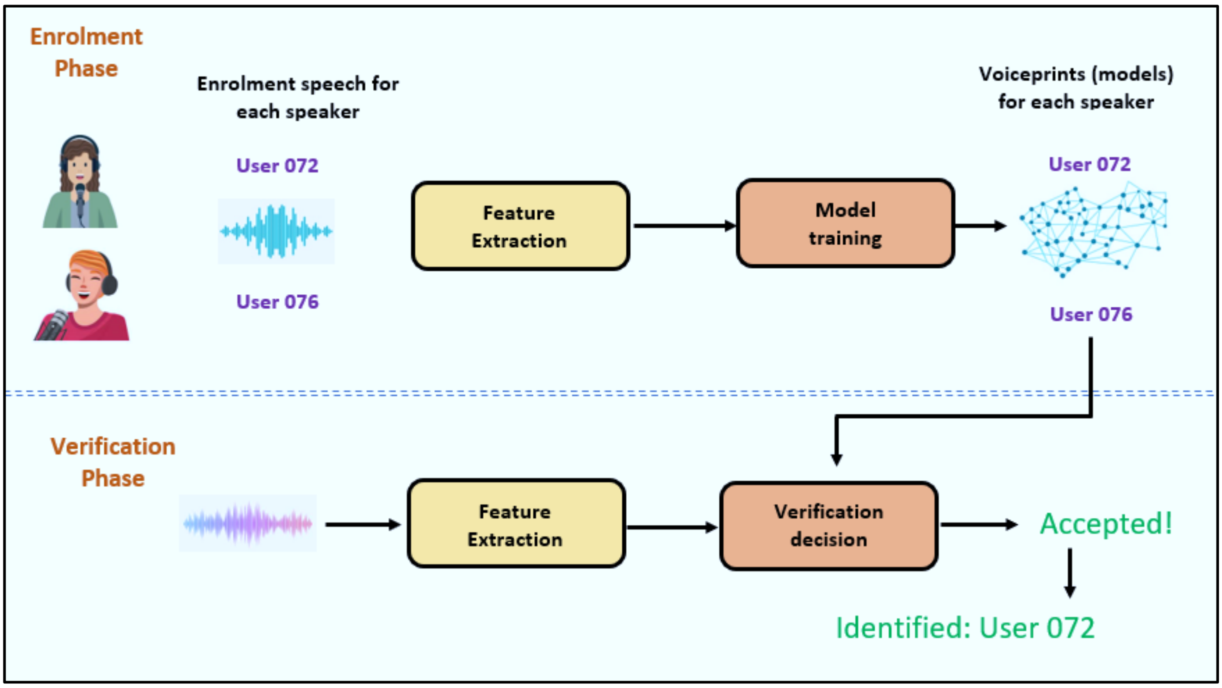

Figure 3.5 High Level Diagram of the system process

The Figure 3.6 shows the flow of actions of the system designed in the user's perspective.

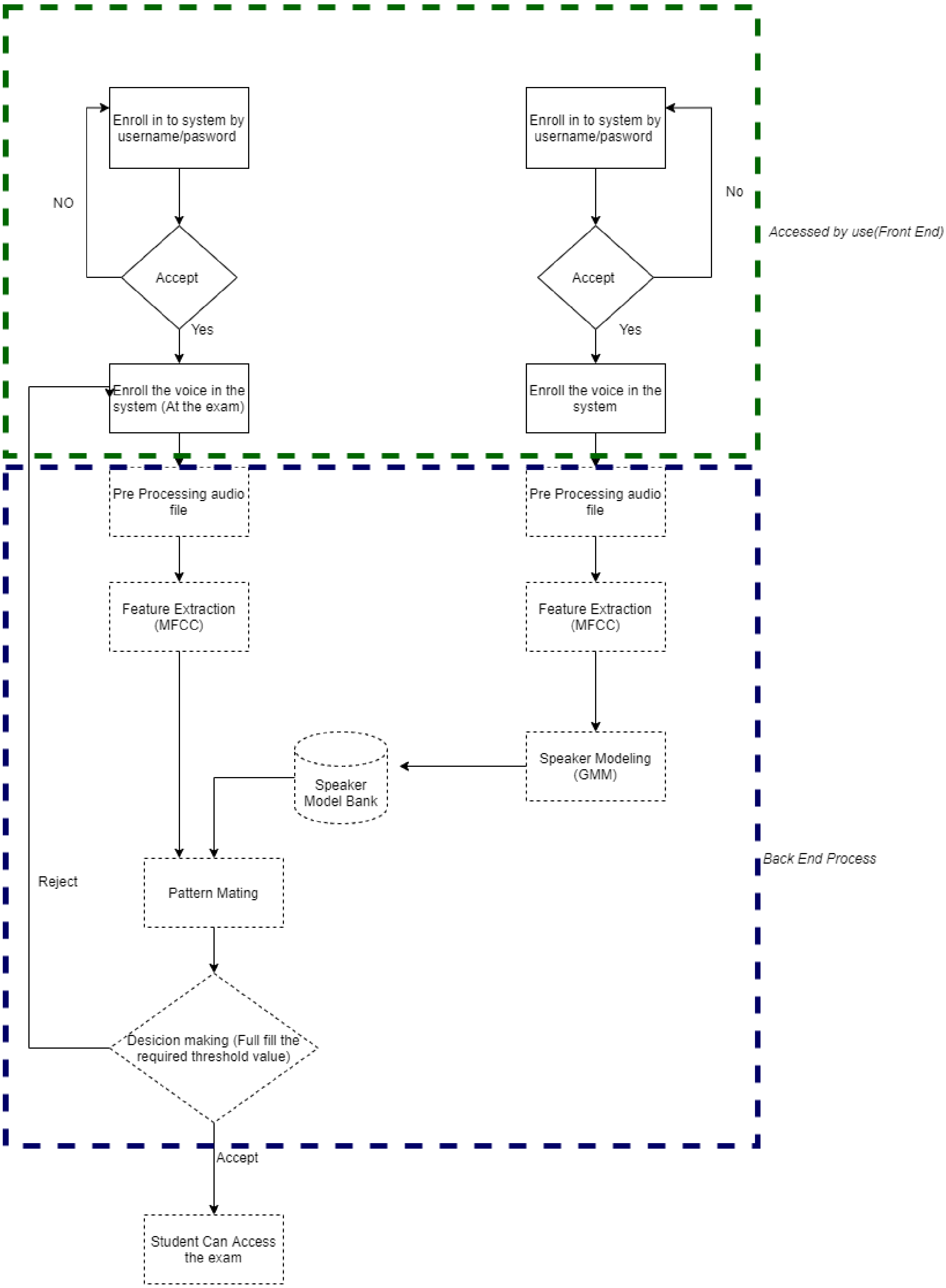

Figure 3.6 Flow Chart of functions of Proposed User Authentication System

3.3 Speaker Verification (Voice Authentication)

Speaker verification can be addressed as a binary class problem of a speaker is either accepted or rejected based on the test speech utterance and previously provided speakers train speech utterance (E.Chandra, et al., 2014). It is a one-to-one match, where the voice print is matched with one template. The unknown speaker's voice sample is compared with the model for the speaker who it claims to be. If the match is above a specified threshold the verification is accepted. In order to avoid imposters to access the system always a higher threshold value is maintained in such a system. A voice verification system consists of following steps in common behalf of the techniques and models used.

a) Input data gathering & Preprocessing
b) Extracting Features
c) Speaker Modeling
d) Pattern Matching for Decision / Verification

The following Figure 3.7 displays the basic steps of a voice/speaker verification system. These steps are discussed in detail in the following sections.

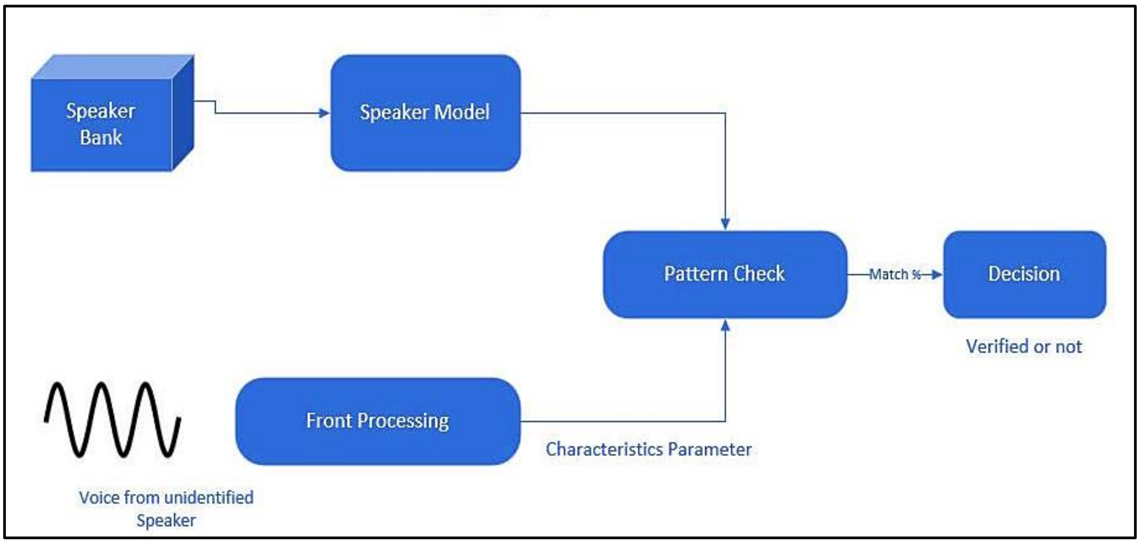

Figure 3.7 Steps of Speaker Verification System

3.4 Voice Signal Characteristics and Pre-Processing

Here the users voice is captured by the system or feed to the system at the first phase. The characteristics and features of the speech of the speaker's voice are included in the audio signal. Voice signals change randomly, continuously with time and are unique to everyone even if they are speaking the same text.

The below graphs (Figure 3.8 & 3.9) show the amplitude and the frequency of two persons speaking out the same text. There we can see the variations of the amplitude and the frequency with time, and we can identify that they possess different features.

