LAUTECH Journal of Computing and Informatics (LAUJCI) – ISSN: 2714-4194
Volume 2 Issue 1, May 2021 – www.laujci.lautech.edu.ng

Development of a Microcontroller Based Voice Recognition System for Accessing Bank Vault

Okomba, N. S.,
Department of Computer Engineering, Federal University Oye-Ekiti,
nnamdi.okomba@fuoye.edu.ng

Esan, A.,
Department of Computer Engineering, Federal University Oye-Ekiti
adebimpe.esan@fuoye.edu.ng

Omodunbi, B. A.,
Department of Computer Engineering, Federal University Oye-Ekiti
bolaji.omodunbi@fuoye.edu.ng

Adeyanju, I. A.,
Department of Computer Engineering, Federal University Oye-Ekiti,
ibrahim.adeyanju@fuoye.edu.ng

Olaniyan, O. M.
Department of Computer Engineering, Federal University Oye-Ekiti.
olatayo.olaniyan@fuoye.edu.ng

ABSTRACT
Security is a very essential aspect in the banking industry, and bank vaults access is limited to a certain number of individuals for proper management and accountability. Many security systems have been used to ensure the safety of customer's assets ranging from password-based system, finger print based system, and combinational locks in the bank vaults. These forms of protection can be hacked, and duplicated. The use of combinational lock doesn't favour the physically challenged ones (handicap) and people with cognitive disability. The developed work aims towards overcoming the challenges in security of the current vault systems in the banking industry, by using voice recognition system approach with a relative spectra frequency coefficient- perceptual linear predictive (RASTA-PLP) as the applied feature extraction technique in order to aid access to the bank vault to authorized persons with a convenient and more secure means without any need of remembering a password, lock combination, and the use of magnetic card of which can be easily stolen. The functionality of this work was implemented and tested using a model bank vault system built and controlled by Arduino ATMEGA328p microcontroller and a voice recognition system designed on MATLAB R2015a which controls the access into the bank vault system and thereby protecting the vault from robbery. Vector Quantization (VQ) approach using LBG algorithm (VQLBG) was applied for feature matching. VQ distortion between the resultant codebook and RASTA-PLP coefficients of an unknown speaker was taken as the basis for determining the speaker's authenticity. Euclidean distance was calculated for each signal to be recognized and matched to the closest speech signal in the speech database.

Keywords: Voice recognition, Speech recognition, Feature extraction, Microcontroller

1. INTRODUCTION

Bank vault is a very important part of banking industry that provides security for storage of treasures, valuables, records, money, and documents. It protects the contents from unauthorized users, theft, fire, and natural disaster. Irrespective of the level of security applied in bank vault system, there are still shortcomings in the existing password, and combinational lock-based bank vault systems. In area of verification of identities and manual method of input which exhibit sluggish and error prone characteristics, Voice recognition will create room for more flexible and time conscious system.

Voice recognition is a signal processing technique for identifying, understanding and converting voice signals into text or commands. Voice recognition comprises all those activities which attempt to link a speech sample to its speaker through its acoustic or perceptual properties (Ramachandran *et al* 2002). The speaker voice characteristics such as expression in spoken words, emotion of the speaker and slowness or loudness are represented by acoustic wave speech signal. The human voice is unique personal trait. Since some amount of variations occurs in the speech samples obtained from the same speaker. This is due to the fact that a speaker cannot exactly imitate the same utterance again and again; even the signature of an individual also shows variation from trails to trials (Mathur *et al* 2013*)*. Voice recognition consists of two different categories such as speaker recognition and speech recognition (Paul *et al*, 2013).

Speech recognition is a signal processing technique that is required to recognize and understand spoken words, by activating a device that digitizes sound waves and matching its pattern against already stored patterns. Speech is the most natural form of human communication and has made it possible for computer to follow human voice commands and understand human languages (Deepak and Vikas, 2013). Speaker recognition helps in the basic purpose of speaker identification which forms a difficult domain (Chaudhary *et al* 2015). Examples may include, security systems users having to speak a PIN (Personal Identification Number) number or to speak their credit card number to verify their identity. By checking the voice uniqueness of the input utterance with the help of Speaker recognition, the system can add an extra level of security. Speech feature extraction which attempts to obtain a parametric representation of an input speech signal plays a crucial role in the overall performance of a Speech Recognition system (Hyunsin *et al*, 2009).

Speech feature extraction techniques are signal processing procedure which convert the speech waveform into some useful parametric representation called features (Davis and Mermelstein, 1980).

Perceptual Linear Prediction (PLP) is a feature extraction technique originally proposed by H. Hermansky as a way of warping spectra to minimize the differences between speakers while preserving the important speech information (Hermansky, 1990) while RASTA is a technique that applies a band-pass filter to the energy in each frequency sub band in order to smooth over short-term noise variations and to remove any constant offset resulting from static spectral coloration in the speech channel (Hermansky, 1994). Relative Spectral Transform-Perceptual Linear Prediction (RASTA-PLP) feature extraction technique is a combined RASTA and PLP technique which performs better in noisy environment, and has the ability to encode low volume voice samples with low number of bits.