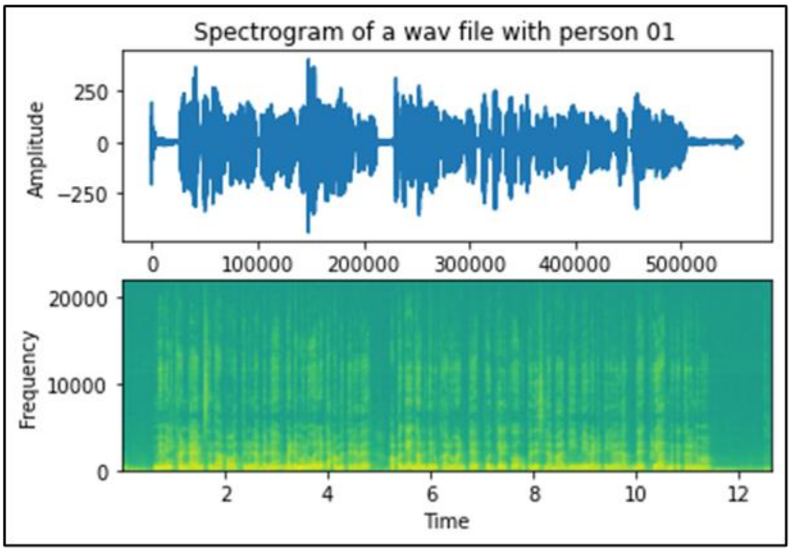

Figure 3.8 Spectrogram of Person 01 speaking sample text 01

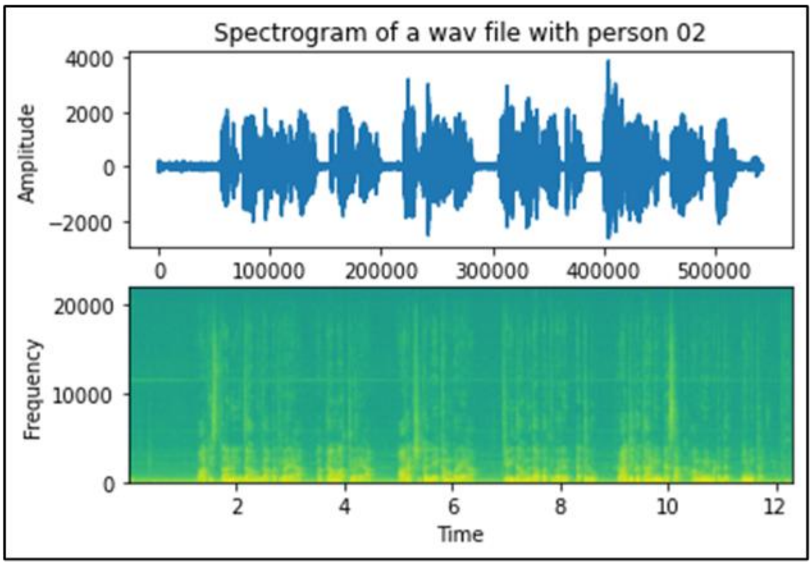

Figure 3.9 Spectrogram of Person 02 speaking sample text 01

3.5 Dataset

The dataset that is used to evaluate the speaker verification system that is being implemented is obtained from the UCSC/LTRL Speech corpus. The dataset contains voice samples of 60 individuals (male and female) with unique identities.

The dataset contains voice samples with duration around 1s-15s speech utterances, where speaker is speaking a different text at each voice sample. According to the requirement, portions of the dataset is acquired varying the duration and the number of samples considered.

These voice samples are generated by reading texts based on Sinhala language. Since the system implemented is based the Sri Lankan education environment, evaluating the system with the native language of Sri Lankans is important since there could be users that are not familiar with English language, where most available speaker verification systems are based on.

3.6 Feature extraction

Chen defines feature extraction as an approach that is used in converting a voice signal to a series of acoustic feature vectors that carries unique characteristic information about the signal, which can recognize the speaker (Chen, 2009) . The feature extraction is required to represent. the speaker information present in the speech signal. A characteristic feature vector with lower physical or spatial properties is generated by the feature extraction algorithms.

Many feature extraction algorithms have been used in previous research, such as Mel Frequency Cepstral Coefficient (MFCC), Linear Predictive Cepstral Coefficient (LPCC), and Perceptual Linear Predictive Cepstral Coefficient (PLPC) discussed in the previous chapters. According to the recognition accuracy obtained in the previous work, Mel Frequency Cepstral Coefficient (MFCC) is used for feature extraction in this study.

3.6.1 Mel frequency cepstral coefficients (MFCC)

Sounds can be represented in two ways, Linear Cepstral and Nonlinear Cepstral. In deriving the MFCC, the nonlinear cepstral of the sound is used. Than linear cepstral representation, Mel scale used in MFCC is more considerable in representing the human auditory system (Sunitha & Chandra, 2015) .This method is focused on variations in critical bandwidths and frequency in the human ear. Low-frequency signals are spaced linearly, whereas high-frequency signals are spaced logarithmically. This is used to record phonetically relevant aspects of speech. Following Figure 3.10 shows the steps in MFCC approach.

Figure 3.10 Steps in MFCC Approach

After the voice sample is acquired from the user, the following process is feature extraction from the voice. Therefore, as mentioned above MFCC approach is used which contains the following steps.

- **Pre-emphasizing:** In pre-emphasizing higher frequencies are filtered in balancing the spectrum of sound sample that lies in the outrageous higher frequencies range.

- **Framing**
  Framing divides the voice signal into frames with 30~20 ms to overcome FFT producing distortions. Since an audio is non-stationary, audio is assumed to be stationary for shorter periods of time.

- **Windowing**
  In the windowing process the audio is converted to frequency domain process to time domain process. It ensures overcoming the discontinuity of the audio since FFT assumes that the audio is periodic and continuous. The graphs in Figure 3.11 shows the original frame and the frame after windowing.

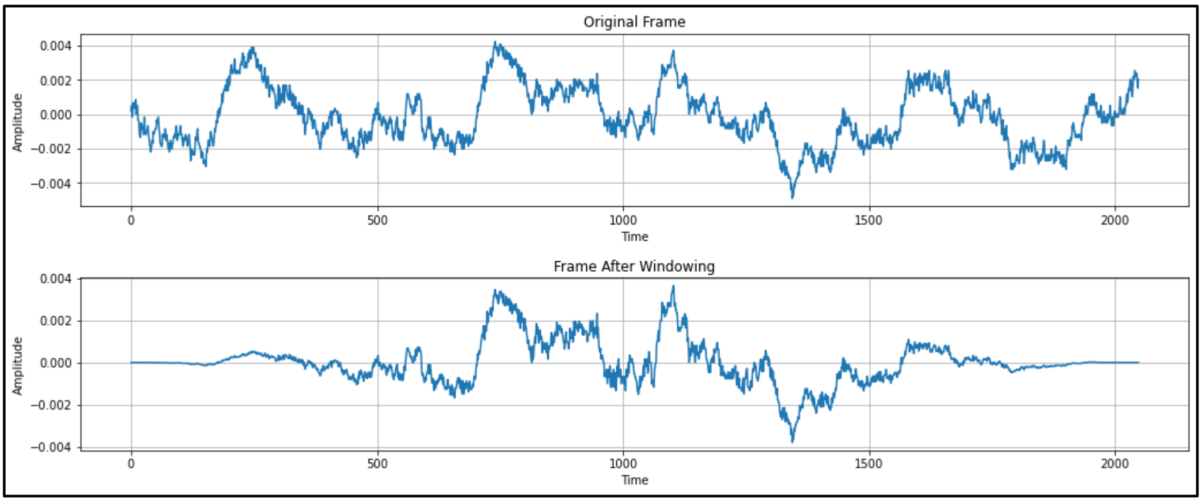

Figure 3.11 Original Frame and Frame after Windowing

- **Fast Fourier Transform**

  There are two main usages of FFT in this process in converting the signal from time domain to frequency domain and in obtaining magnitude responses of every frame. Within FFT we assume that the signal is periodic and continuous in wrapping around in each frame.

  To determine the star and end points, filter points are defined at first. In order to do that filter bank edges are converted to MEL space. Next, between two MEL frequencies linearly spaced array is constructed and the array is converted to frequency space. At the final step the array is normalized to the FFT size and relative FFT values are chosen.

  This process is shown in the Figures 3.12 and 3.13 below:

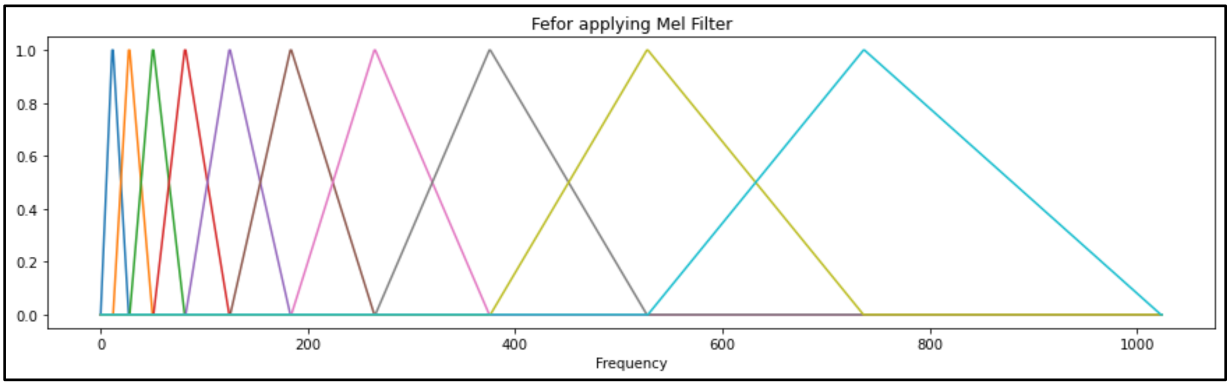

Figure 3.12 Before Applying Mel Filter

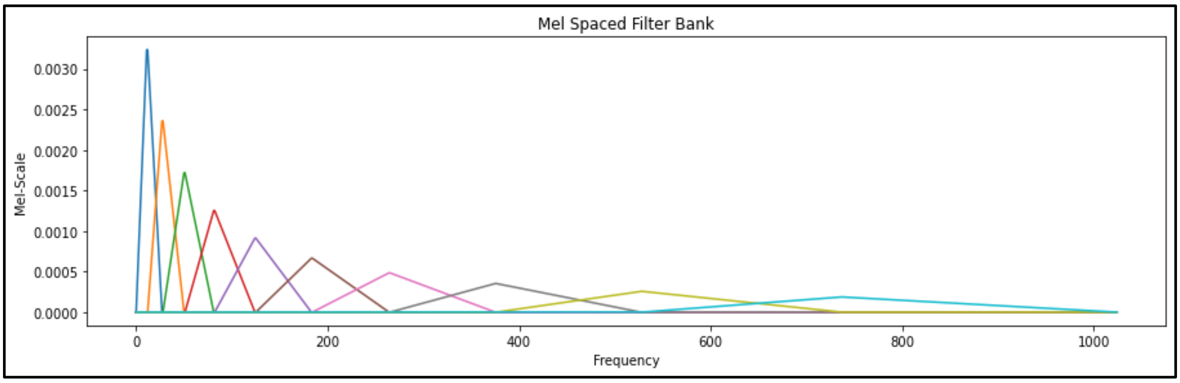

Figure 3.13 After Applying Mel Filter

- **Mel Frequency Wrapping**

  Mel frequency is based on the studies of frequency range that can be acquired by a human. Since the humans are less sensitive to frequencies in higher ranges. So in this step above generated spectrums are mapped in to Mel scale, making it easier to get estimations of the energy at certain points.

- **Discrete cosine transforms**

  The reason for performing DCT is to convert the log Mel scale cestrum from the frequency domain to the time domain. The obtained feature is called the Mel frequency cestrum coefficient. The Figure 3.14 below is the MFCC obtained for the voice sample.