2. RELATED WORKS

Bhargavi *et al* (2015), worked on a voice control wheel chair project to help the physically challenged and old people who face many problems in life regarding to mobility that will enable them not to be dependent relatives. The invention of the wheel chair is of a great help to the elderly, though it still needs improvement. Voice controlled wheel chair was made using HM2007 voice recognition kit.

Haitao, and Xiaojun (2009) designed a Hidden Markov Model (HMM) speech recognition system which is based on Field Programmable Gate Array (FPGA). The developed system improved the conventional Viterbi algorithm and recognition speed has been increased. The core part of the hardware is EP2S60F1020C3 FPGA chip. The experimental result of this system shows that the speech recognition accuracy reaches 94% when ten numbers are being recognized, and the average recognition time is 0.669s.

Godfrey *et al,* (2019) worked on Voice activated Home System for the Movement Impaired Aged Persons; the developed system is integrated as a single portable unit that allows the user to wirelessly control lights, fans, electronic doors, electronic windows and robotic wheelchair. The system is portable and constructed in a way that it is easy to install, configure, run, and maintain. Here, a voice controlled smart home system is designed, so that the users can perform their domestic tasks by just the use of their voices to speak their command words.

Apsana.S, and Renjitha G Nair (2016) worked on a Voice Controlled Wheel Chair Using Arduino. The project is to help the physically challenged people and the elderly people who need the assistance of others to help them in moving from one spot to another. The project makes use of Arduino Uno board and other components. Voice Recognition Kit (HM2007 Module) is being used to recognize the voice command. The voice command given is converted to binary numbers by Voice Recognition Kit and those binary data is given to the Arduino board for the control of the wheel chair.

Muhammed (2008) worked on Robot Control with Voice Command. Speech commands are taken by a microphone and features of the commands were extracted with MFCC algorithm. The commands are recognized using Neural Network. The recognized command converted to the form in which the robot can recognize. The final form of the commands is sent to the robot and the robot move accordingly. The set of command the robot was trained to recognize are; Move Forward, Move Backward, Turn Right, Turn Left, Stop and Continue.

Akshay *et al* (2016) worked on Voice Controlled Home Automation. With the rapid increase in the number of elderly ones in the society and the handicapped, there is great need for speech control systems. This project is about controlling home appliances with the aid of speech. The automation recognizes voice commands given by the user and transfers it to a microcontroller which detects the voice command and proceeds with the switching accordingly. Raspberry Pi microcontroller module & Bluetooth module HC05 was used to implement the vision. The home automation system is intended to control all lights and other electrical appliances in a home or office using voice commands.

Oyebola B. O. (2015) worked on the development of a microcontroller-based security system for a bank vault, home and offices facilities. The system uses a keypad to send signal to the control unit in order to access a lock when correct pin combinations are presented and triggers an alarm system when wrong pin combinations are entered. The system is designed to allow authorized users make changes of the pin combination at will. The system components and connections were tested for continuity in order to detect any fault and the circuit worked well as designed. This device allows a user access the door to a bank vault with a password, reducing the stress and force applied while using a manual technique.

The manual and password-based approach for locking bank vault have challenges that involves cloning the keys and having the password hacked by multiple trials which creates a limitation in existing systems. Due to the biometric characteristics of the human voice and the application of pattern recognition, Voice recognition for bank vault system will go a long way in enhancing the security system of the banking industry and reduce the rigidity and time constrain in the manual, password, and combinational lock approach of banking security system.

3. PROPOSED METHOD

This project is on the development of a voice recognition system to gain access into a bank vault. This project was implemented with the aid of MATLAB R2015a software to produce the voice recognition system which comprises of voice acquisition stage, pre-processing, recognition stage and the hardware development, using Arduino ATMEGA328P microcontroller which receives signal from the developed voice recognition system. Figure 1 illustrates the block diagram of the developed system

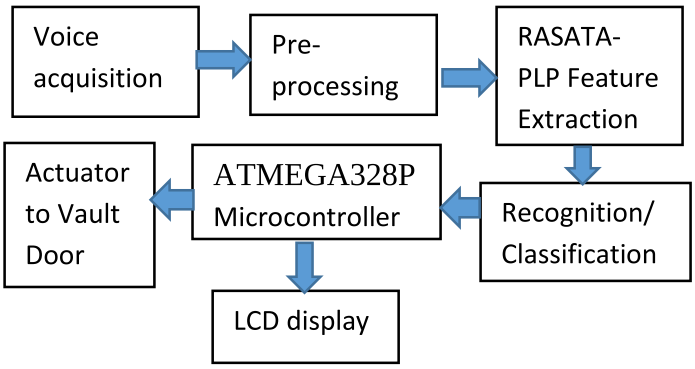

Figure 1: Voice recognition system block diagram

3.1. Voice acquisition and Pre-processing

A microphone was used to capture ten voice samples mentioning the words open and close for the purpose of vault control, and convert the acoustic sound wave into an analog signal. This analog signal is conditioned with filtering to compensate for any channel impairments. The filtering limits the bandwidth of the signal to approximately the Nyquist rate which is half the sampling rate before sampling. The conditioned analog signal is then sampled to form a digital signal by an analog-to-digital (A/D) converter. The realized digital signal is further pre-processed by pre-emphasic filtering using matlab in order to boost the signals high frequency components. The equation for pre-emphasis is shown in Equation (1):

$$H(z) = 1 - az^{-1}$$

3.2. Feature Extraction

The Pre-processed voice samples has features that characterize them, which were extracted using RASTA -PLP feature extraction technique and stored in the database as SOUD ID for recognition and classification purposes. The RASTA filtering serves as a modulation-frequency band pass filter, that emphasizes the modulation frequency range most relevant to speech while discarding lower or higher modulation frequencies, and PLP incorporates critical-band spectral-resolution into its spectrum estimate by remapping the frequency axis to the Bark scale and integrating the energy in the critical bands to produce a critical-band spectrum approximation. figure 2 illustrates the RASTA-PLP model diagram.

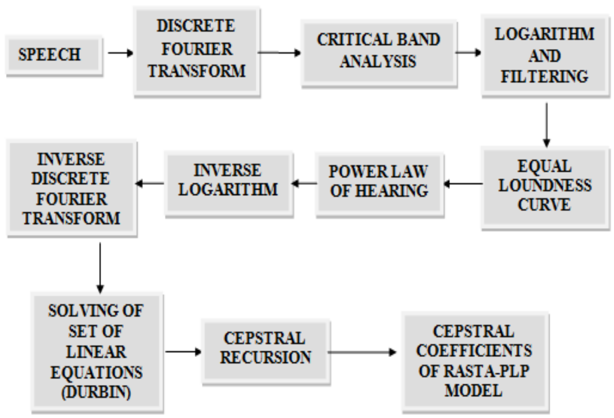

Figure 2: RASTA-PLP Model diagram

3.3 Software Design

The voice system database was programmed using structural programming method whereby each section (function) leads to another in the whole speech program in MATLAB R2015a software due to its high-level dynamic nature. The MATLAB code controlling the voice recognition system, allows user interaction through the GUI rather than thousands of codes. This GUI has some press-buttons menu that represent the key steps taken in achieving high efficiency and optimization of the system such as Add a new speech from microphone, Speaker recognition from microphone, RASTA-PLP representation, Database information, Delete database, and Exit.