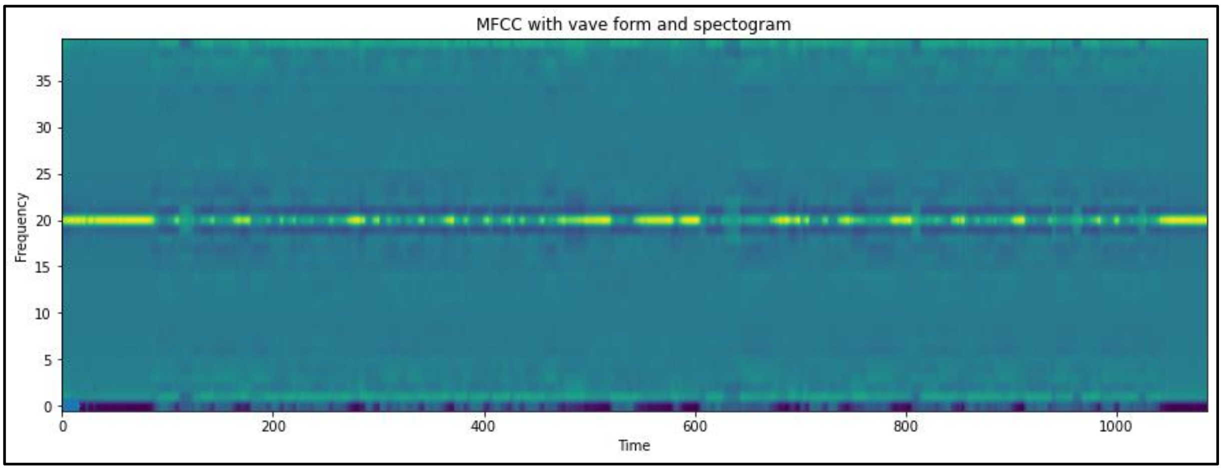

Figure 3.14 MFCC wave form & Spectrogram

MFCC Implementation

Let us consider each frame consist of 'N' samples and let its adjacent frames be separated by 'M' sample where M is less that N. Hamming window is used in which each frame is multiplied. Mathematically, window equation is given by:

$$\boldsymbol{W(n) = 0.54 - 0.46\ \cos(\frac{2\pi n}{N-1})}$$

Now, Fourier Transform (FT) is used to convert the signal from time domain to its frequently domain. Mathematically, it is given by:

$$\boldsymbol{X_k = \sum_{i=0}^{N-1} x_i\ e^{\frac{2\pi ik}{N-1}}}$$

$$\boldsymbol{M = 2595 \log_{10}(1 + \frac{f}{700})}$$

In the next step log Mel scale spectrum is converted to time domain using Discrete cosine Transform (DCT). Mathematically, DCT is defined as follow:

$$\boldsymbol{X_k = \alpha\ \sum_{i=0}^{N-1} x_i\ (2i + \frac{1}{2N})}$$

The result of the conversion is known as MFCC and set of co-efficient is called acoustic vector.

Here in this study, 40 dimensional features are extracted from the voice frames: 20 MFCC features and 20 derivatives of MFCC features with 25ms frame size and 10ms overlap between the frames. These derivatives review the characteristics and details of MFCCs with the time. The calculation of the delta MFCC and joining them with the 20 dimensional original MFCCs effecting in increasing the efficiency and accuracy in performance in voice and speech analysis

For implementing MFCC in this research, python speech features package is used. For each wav file that is used in training, this python function extracts features and build a 40-dimensional feature vector.

Python Implementation of MFCC

```
Import python_speech_features as mfcc

def calculate_mfcc(ary):

#Generate delta of eigenvector matrix

    Row, column = ary.shape
    delta_mtx = np.zeros((rows,20))
    N = 2
    For i in range (row):
        array_idx = []
        j = 1
        And j <= N:
            If i-j <0:
              f = 0
            Other:
              f = i-j
            If i+j> rows-1:
                s = line -1
            Other:
                s = i+j
            arr.append((s,f))
            j+=1
        delta_mtx[i] = (ary[array_idx[0][0]]ary[array_idx[0][1]] + (2 * (ary[array_idx[1][0]]-
ary[array_idx[1][ 1]]))) / 10

          Return delta_mtx

def mfccfeats (audio, rate):

# Generate 40-dimensional feature vector by combining dekta and 20-dimensional fear vector from
audio

    mfccfeat = mfcc.mfcc(audio,rate, 0.025, 0.01,20,nfft = 1200, appendEnergy = True)
    mfccfeat = preprocessing.scale(mfccfeat)
    deltas = calculate_delta(mfccfeat)
    dim40_vec = np.hstack((mfccfeat,deltas))
    Return dim40_vec
```

3.7 Speaker Modeling

During the enrollment process, the feature vector which is obtained through feature extraction process is used to construct the speaker model that possess the following characteristics:

- Theoretical approach of modeling with a mathematical model for system implementation
- Appropriate size and computation representation.

For speaker verification systems, there are a variety of modeling techniques that depend on both these characteristics. The modeling technique used highly rely on the characteristics of the speech samples acquired for training and expected performance depends on training, storage and computation considerations.

Through referring the previous literature, it was identified that GMM most widely used classifier and it has performed well when compared with classification models since it has shown a better

3.7.1 Gaussian Mixture Model

Gaussian Mixture Model is used to modeling the probability density function of a multi-dimensional feature vector. GMM is a type of probabilistic clustering model that is used in representing the existence of sub populations within a entire population. In training a GMM model, the probability distribution of a class is approximated by combining 'c' Gaussian clusters (components of GMM) .

Given a speech feature vector $X = \{X_i\}$ of dimension F, thet probability density of X_i given a C GMM speaker model λ is given by:

$$\boldsymbol{p(X_i|\lambda) = \sum_{C=1}^{C} W_c g(x_i, \mu_c, \Sigma_c)}$$

$$\boldsymbol{\sum_{C=1}^{C} W_c = 1}$$

Following Figure 3.15 displays how GMM acts in speaker recognition system in speaker modeling and scoring.

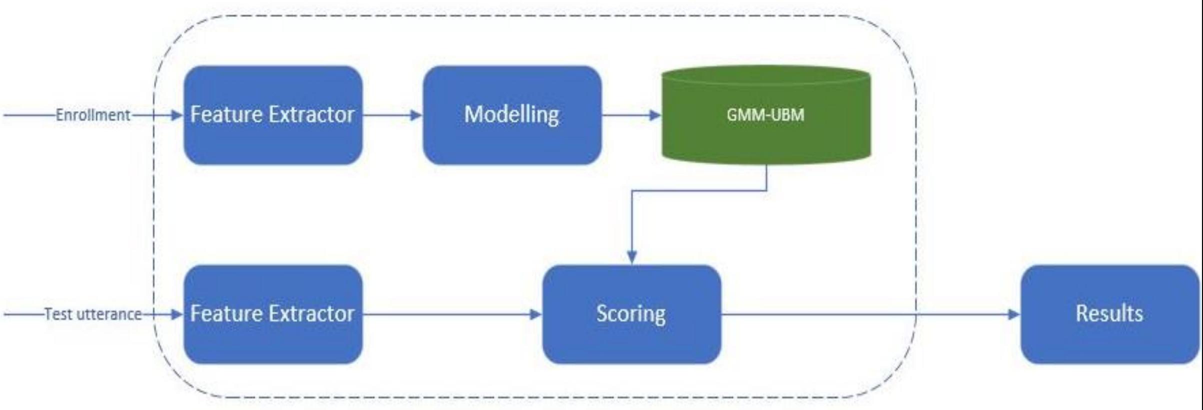

Figure 3.15 GMM Steps in a block diagram

As discussed above feature matrix with 40 dimensional MFCC and Delta MFCC features is generated in feature extraction process and a GMM is learned from that using the Python sklearn.mixture package. A GMM speaker model with 16 gaussian components is trained using the below python file and unique models are created for each speaker and speaker models are dumped as .gmm files using python _pickle package. According to the data available in the da data set, models with following characteristics displayed in the Table 3.1, below are generated in this study

Table 3.1 Generated GMM Models

Model	No. of Speakers considered	Voice Signal Length used for building Model	No. of Voice Samples Used for each model	No. of Individual voice models generated
Model 1	60	4s	5	60
Model 2	60	4s	10	60
Model 3	60	10s	5	60
Model 4	60	10s	10	60

The generating of GMM models in python is implemented as below.

Python Implementation of GMM Models

```
import _pickle as cPickle
import numpy as np
from scipy.io.wavfile import read
from FeatureExtraction import mfccfeats
from sklearn.mixture import GaussianMixture as GMM

n = 1
# Using 10 sound files form each speaker to extract mfcc fedatures

feat_array = np.asarray(())
for paths in file_pathss:
   paths = paths.strip()
# reading the audio file
   sr,audio = read(train_set+paths)

   # Generate 40-dimentional feature feat_vectors by combiing dekta and the 20 dimnetional fearure
feat_vector from a audio
   feat_vector = extract_feat_array(audio,sr)
   print( "Audio feat_vector Dimentions =", feat_vector.shape)

   if feat_array.size == 0:
      feat_array = feat_vector

   else:

      feat_array = np.vstack((feat_array, feat_vector))
# As 10 files are consideres and concatinated, proceed to train the model

   if n == 10:

      gmm = GMM(n_components = 16, max_iter = 200, covariance_type='diag',n_init = 10)
      gmm.fit(feat_array)


      # Whne the training is colpleted, dump the Gaussian Mixture Model

      pickle_file = paths.split("/")[0]+".gmm"
      cPickle.dump(gmm,open(dest + pickle_file,'wb'))

      print ('*** GMM model generated for Speaker',pickle_file," With Dimentions
",feat_array.shape)
      feat_array = np.asarray(())
      n = 0
   n = n + 1
```

3.8 Decision Making

When an unidentified voice sample is submitted for speaker identification, the procedure begins by extracting the 40-dimensional feature vector with 25ms frame and 10ms overlap between the frames for the relevant voice.

As the next step, log likelihood scores for each frame are required to be calculated for each frame in the sample voice against each speaker.

This process is repeatedly done for all the ‘c’ Gaussian components of the model, and the weighted sum of ‘c’ likelihoods are taken as the *‘W’* parameter in the equation.

$$\boldsymbol{p(X_i|\lambda) = \sum_{C=1}^{C} W_c g(x_i, \mu_c, \Sigma_c)}$$

$$\boldsymbol{\sum_{C=1}^{C} W_c = 1}$$

When the above values are applied and operated, the algorithms produce the log likelihood value for the frame, which is repeated for all the frames in voice sample and added together. The speaker is identified by considering the most significant highest value in likelihood scores.

Python Implementation for Decision Making

```
Import operating system
Import _pickle as cPickle
Import numpy as np
Import and read from scipy.io.wavfile
Import extract_features from FeatureExtraction
Import time

Define test record ():
#Training data path
test_dir = "Testing_test_audio/"
model_file = "Speaker_gmm_spkr/"

  spkr_model = [os.paths.join(model_file,file_name) for file_name in
        os.listdir(model_file) if file_name.endswith('.gmm')]

  # Load the trained voice model (gmm)
  gmm_spkr = [cPickle.load(open(file_name,'rb')) for file_name in spkr_model]
  Speaker = [file_name.split("/")[-1].split(".gmm")[0] for file_name
        In spkr_model]
```

```
  # Read test audio file

  Source, test_audio = read (test_dir + path)
  feat_vector = extract_features(test_audio,sources)
  #print(feat_vector)

  lg_likelihood = np.zeros(len(gmm_spkr))
For me in the range (len(gmm_spkr)):
    Print("i=",i)
    gmm = gmm_spkr[i] # Analyze each model
    Score = np.array(gmm.score(feat_vector))
    Print(scor.sum())
    lg_likelihood[i] = scor.sum()
    Print (lg_likelihood[i])

  win_spkr = np.argmax(lg_likelihood)
  Print ("Highest Scored Speaker", win_spkr)
  print ("\t The speaker is recognized as-",speakers[win_spkr])

  Time.sleep(1.0)

  Print ("Speaker recognition succeeded")

  Back (speaker [win_spkr])
```

3.9 System Implementation

For the purpose of demonstration, a prototype system was designed, which presents how the user interacts with the system. HTML, CSS, JavaScript and PHP was used in the development of this web-based prototype system.

As we can observe there are two main parts of the system, as enrollment and testing phase. The below Figure 3.16 shows the basic architecture of the proposed system.