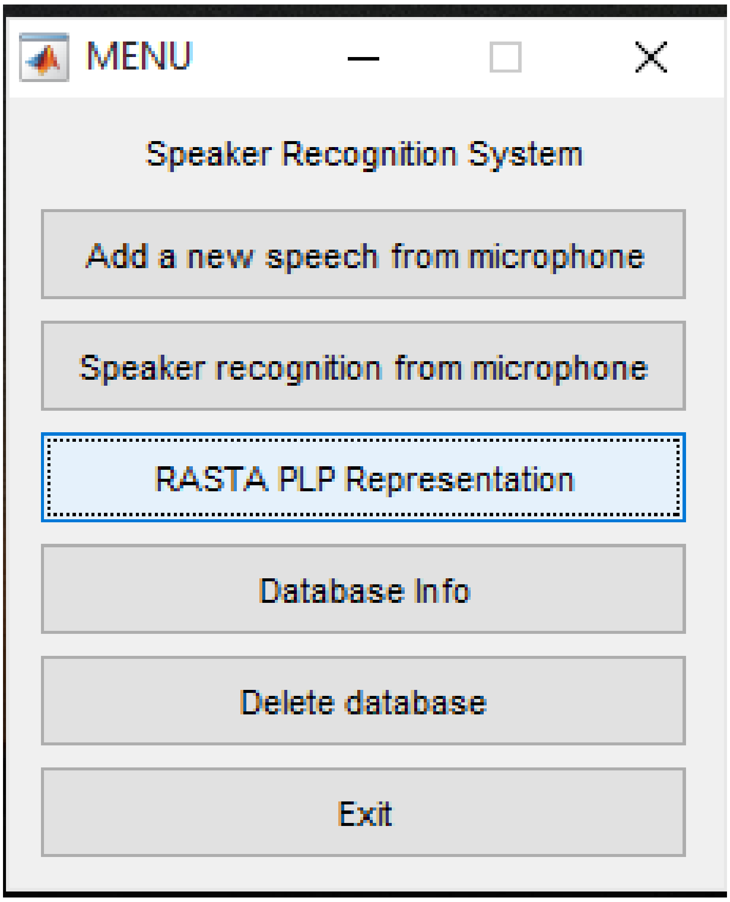

Figure 3: The GUI for the Speaker Recognition System

3.3.1 Add a New Speech from Microphone

This is the first press-button on the pop-up menu of the speaker recognition system where individual user speech signals or data in the form of speech is added to the database of the system. This section represents the training phase of the speaker recognition system where the system is trained with different user speech signals and then stored in the speech sample database that will be used for reference model in the recognition phase of the system. In this stage, during the acquisition of the speech signal, some parameters are to be inputted for good acquisition of speech signal, these parameter are; a class number (sound ID) that will be used to differentiate different speech in database and also will be used for recognition; the sampling frequency; bits per sample; duration of recording which is in seconds. There are some recommended values for these parameters listed above, sampling frequency (22050Hz recommended), sampling bits (8 recommended), and duration of recording in seconds (2 seconds recommended). After recording, the user(s) speech is added to the system database.

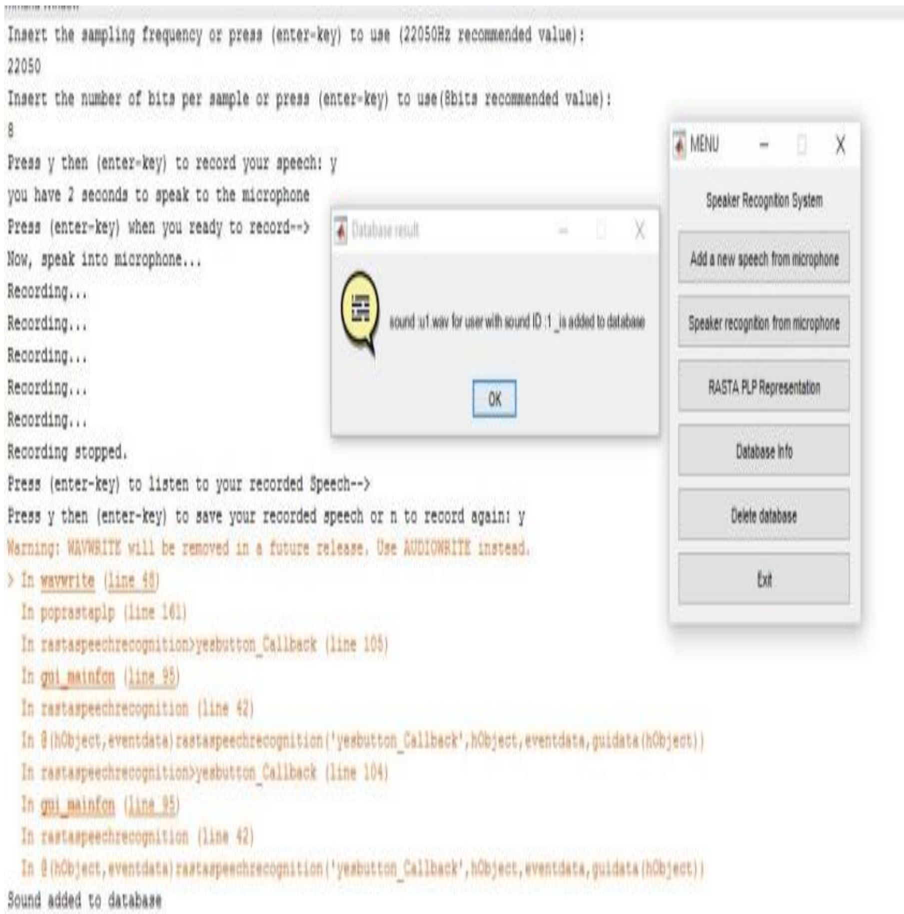

Figure 3.4: Command Prompt and Pop-up Message Box display for the Speech Acquisition

3.3.2 Speaker Recognition from Microphone

This is the second press-button of the pop-up menu which is the important section of the whole speaker recognition system. This section controls the access into the bank vault after the user's Sound ID is recognized; it recognizes the sound ID with open and close. During the recognition phase of the speaker recognition system, speech signals recorded into the system are processed and compared to see if there is a match with the speech signal in the system database. The mathematical tools and algorithm used while processing the user(s) speech signal are; Relative Spectral Analysis with Perceptual Linear Prediction (RASTA-PLP) algorithm for feature extraction and VQLBG (vector quantization using LBG algorithm) for easy recognition of user speech signals. In recognition, VQLBG was used instead of HMM, ANN, Gaussian method as it has faster processing and dynamic to changes in the use of different speech signals. VQLBG makes erasing and re-addition of speech signals very easy and faster unlike HMM or ANN which needs most training and is not easy to erase data, they are not as dynamic in change of data during erasing process as the system must be configured to be able to erase speech signals and also reprogram new speech signals countless number of times.

\

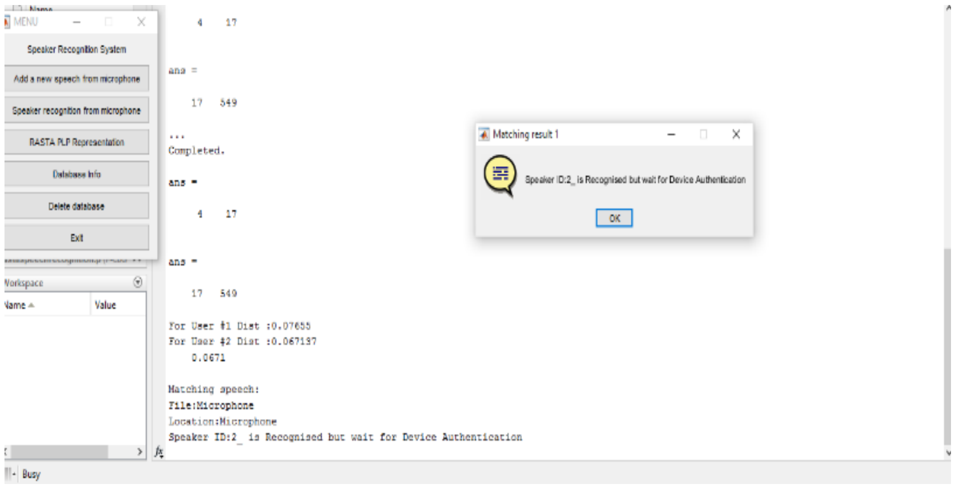

Figure 3.5: Command Prompt and Pop-up Message Boxes display for the Speaker ID using RSFC - PLP computation and VQLBG method

3.3.2 Database Information

This fourth press-button of the pop-up GUI menu of the speech recognition system which is used to display the details of the user(s) speech stored in the database of the system during the acquisition phase. Such that when any Speaker ID is supplied, this section plays the sound wav file pertaining the ID number and if there is no speech signal in the database, it prompts up a message saying “Database is empty”. Else it displays the content of the Sound ID in the database.

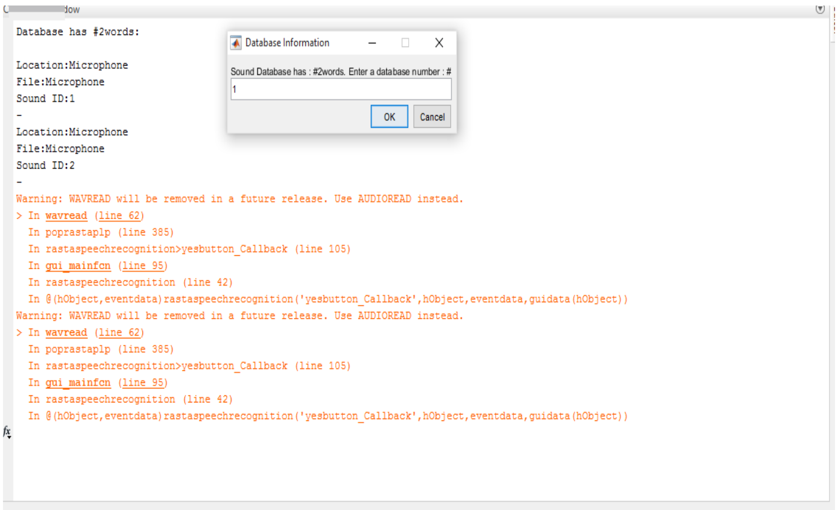

Figure 3.7: A description of database information

3.3.2 Delete Database

This fifth press-button on the popup menu of the speaker recognition system is used to erase/delete the entire content of the database. This is included to address the need of capturing different set and countless number of speech signals also to remove sets of speech sample on the users will. On deletion of speech signals from the database, it prompts up a message saying “Database was successfully removed from the current directory.” When deletion is attempted on an empty database, it also prompts a warning message “Database is empty”