Enrollment to Authentication Process (Training Phase)

Login for the examination (Testing Phase)

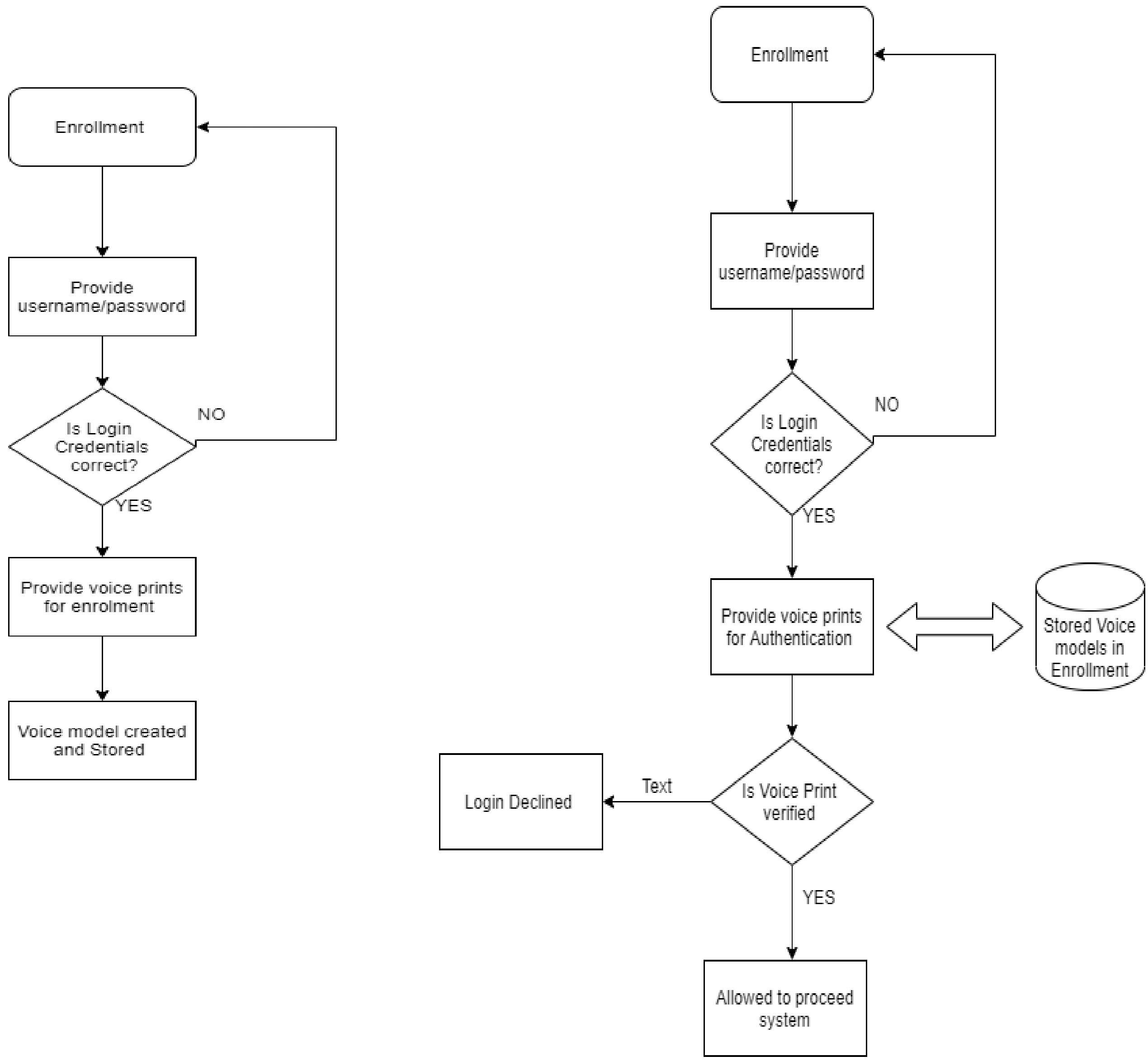

Figure 3.16 System Implementation

The general home page displayed as below Figure 3.17 of the system is presented with a set of instructions given to the users to address the users about the process of authentication.

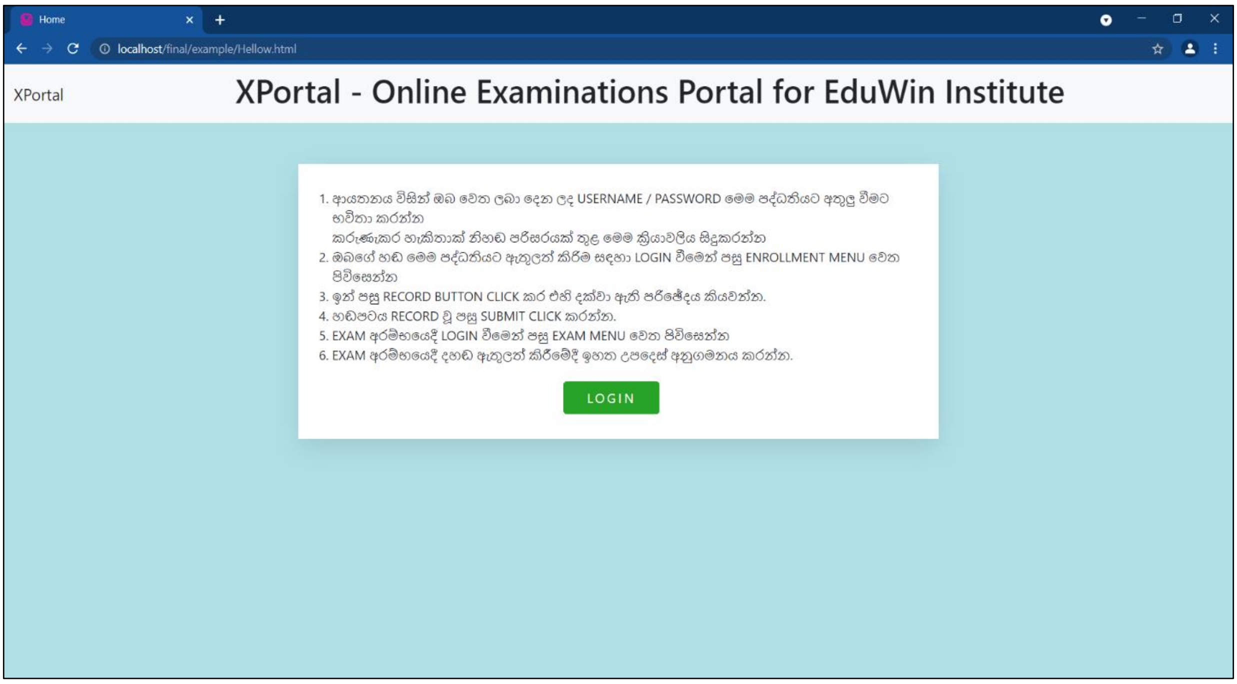

Figure 3.17 Welcome Page

To improve the security of the authentication process two-step verification is process is taken into consideration in order to add an extra layer of security. Therefore, users are provided with a previously defined username/password to log in to the system at both the enrollment and exam initiation as Figure 3.18.

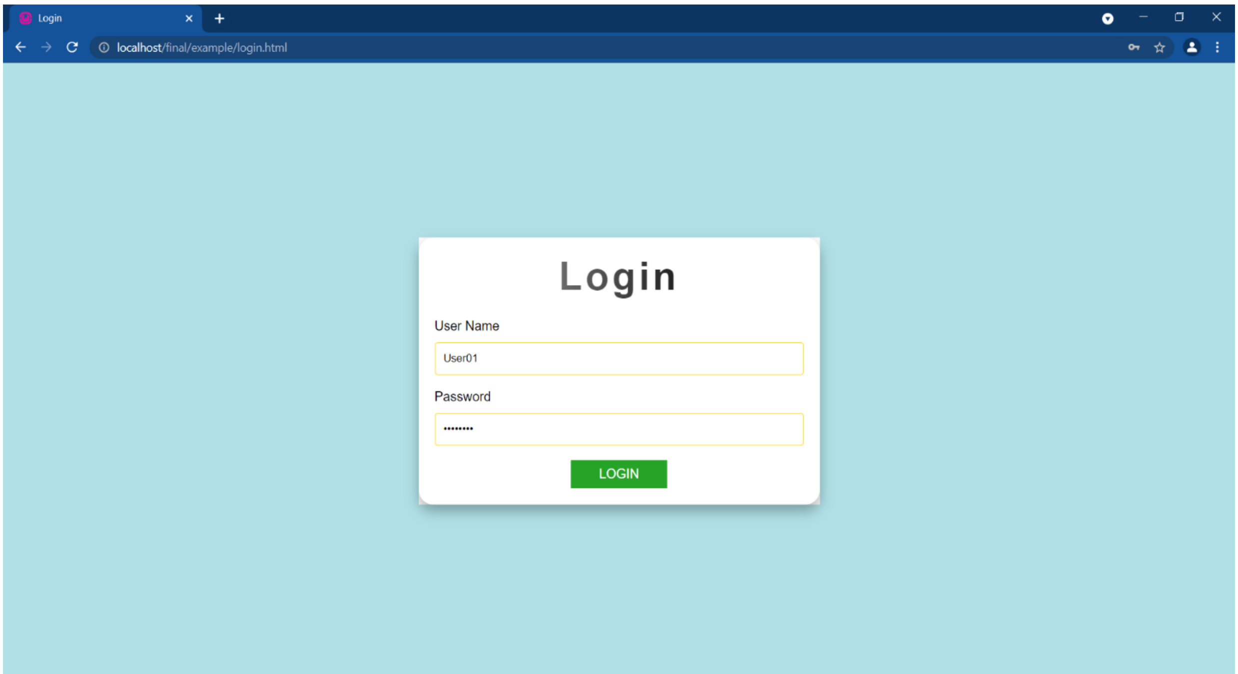

Figure 3.18 Login Page

Next the user is directed to the Main Home page (Figutr 3.19) , where the user is required to enroll with the system at first. Then the unique speaker model is generated with the users voice samples and stored.