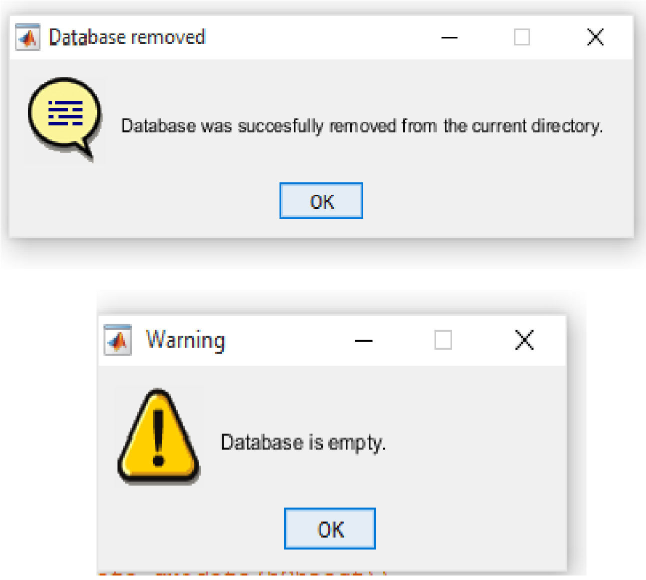

Figure 3.8: Status Display of the database deletion process

3.3.2 Exit

This is the final section of the speech system database which is used to close all operation of the speaker recognition system GUI

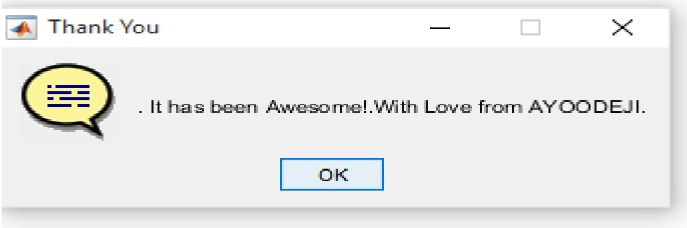

Figure 3.9: Exit Operation Display

3.4 Hardware Design

The microcontroller ATMEGA328 on the Arduino Uno development board receives the signal from the MATLAB R2015a software through the RS232 cable and sends a signal to the bank vault prototype to grant access into the bank vault. Arduino Uno board has some on-board electronic components such as: A 20MHz crystal oscillator which provides speed for the microcontroller and is connected to two 22uF capacitor connected in parallel. A 10k variable resistor which is used to provide the contrast for the LCD. A 22K pull-up resistor which is used for the microcontroller, it has a port with open circuit, they are connected internally, and they are used to provide the missing filter when the port is in a high state. A casing was fabricated for the Liquid Crystal Display (LCD), Universal Serial Bus Port (USB).

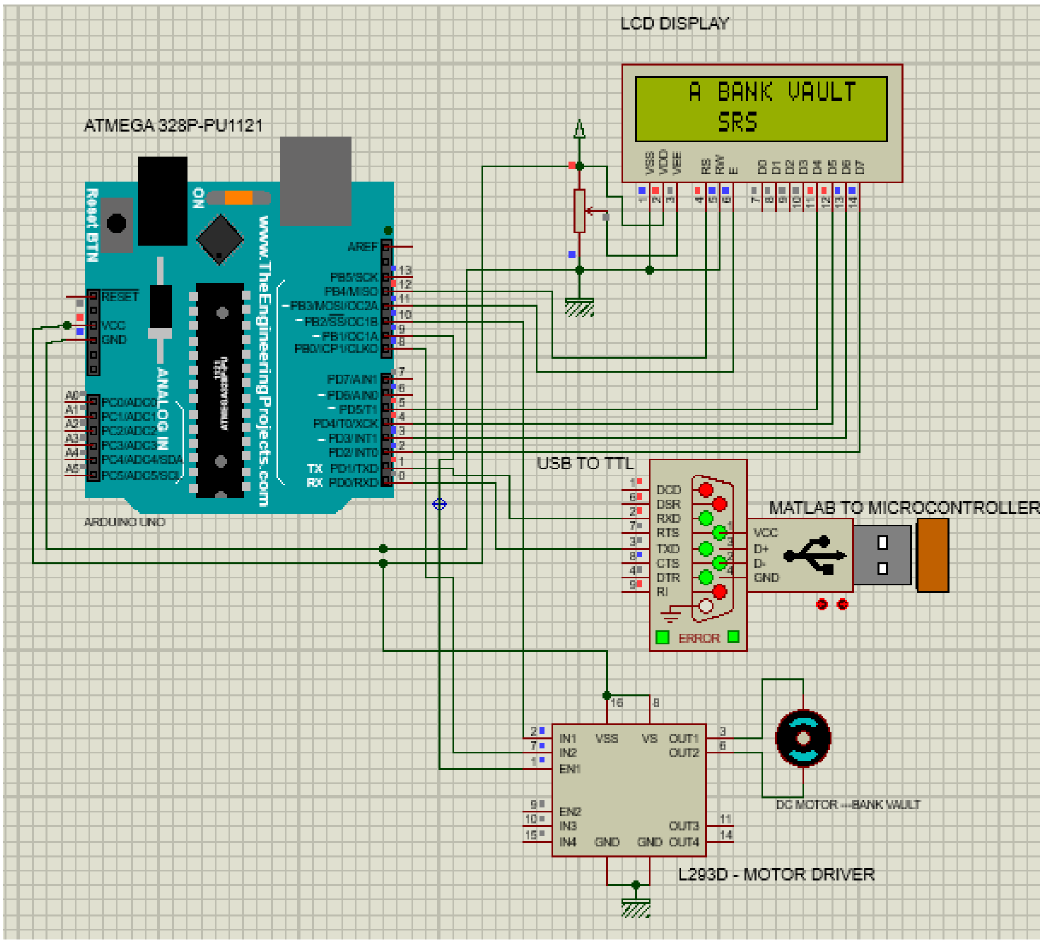

Figure 3.10: Schematic diagram of the circuit hardware for access into a bank vault