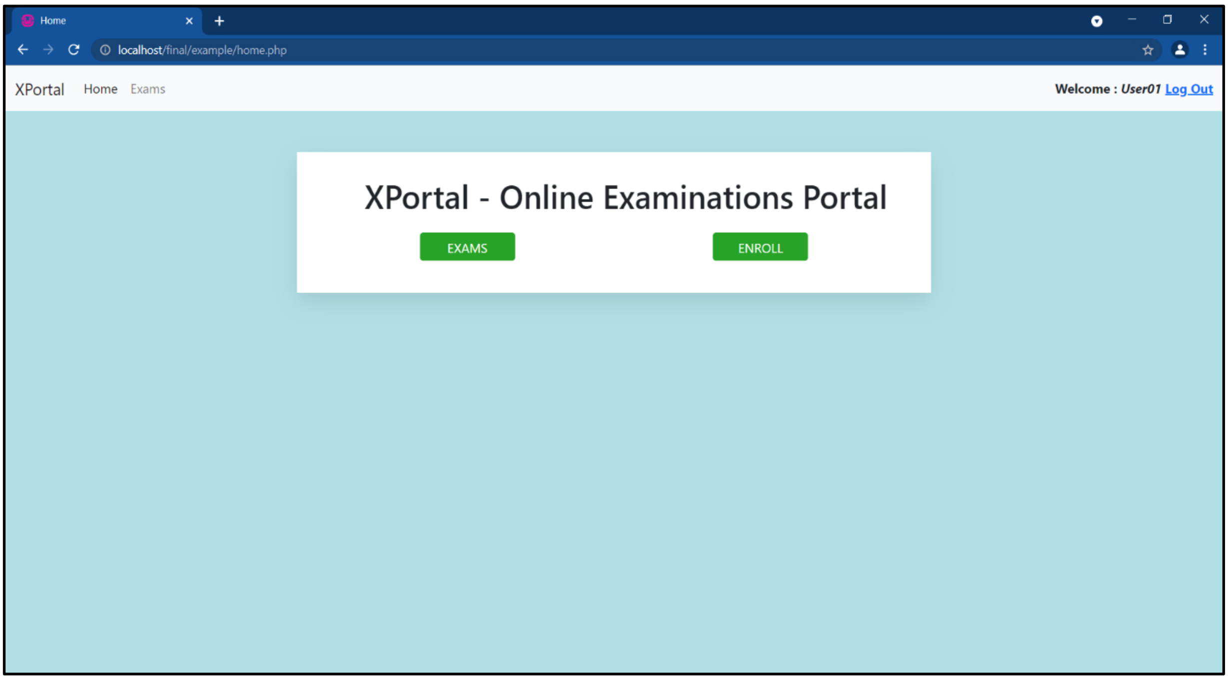

Figure 3.19 Home Page

Random texts are displayed in the screen as Figure 3.20 for acquiring the speech utterances from the user in the front end and are submitted for feature extraction and speaker modeling

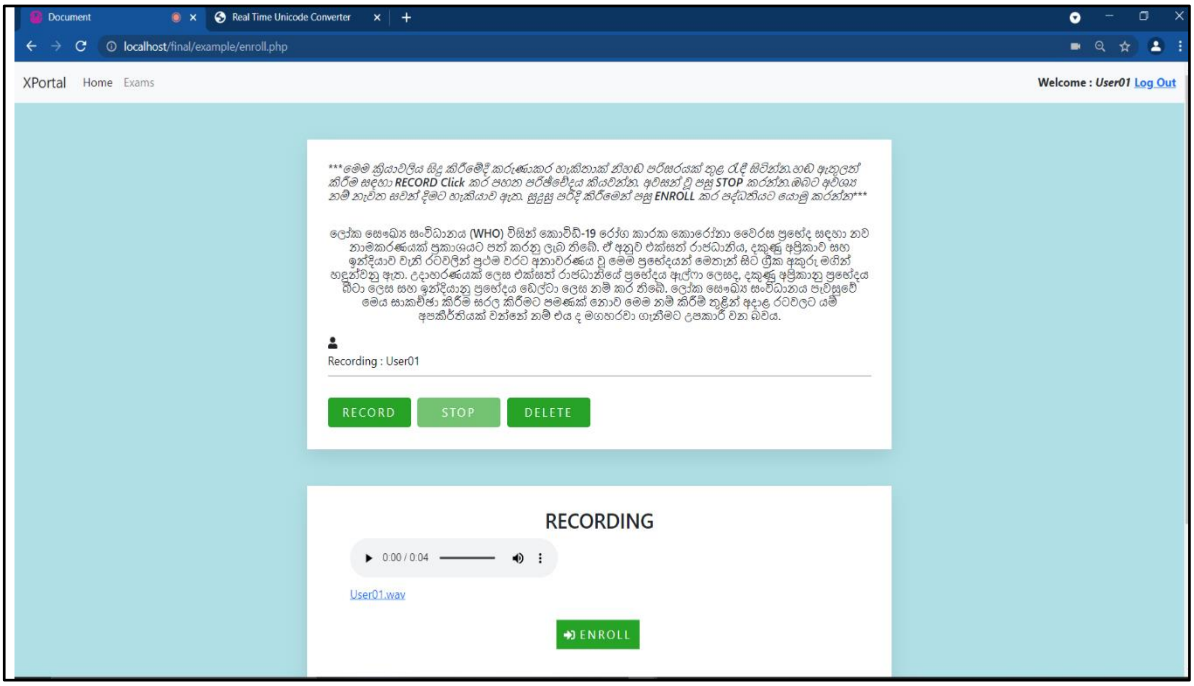

Figure 3.20 Enrolment Page

In authentication step, the procedure of acquiring the voice is same as the enrolment step shown in Figure3.21.

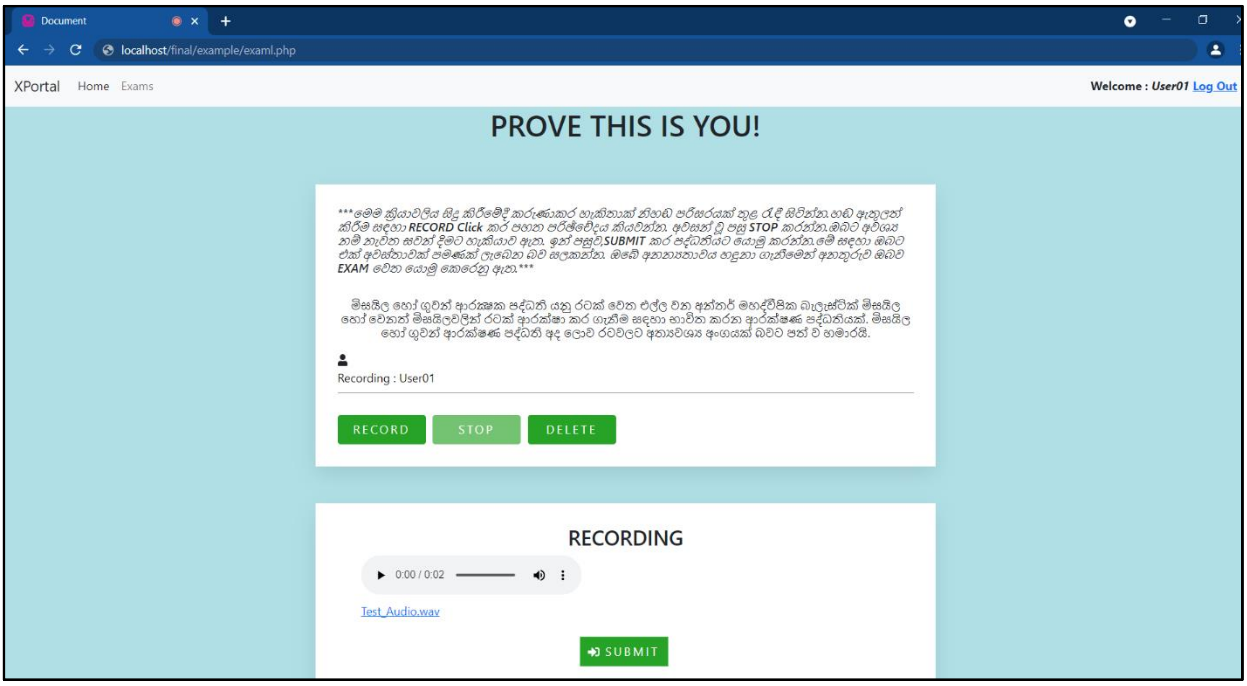

Figure 3.21 Authentication Page

The correctly identified useres are allowed to proceed to exam as Figure 3.22 after verifying the identity using both steps in two-step verification : username/password and speaker identification

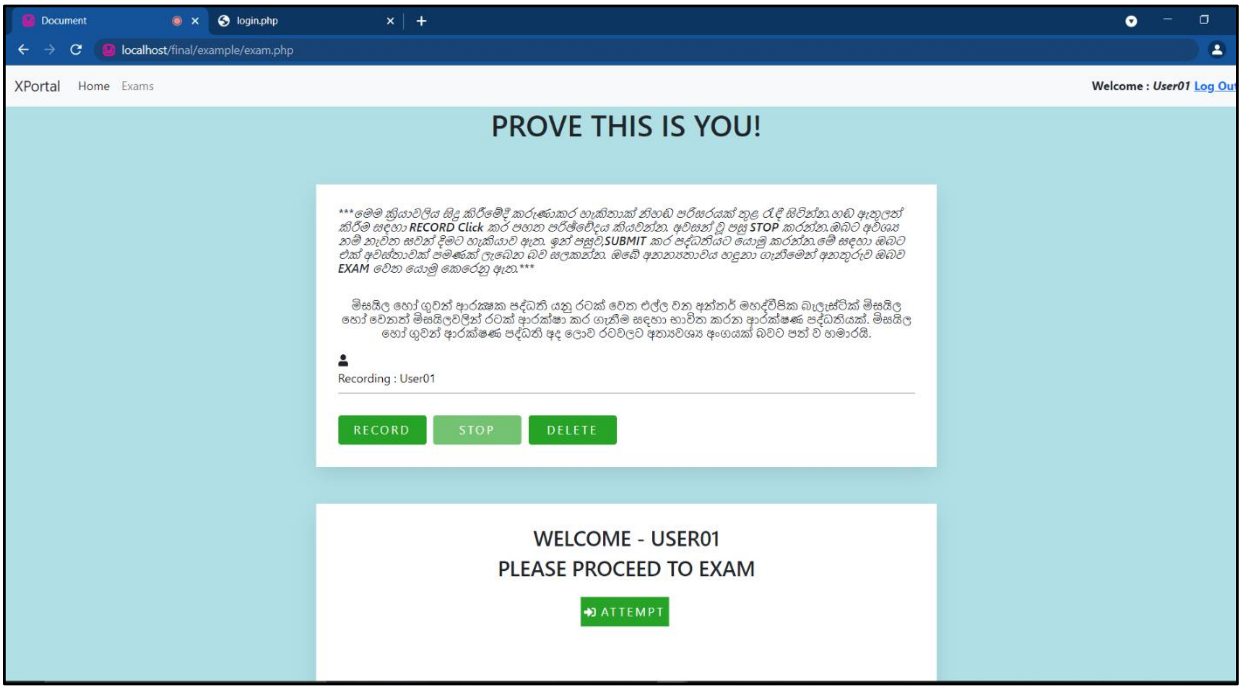

Figure 3.22 Welcome message after verification

After sucessful 2-step verification, user is allowed to view the questions as in Figure 3.23.

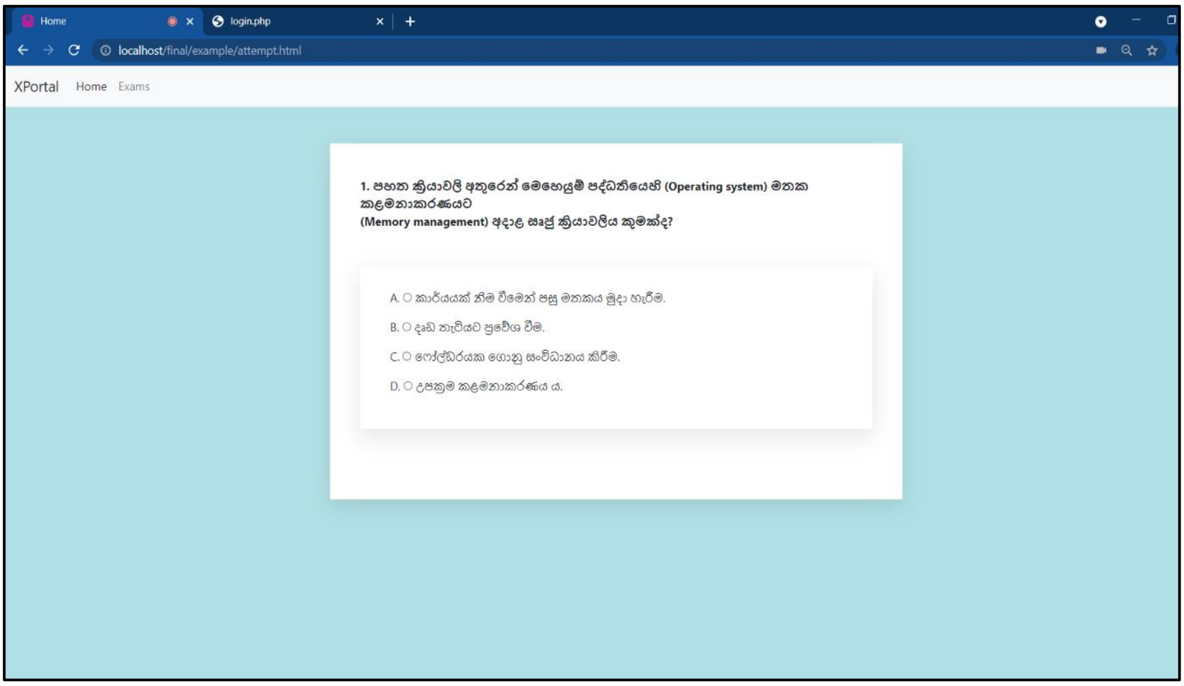

Figure 3.23 Questions page sample

The designed system is a prototype as mentioned above, which address the basic requirements of the presented approach for user authentication in an online examination. The data acquired from front end are passed to backend using python http (Hypertext transfer protocol) service and processed in required phases.

CHAPTER 4

EVALUATION AND RESULTS

This project proposed a Voice Recognition Model for Online Examinations. The model is designed so that it is easier to be used in remote location, without any specialized equipment's or supportive technologies are not required in implementing. This section presents how the how the testing and evaluation procedure is carries out in using the voice recognition models.

The results are tested with the specified objectives of the proposed system. The developed approach is tested against the dataset that was obtained from the UCSC/LTRL speech corpus which contains different speech utterances of Sinhala Language from individual speakers. The accuracy is measured based on the no. of correctly recognized speakers against the available trained speaker models along with different types of models defined with variant characteristics.

The experiments were conducted on an Intel® Core (TM) i5 – 10300H CPU(2.50-4.5)GHz with 8 GB RAM and NVIDIA GTX 1650 Graphics. The model is implemented in Python on Windows 10.

Due to the prevailing restrictions and difficulties in accessibility the system was tested using a previously gathered dataset as mentioned above. Therefore, a separate approach for the evaluation was designed which supports uploading a testing audio to the system. The uploaded file is tested against the previously generated GMM models, and the identified user is returned. The identified user is confirmed by checking the user that is logged in to the system is the same person identified. This alteration was only made for testing and evaluation purpose and stays as a different section of the prototype designed for system implementation.

Following Figures 4.1 and 4.2 show how the system is altered according to the current requirements

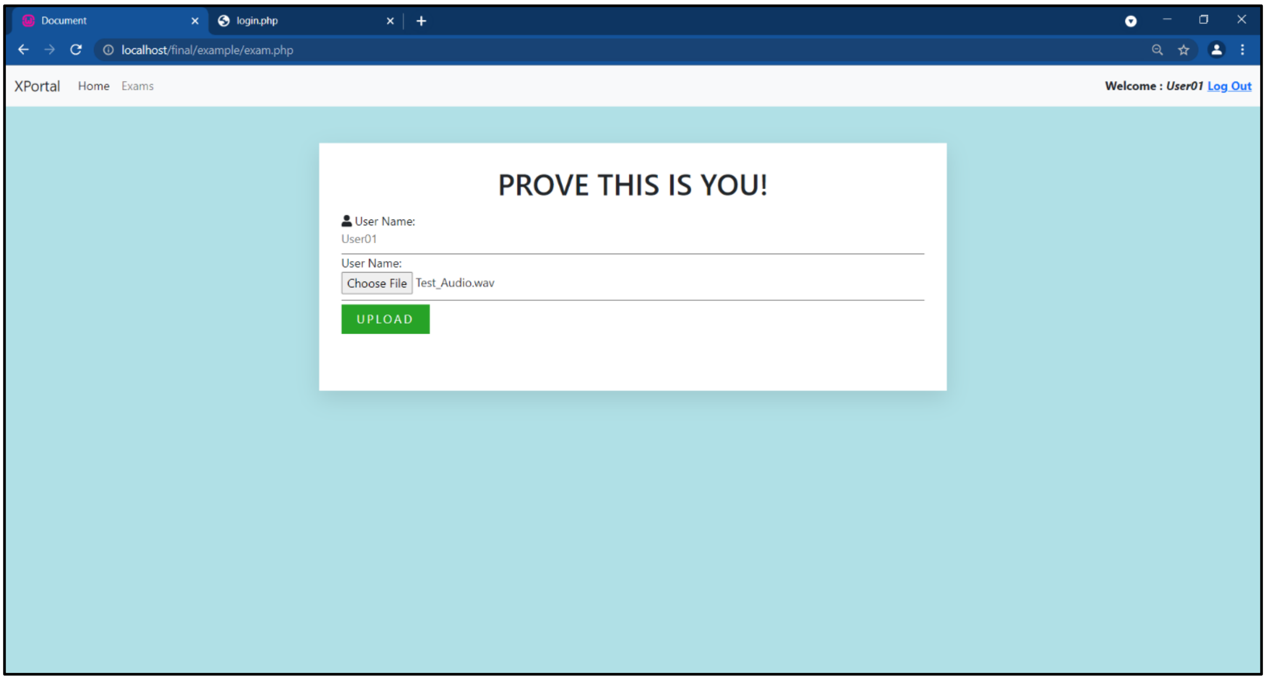

Figure 4.1 Page uploading test voice samples

Correctly identified users are displayed as below and others are redirected back to the home page.

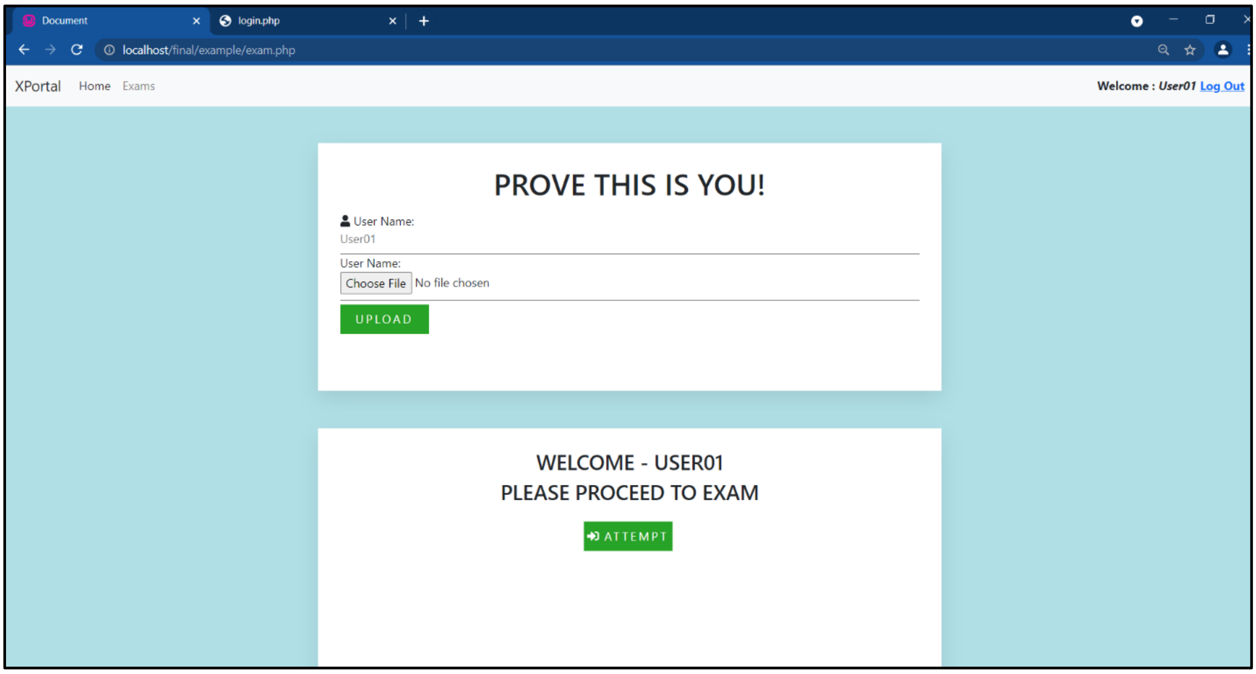

Figure 4.2 Welcome message for confirm identification

Features Extracted

Following are the feature vectors obtained from MFCC feature extraction. Two feature vectors extracted from the same person (072), for two different voice samples are given below. It can be observed that, though the voice samples belong to the same person, the feature vectors are different. Likewise, features are extracted from each user to generate the speaker models.

Feature Extracting from user 072

Features extracted from User 072 , voice sample 01

```
[[-1.08473051e+00 -8.87010687e-01  1.17950717e+00 ...
1.86208260e-01
   1.39594685e-01  3.60891069e-01]
 [-1.72484770e+00 -4.45250192e-01  1.69560822e+00 ...
2.74221765e-01
   1.74904629e-01  2.78464898e-01]
 [-1.60881666e+00 -4.29232610e-01  1.78943363e+00 ...
2.57876136e-01
   4.11797132e-02  1.12039290e-01]
 ...
 [-8.32619336e-01 -2.30661680e+00  2.04500045e+00 ...
1.95720768e-01
   4.45279731e-01 -5.84833663e-02]
 [-1.07911132e-01 -2.32307094e+00  6.92530293e-01 ...
6.92134950e-02
   4.06231270e-01 -1.90424079e-02]
 [-9.23990734e-02 -2.49015905e+00  4.08807475e-01 ...
1.33227736e-03
   2.85323009e-01  1.40716418e-02]]
Audio Vector Dimentions = (1072, 40)
```

Features extracted from User 072 , voice sample 02

```
[[-0.91633931 -0.36801379  1.80053463 ...  0.12736839 -
0.15154584
  -0.02867407]
 [-1.10911008 -0.36083061  1.54401225 ...  0.24435469 -
0.11862512
   0.05856831]
 [-1.18877635 -0.51696897  1.50874369 ...  0.3215714  -
0.06405392
   0.15608178]
 ...
 [-1.66639872 -1.13008818  1.13017184 ... -0.08003002 -
0.02436209
   0.07757806]
 [-1.66820986 -1.09658256  1.0898229  ... -0.03746659
0.02948035
   0.03988005]
 [-1.75488735 -1.0583746   1.11303413 ... -0.04779263 -0.01686
  -0.04561303]]
Audio Vector Dimentions = (1815, 40)
```

Features extracted from User 072 , voice sample 01

```
[[-0.77909757 -0.11890273  0.93211055 ...  0.14128113
0.24762661
   0.33352062]
 [-1.09026261 -0.21745074  2.08816572 ...  0.24072504
0.31733811
   0.39140305]
 [-1.2134993  -0.26538125  2.13027188 ...  0.31906241
0.34551556
   0.36484232]
 ...
 [-1.20389469 -0.41142687  0.53794005 ... -0.03409937 -
0.16293943
   0.01487046]
 [-1.18168471 -0.45579137  0.35610547 ... -0.02890975 -
0.09830564
   0.06861528]
 [-1.52245384 -0.40473205  0.49193793 ...  0.03164303 -
0.00393248
   0.08822047]]
Audio Vector Dimentions = (1258, 40)
```

Features extracted from User 074 , voice sample 02

```
[[-1.007489    0.04332664  1.87975104 ...  0.02962123
0.04874434
   0.07557722]
 [-1.71467837 -0.26643401  1.96923403 ... -0.02731891 -
0.07020341
   0.0991265 ]
 [-1.73603527 -0.43160963  1.66229767 ...  0.04682993 -
0.06754905
   0.03254996]
 ...
 [-1.83684292 -1.15717272  1.43802138 ...  0.03642377 -
0.25044341
  -0.24270543]
 [-2.03739383 -1.1080042   1.29817893 ...  0.09805845 -
0.33830926
  -0.37270045]
 [-2.4344688  -1.29509311  0.9809771  ...  0.18041749 -
0.15339082
  -0.2818969 ]]
Audio Vector Dimentions = (1100, 40)
```

Speaker Models

With the extracted features, separate GMM speaker models are trained for each speaker. Below mentioned are two models generated for speakers 072 ,074 and 075. .