From the schematic diagram of the system control circuit shown above. A 5V DC is supplied into the circuit by a Computer System 1 through a Universal Serial Bus cable interfaced with a serial port connector. The voltage in the circuit passed through 2200uF capacitor (a condenser microphone), which work by removing the lumps/ripples present in the voltage. The connections of pins between the different materials are explained as follows;

Pins 12,13,5,4,3, and 2 of the ATMEGA328 is connected to the RS, E, D4, D5, D6, and D7 respectively for the displaying status and operation of the speech recognition system. The VSS, VEE and R/W are grounded. Vo of the LCD is connected to 10k ohms variable resistor for the contrast of the LCD. Pins 0 and 1 of the Arduino Uno is used as UART receiver pin (RX) and Transmitter pin (TX) pin which serve as USART receiver pins which in which an RS232 is interfaced with a Universal Serial Bus module through which the control of the hardware operation was implemented. Pins 10, 9, 8 are connected to L293D motor driver IC as enable, input1 and input2 for controlling the bidirectional movement of the DC motor.

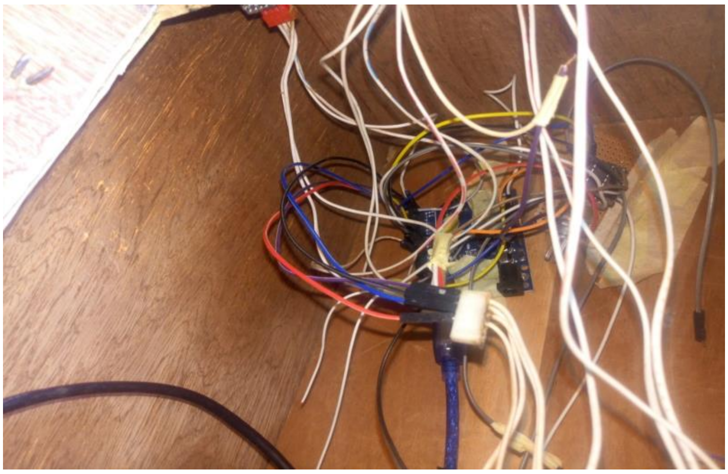

Figure 3.11: Diagram of the Internal Arrangement of the Prototype/Model Bank Vault circuit

Figure 3.12: The Prototype/Model Bank Vault to be controlled by the Voice Recognition

4.0 RESULTS AND DISCUSSIONS

The table1 below summarize the result and evaluate the system based on the Euclidean distance of the individual speech signal to be recognized. The, parametric representation of the distances of the speech signal is shown as calculated when the RASTA-PLP and VQ codebook training function are applied. The speaker having the shortest feature distance is matched to a user stored in the database of the speech system, and thus, is the recognized user. In the table, the red sections simply indicate the recognized users.

TABLE 1: Euclidean Distance of Speakers to Speech Samples [OPEN AND CLOSE]

Speech Sample: [OPEN]				
User in Database	**Distance User-1**	**Distance User-2**	**Distance User-3**	**Distance User-4**
Voice 1	**0.28121**	**0.29036**	**0.30107**	**0.30516**
Voice 2	**0.25997**	**0.28216**	**0.32551**	**0.33107**
Voice 3	**0.31004**	**0.3236**	**0.33601**	**0.35148**
Speech Sample: [CLOSE]				
User in Database	**Distance User-1**	**Distance User-2**	**Distance User-3**	**Distance User-4**
Voice 1	**0.291993**	**0.27754**	**0.28589**	**0.25906**
Voice 2	**0.30569**	**0.28911**	**0.32242**	**0.28076**
Voice 3	**0.33461**	**0.30088**	**0.60194**	**0.30036**

5.0 CONCLUSION

The developed system is based on recognizing an unknown speaker from a given set of registered speakers in the system database. This project was accomplished using MATLAB R2015a software in creating the speaker recognition system. In the first step of generating the speaker recognition model, speech data acquisition and pre-processing was done, we then went for feature extraction using Relative Spectral Analysis with Perceptual Linear Prediction. These features act as a basis for further development of the speaker recognition process. Next on recognition phase, we went for feature mapping using the vector Quantization using LBG algorithm. The results obtained using RASTA-PLP and VQ are appreciable. The RASTA-PLP coefficients and vectors for each speaker were computed and quantized respectively for efficient representation. The code books were generated using LBG algorithm which optimizes the quantization process. VQ distortion between the resultant codebook and RASTA-PLP coefficients of an unknown speaker was taken as the basis for determining the speaker's authenticity. Euclidean distance was calculated for each signal to be recognized and matched to the closest speech signal in the speech database. Accuracy of 96% was obtained using VQLBG algorithm and Euclidean distance.