GMM model for Speaker 072 (10 training audio samples)

```
[[-1.08473051 -0.88701069  1.17950717 ...  0.18620826
0.13959469
   0.36089107]
 [-1.7248477  -0.44525019  1.69560822 ...  0.27422176
0.17490463
   0.2784649 ]
 [-1.60881666 -0.42923261  1.78943363 ...  0.25787614
0.04117971
   0.11203929]
 ...
 [-1.94236605 -1.44842444  1.26942716 ... -0.05186842 -
0.04784436
   0.16498338]
 [-1.93941732 -1.50344721  1.46782347 ...  0.11620192 -
0.03556226
   0.10122362]
 [-2.07935752 -1.68852505  1.26845275 ...  0.14797687
0.00916301
   0.08151565]]
*** GMM Model generated for Speaker 072.gmm With Dimensions =
(14024, 40)
```

GMM model for Speaker 121 (10 training audio samples)

`[[-7.79097567e-01 -1.18902734e-01  9.32110551e-01 ... 1.41281130e-01`
`   2.47626610e-01  3.33520618e-01]`
` [-1.09026261e+00 -2.17450738e-01  2.08816572e+00 ... 2.40725041e-01`
`   3.17338108e-01  3.91403053e-01]`
` [-1.21349930e+00 -2.65381254e-01  2.13027188e+00 ... 3.19062406e-01`
`   3.45515561e-01  3.64842318e-01]`
` ...`
` [-1.77375466e+00 -1.85641218e+00  3.39240292e-01 ... 9.93119132e-02`
`  -1.86865903e-01 -1.14457577e-01]`
` [-2.01817837e+00 -1.66289231e+00  6.49101830e-01 ... 2.33716054e-01`
`  -4.83442049e-02 -1.58454754e-01]`
` [-2.18399997e+00 -1.68909413e+00  8.39430241e-01 ... 1.77441416e-01`
`  -1.47067792e-03 -1.22742113e-01]]`
`*** Speaker Model Generated for 074.gmm   With Dimensions = (9848, 40)`

GMM model for Speaker 148 (10 training audio samples)

`[[-1.03600502 -0.62002154  1.3678865  ...  0.16045773 0.19827585`
`   0.31633485]`
` [-1.35698839 -0.12854387  1.96608604 ...  0.14440207 0.27389171`
`   0.26052887]`
` [-1.32172428 -0.11045739  1.97063161 ...  0.22256635 0.30560258`
`   0.1511345 ]`
` ...`
` [-1.97313594 -1.17685684  1.44051878 ...  0.05589005 -0.09460351`
`  -0.15463278]`
` [-1.98695507 -1.11142413  1.62840925 ...  0.06222338 -0.1470232`
`  -0.06906416]`
` [-2.11494271 -1.1247889   1.73784062 ...  0.12342813 -0.05449403`
`   0.07258851]]`
`*** Speaker Model Generated for 075.gmm   With Dimensions = (9940, 40)`

In evaluating the application, voice models were generated using different approaches by changing the no. of voice samples considered and the duration of the voice samples that are used. Different portions of the data set were used in this, and audio were individual to each model group. The Table 4.1 shows how the models were designed.

Table 4.1 Generated GMM Models in the evaluation

Model	No. of Speakers considered	Voice Signal Length used for building Model	No. of Voice Samples Used for each model	No. of Individual voice models generated
Model 1	60	4s	5	60
Model 2	60	4s	10	60
Model 3	60	10s	5	60
Model 4	60	10s	10	60

For testing the implementation, 4 sets of training data were generated with the following specifications as shown in Table 4.2.

Table 4.2. Training Sets

Testing Set	No. of Testing Samples Considered	Duration (d) of training voice samples
A	18	2s
B	18	4s
C	18	10s
D	18	15s

The testing voice samples are collected from the following speakers identified as follows in Table 4.3, and contains different voice samples of the users speaking

Table 4.3 Speakers of the training sets

072	092	123
076	098	130
078	104	140
081	110	143
085	119	148
089	121	150

The Table 4.4 shows the results of testing user 072 voice samples against the generated 60 speaker models

Table 4.4 User 072 Results

Audio File (User 072)		**Model (Duration of the training voice/ no of voice samples)**			
		Model1	**Model 2**	**Model3**	**Model4**
Sample	**Length**	**4s /5**	**4s/10**	**10s /5**	**10s /10**
A	2s	N	N	N	Y
B	4s	N	Y	N	Y
C	10s	N	Y	Y	Y
D	15s	N	Y	Y	Y

The Table 4.5 shows the results obtained by testing user 121 voice samples against the generated 60 speaker models

Table 4.5 User 121 Results

Audio File (User 121)		**Model (Duration of the training voice/ no of voice samples)**			
		Model1	**Model 2**	**Model3**	**Model4**
Sample	**Length**	**4s /5**	**4s/10**	**10s /5**	**10s /10**
A	2s	N	Y	N	N
B	4s	Y	Y	N	Y
C	10s	N	Y	Y	Y
D	15s	N	Y	Y	Y

The Table 4.6 shows the results obtained by testing user160 voice samples against the generated 60 speaker models

Table 4.6 User 148 results

Audio File (User 148)		**Model (Duration of the training voice/ no of voice samples)**			
		Model1	**Model 2**	**Model3**	**Model4**
Sample	**Length**	**4s /5**	**4s/10**	**10s /5**	**10s /10**
A	2s	N	N	N	N
B	4s	Y	N	N	N
C	10s	Y	Y	Y	Y
D	15s	N	Y	Y	Y

The following Table 4.7 shows how the test voice samples of training set D were identified by the model generated with 10 training voice samples with duration 10s

Table 4.7 Test Speaker Identification on Model 4 (10/10s) with Training Set D

Test Sample	Speaker	Correctly Identified	Incorrectly Identified
1	072	Y	
2	076	Y	
3	078	Y	
4	081	Y	
5	085		Y
6	089	Y	
7	092	Y	
8	098	Y	
9	104	Y	
10	110	Y	
11	119	Y	
12	121	Y	
13	123	Y	
14	130	Y	
15	140	Y	
16	143	Y	
17	148	Y	
18	150	Y	

The following Table 4.8 shows how the test voice samples of training set B were identified by the model generated with 10 training voice samples with duration 4s

Table 4.8 Test Speaker Identification on Model 2 (10/4s) with Training Set B

Test Sample	Speaker	Correctly Identified	Incorrectly Identified
1	072	Y	
2	076	Y	
3	078		Y
4	081	Y	
5	085	Y	
6	089	Y	
7	092	Y	
8	098	Y	
9	104	Y	
10	110	Y	
11	119	Y	
12	121	Y	
13	123	Y	
14	130	Y	
15	140	Y	
16	143	Y	
17	148		Y
18	150	Y	

Overall accuracy in identifying an unknown speaker is obtained as follows against the user models generated with different training voice samples.

$$\textit{Accuracy} = \frac{\textit{No.of Correctly Identified Test Samples}}{\textit{Total No.of Test Samples}}$$

Table 4.9. Overall Results

Audio File			**Model (Duration of the training voice/ no of voice samples)**			
No. of total training models	**Testing Sample**	**No. of test samples tested**	**4s /5 Model1**	**4s/10 Model 2**	**10s /5 Model 3**	**10s /10 Model 4**
			Accuracy	**Accuracy**	**Accuracy**	**Accuracy**
60	A	18	27%	22%	16%	11%
60	B	18	55%	83%	61%	66%
60	C	18	11%	77%	83%	88%
60	D	18	16%	88%	88%	94%

It is observed in Table 4.9 that, different combinations of training models and testing data generates different results. The accuracy is higher when the no. of voice samples considered for the model generation is high. The accuracy overall is decreased with the incorrectly identified data. Also, it can be observed that if the training voice samples and the testing voice samples have the similar specifications the accuracy is increased.

In the implemented MFCC and GMM based approach and the similarity score defined in the decision module, identifies the user by analyzing the log probability and then identifies the speaker by the highest probability. But, with the two- step verification implemented, the user's identity is double checked and imposters getting access is avoided up to a considerable extent.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

This chapter conclude the document by presenting the overall progress, views and with the extensions that are to be done as future work. The main objective of this study was to implement a voice authentication approach to be used in online examinations based on Sinhala Language.

With the sudden outbreak of the COVID-19 pandemic and closure of educational institutes the necessity of online education facilities has risen, and examinations play a key role in every education sector. With adapting to online examinations, the major obstacle that comes across is user authentication. Though there are various approaches used in user authentication, there are many challenges of implementing them developing country like Sri Lanka. The inaccessibility of continuous and strong network/internet facilities that allows continuous video streaming for video-based authentication, and other required instruments is the main challenge that is faced in this type of environment.

As a solution to this above problem, a system with the capability of implementing voice authentication is presented with the study. There were four main phases carried out, acquisition data and generating required datasets, feature extraction, speaker modeling and evaluation. After analyzing the data attainment and considering the factor that the system implemented should function efficiently in real time, a suitable voice authentication approach was developed.

For improving the security an optional feature that adds more security is implemented along with voice authentication where a username and password that is previously shared with the user is required in accessing the system. In voice authentication MFCC is used in feature extraction and GMM in speaker modeling using the extracted features.

Different types of speaker models are generated by variating the number of voice samples considered and the duration of the testing and training voice signals. We can observe that higher accuracy is obtained with the models generated when the number of voice signals used in creating one model and the audio signal duration is increasing. Based on above, this system can be used to facilitate the online education platforms as expected in the objectives and as well as in any other online platform where security should be intensified.

5.2 Future Work

We can expect that the online education will become one of major sector in education system in the society. Considering that we can suggest few extensions to expand the study and provide a solution that is more secure and efficient.

The system can be supported with a strong voice activity detection approach that removes noise and truncate silence where in real time data could be unclean to maintain the robustness of the system.

In increasing the security, the users' answers can also be captured as voice and analyzed in two ways: to authenticate the user continuously and identify the answer by speech recognition. This will be more in maintaining the expected security levels and avoiding any possibilities of cheating on real time examinations. Also, voice authentication can be combined with image authentication using a face recognition approach if the required facilities can be acquired by the users.

APPENDIX A

Survey On Online Examinations

The COVID-19 pandemic changed learning in many unprecedented ways. While online education systems are widely being adopted by a number of institutions these days to continue with their academic work, online examinations also have come to forth. One major limitation in the online examinations is the difficulty of providing true user authentication.

This questionnaire focus on the identifying the currently used user authentication methods in Sri Lankan higher education institutes.

1. Specify your Education Status *

Mark only one oval.

- ( ) Post Graduate (Masters/ Post Graduate Diploma)
- ( ) Undergraduate
- ( ) Higher Diploma/ Diploma
- ( ) Other: ______________________

2. What is the type of the institute you are attending *

Mark only one oval.

- ( ) Government Institute
- ( ) Semi Government Institute
- ( ) Private
- ( ) Other: ______________________

3. Does your institute conduct online classes? *

Mark only one oval.

- Yes
- No

4. Does your Institute conduct online examinations? *

Mark only one oval.

- Yes
- No

5. If "Yes", what is the authentication method used to verify the student's identity at the online examination?

Mark only one oval.

- Using specific User Name / Password provided by the institute
- Image Capturing/ Video Streaming throughout the exam
- Using Voice Recognition
- Face Recognition
- Other: ____________________

6. How would you describe the existing mode of conducting examinations in your institute?

Mark only one oval.

- Exams are administered online and can be attempted by students online remotely with remote invigilation
- Exams are administered online and can be attempted by students online remotely without any remote invigilation

7. If " Voice Recognition" is used for user authentication in an online examination, Do you posses the reliable hardware required ? i.e. Headphone with Mic / Headset

Mark only one oval.

- Yes
- No

8. Any other comments on existing mode of conducting examinations in your institute?

____

APPENDIX B

Speakers Identity as included in The Dataset

(How speakers are named)

72	95	95
73	96	96
74	97	97
75	98	98
76	103	103
77	104	104
78	105	105
79	106	106
80	108	108
81	109	109
82	110	110
83	111	111
84	112	112
85	113	113
86	114	114
87	115	115
88	116	116
89	117	117
90	118	118
92	119	119

REFERENCES

Allen, I. E. & Seaman, J., 2008. *Staying the Course.* United States of America: sloanconsortium.org.

Kumar, P. M., 2016. A New Human Voice Recognition System. *Asian Journal of Science and Applied Technology,* Volume 5, pp. 23-30.

Maloney-Krichmar, D. & Abras, C., 2003. History of emergence of online communities.

Vuppala, A. & Rao , K., 2014. Speech Processing in Mobile Environments. *SpringerBriefs in Electrical and Computer Engineering.*

Abdul-Hassan, A. K. & Hadi, I. H., 2019. CENTRAL INTELLIGENT BIOMETRIC AUTHENTICATION BASED ON VOICE RECOGNITION AND FUZZY LOGIC. *Journal of Theoretical and Applied Information Technology,* Volume 97.

Ahmed, I., Mohamed, . M. A. & Noma, A. M., 2018. A Framework For Secure Online Exam Using Biometric Fingerprint And Steganography Techniques. *International Journal of Engineering & Technology.*

Awojide, S., Awe, O. & Babatope, T., 2018. Biometric Fingerprint System Using an Online Based PAttern Recgonition for Candidates Authentication in Nigeria INstitution Examinations. *International Journal of Science & Engineering* .

Azeta, A., Inam, A.I , A. & Daramola, O., 2018. A Voice-Based E-Examination Framework For Visually Impaired Students In Open And Distance Learning. *Turkish Online Journal of Distance Education.*

Bhimani, N. V., 2014. Speaker Recognition System Based On MFCC. *International Journal of Engineering Research & Technology,* 3(2).

Carey, K., 2020. *s everybody ready for the big migration to online college? Actually, no.,* s.l.: The New York Times.

Chelali, F. Z. & Djeradi, A., 2017. Speaker identification system using LPC -Application on Berber language. *International Journal of Speech Technology,* 20(3).

Chenafa, M., Istrate, D., Vrabie, V. & Her, M., 2008. Biometric System Based on Voice Recognition Using Multiclassifiers. *European Workshop on Biometrics and Identity Management.*

Chen, Y. S.-H., 2009. “Speaker verification using MFCC and support vector machine. *International Multimedia conference of Engineers and Computer Scientisits* , Volume 1.

Dehak, N., Kenny, P., Dehak, R. & Dumouche, P., 2011. Front-end factor analysis for speaker verification. *IEEE Trans. on Audio, Speech, and Language Processing,* pp. 788-798.

E.Chandra, E., Manikandan, K. & Kalaivani, M., 2014. A Study on Speaker Recognition System and Pattern classification Techniques. *INTERNATIONAL JOURNAL OF*

INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING, 2(2).

Fayyoumi, A. & Zarrad, A., 2014. Novel Solution Based on Face Recognition to Address Identity Theft and Cheating in Online Examination Systems. *Advances in Internet of Thing,* Volume 4, pp. 5-12.

Giorgi, R. et al., 2019. An Iris+Voice Recognition System for a Smart Doorbell. *Mediterranean Conference on Embedded Computing (MECO).*

Godfrey , A. M. & Nichie, A., 2013. OICE RECOGNITION USING ARTIFICIAL NEURAL NETWORKS AND GAUSSIAN MIXTURE MODELS. *International Journal of Engineering Science and Technology (IJEST).*

Gunawardena, C. & McIsaac, M., 1996. Distance Education. In: *Jonassen, D. (Ed.) Handbook for Research on Educational Communications and Technology.* New York: s.n., pp. 403-437.

Hayashi, R., Garcia, M., Maddawin, A. & Hewagamage, K. P., 2020. *Online Learning in Sri Lanka's Higher Education Institutions during the COVID-19 Pandemic,* s.l.: Asian Development Bank.

Hayashi, R., Garcia, M., Maddawin, A. & Hewagamag, K. P., 2020. *Online Learning in Sri Lanka's Higher Education Institutions during the COVID-19 Pandemic,* s.l.: Asian Development Bank.

Herrera-Camacho, A. et al., 2019. Design and Testing of a Corpus for Forensic Speaker Recognition Using MFCC, GMM and MLE. *Proceedings of the 2019 International Conference on Video, Signal and Image Processing.*

Ibrahima, N. S. & Ramli, D. A., 2018. I-vector Extraction for Speaker Recognition Based on Dimensionality Reduction. *InProcedia Computer Science* , Volume 126.

Inman, M., Danforth, D., Hangai, S. & Sato, K., 1998. Speaker identification using hidden Markov models. *International Conference on Signal Processing* .

Jung, B.-H. (. F., Umair , M. & Meng, Z., 2020. Active voice authentication. *Digital Signal Processing.*

Khairi, T. W. A., 2019. Secure Mobile Learning Sytem Using Voice Authentication. *Journal of Enginnering and Applied Sicences* , Volume 14, pp. 8180-8186.

Kim , C. & Stern, R., 2010. Feature extraction for robust speech recognition based on maximizing the sharpness of the power distribution and on power flooring. *IEEE International Conference on Acoustics, Speech, and Signal Processing.*

King, C. G. & Guyette, R. W., 2009. Online Exams and Cheating: An Empirical Analysis of Business Students' Views. *Journal of Educators Online.*

Kumari , T. J. & Jayanna, H., 2018. i-Vector-Based Speaker Verification on Limited Data Using Fusion Techniques. *De Gruyter.*

Kydyrbekova, A. M. ,. O. O. M. &. P. D., 2020. Identification and authentication of user voice using DNN features and i-vector. *Cogent Engineering,* Volume 7.

Mahola, U., Nelwamondo, F. . V. & Marwala , T., n.d. HMM Speaker Identification Using Linear and Non-linear Merging Techniques.

Mauryaa, A., Kumara, D. & Agarwalb, R., 2018. Speaker Recognition for Hindi Speech Signal using MFCC-GMM Approach. *Procedia Computer Science,* Volume 125, pp. 880-887.

Mohd. Shah, H. . N., Kamis, . Z. & Abdollah, M. F., 2014. Biometric Voice Recognition in Security System. *Indian Journal of Science and Technolog.*

Nijhawan, G., 2014. Speaker Recognition using Support Vector Machine. *International Journal of Computer Applications,* Volume 32.

Purnama, F., Sukhbaata, . O., Choimaa, . L. & Usagawa , T., 2020. Implementation of real-time online mouse tracking on overseas quiz session. *Education and Information Technologies.*

Rahme, F. a., E.Ahmed, T., A.Saeed, R. & Alhumyanic, H., 2021. Analysis and challenges of robust E-exams performance under COVID-19. *Results in Physics,* Volume 23.

Ramu, T. & Arivoli, T., 2013. A Framework Of Secure Biometric Based Online Exam Authentication: An Alternative To Traditional Exam. *International Journal of Scientific & Engineering Research,* 4(II).

Reedy, A., Pfitzner, D., Rook, L. & Ellis , L., 2021. Responding to the COVID-19 emergency: student and academic staff perceptions of academic integrity in the transition to online exams at three Australian universities. *International Journal for Educational Integrity,* Volume 17.

Renolds, D., 1994. Experimental evaluation of features for robust speaker identification. *Speech and Audio Processing IEEE Transactions,* Volume 3.

Rudrapal , D., Das, S. & Kar, N., 2012. Voice Recognition and Authentication as a Proficient Biometric Tool and its Application in Online Exam for P.H People. *International Journal of Computer Applications,* Volume 39.

Sadkhan, S. B., Al-Shukur, B. K. & Mattar, A. K., 2017. Biometric voice authentication auto-evaluation system. *Annual Conference on New Trends in Information & Communications Technology Applications (NTICT).*

Salameh, N. & Shukur, Z., 2015. Review of user authentication methods in online examination. *Asian Journal of Information Technology* .

Sanjaya, W., Anggraeni, D. & Santika, I. . P., 2018. Speech Recognition using Linear Predictive Coding (LPC) and Adaptive Neuro-Fuzzy (ANFIS) to Control 5 DoF Arm Robot. *Journal of Physics.*

Sharma, V. & Bansal, P. K., 2018. A Review On Speaker Recognition Approaches And Challenges. *International Journal of Engineering Research & Technology* , 2(5).

Shdaifat, A., Obeidallah, R., Ghazal, G. & Sarhan, A. A., 2020. A proposed Iris Recognition Model for Authentication in Mobile Exams. *International Journal of Emerging Technologies in Learning (iJET).*

Shoup, A., Talkar, . T., Jodie, . C. & Jain, A., n.d. An Overview and Analysis of Voice Authentication Methods.

Sithara, A., Thomas, A. & Mathew, D., 2018. Study of MFCC and IHC Feature Extraction Methods With Probabilistic Acoustic Models for Speaker Biometric Applications. *Procedia Computer Science,* Volume 143.

Sukmandhani, A. A. & Sutedja, I., 2019. Face Recognition Method for Online Exams. *International Conference on Information Management and Technology (ICIMTech).*

Sunitha, C. & Chandra, E. H., 2015. Speaker Recognition using MFCC and Improved Weighted Vector Quantization Algorithm. *International Journal of Engineering and Technology ,* Volume 7, pp. 1685-1692.

Taft, S., Perkowski, T. & Martin, L., 2018. A Framework for Evaluating Class Size in Online Education. The Quarterly Review of Distance Education. *The Quarterly Review of Distance Education,* pp. 181-197.

Tanwar, R., Singh, . K. & Malhotra, S., 2019. An Approach to Ensure Security using Voice Authentication System. *International Journal of Recent Technology and Engineering (IJRTE).*

Ullah, A., Xiao, H., Barker, T. & Lilley, M., 2012. Using Challenge Questions for Student Authentication in Online Examination. *International Journal for Infonomics,* Volume 5.

Wu, L. et al., 2019. LVID: A Multimodal Biometrics Authentication System on Smartphones. *IEEE Transactions on Information Forensics and Security ,* Volume 15.

Zhang, X., Xiong, Q., Dai, Y. & Xu, X., 2018. Voice Biometric Identity Authentication System Based on Android Smart Phone. *International Conference on Computer and Communications (ICCC).*