Finally, the functionality of the project was implemented and tested using a model bank vault system built, controlled by Arduino ATMEGA328p microcontroller which receives access signal from the speaker recognition system design on MATLAB R2015a and drives the Bipolar Transistor (NPN) which controls the access into the bank vault system.

5.1 Recommendations

For future works or improvements related to this project, the following suggestions can be applied;

1. A system can be developed which the speech samples to be stored in the database will be represented with the user's exact name for easy identification.
2. Also, the system can be developed to be able to contain wide range of users and still retain its efficiency and accuracy in recognition and feature mapping.
3. The system can also include a voice output which states the operation of the system.
4. The whole project can be made standalone by implementing it on a very high-speed microcontroller interfaced with some necessary component which would cost more to achieve.
5. Along with this system, the additional use of Face-detection can be implemented to enhance the security of the system.

REFERENCES

Akshay Mewada, Ayush Mishra, Manoj Gupta, Rahul Dash, Nilofer Mulla (2016), "a research paper Voice Controlled Home Automation" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 3, March 2016 ISSN: 2277 128X.

Apsana.S, and Renjitha G Nair (2016). "Voice Controlled Wheelchair using Arduino"; International Advanced Research Journal in Science, Engineering and Technology; National Conference on Emerging Trends in Engineering and Technology (NCETET'16) Lourdes Matha College of Science & Technology, Thiruvananthapuram Vol. 3, Special Issue 3, August 2016 Copyright to IARJSET DOI 10.17148/IARJSET 332. ISSN (Online) 2393-8021; ISSN (Print) 2394-1588.

Oyebola B. O. (2015) "Microcontroller Based Keypad Operated Door System with Alarm, A Technical Approach to Improve Commercial Bank Vault Security" International Journal of Research (IJR) e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015 Available at http://internationaljournalofresearch.org, Available online:www.edupediapublications.org/journals P a g e | 635.

Bhargavi .Y, Hariprasad Reddy, Ravi TejaCh.V (2015). Paper named as "*Voice based wheel chair for physically challenged".*

Chaudhary, P. J., & Vagadia, K. M. (2015). A Review Article on Speaker Recognition with Feature Extraction. International Journal of Emerging Technology and Advanced Engineering, IJETAE, 5(2).

Davis S.B., and Mermelstein P. (1980): "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences." IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP **28**: 357–366.

Deepak, M. Vikas (2013), "Speech Recognition using FIR Wiener Filter", International Journal of Application or Innovation in Engineering & management (IJAIEM), pp.204-20,2013.

Godfrey N. Okorafor, Felix K. Opara, Nkwachukwu Chukwuchekwa, and Chigozie C. Ononiwu, (2019); "Voice activated Home System for the Movement Impaired Aged Persons"; European Journal of Engineering Research and Science (EJERS) Vol. 4, No. 11, November 2019

H. Hermansky (1990), ―Perceptual linear predictive (PLP) analysis for speech‖, J. Acoustic Soc. Am., pp. 1738-1752, 1990.

H. Hermansky and N. Morgan (1994), ―RASTA Processing of Speech‖, IEEE Trans. On Speech and Audio Processing, Vol. 2, 578-589, Oct. 1994.

Mathur, S., Choudhary, S. K., & Vyas, J. M. (2013). Speaker Recognition System and its Forensic Implications. Open Access Scientific Reports, 2(4).

Haitao Zhou, And Xiaojun Han (2009), "Design and Implementation of Speech Recognition System Based on Field Programmable Gate Array"; Journal Of Modern Applied Science; Vol. 3, No. 8

Hyunsin Park, Tetsuya Takiguchi, and Yasuo Ariki (2009), Research Article Integrated Phoneme SubspaceMethod for Speech Feature Extraction, Hindawi Publishing Corporation EURASIP Journal on Audio, Speech, and Music Processing Volume 2009, Article ID 690451, 6 pages doi:10.1155/2009/690451

Muhammed Zahit ÖZDEM_RCAN (2008), "Robot Control with Voice Command" Yildiz Technical University Faculty of Electric and Electronics Department of Computer Engineering.

Paul, Teenu Therese, and Shiju George (2013) "Voice recognition based secure android model for inputting smear test result." *International Journal of Engineering Sciences & Emerging Technologies, ISSN*: 2231-6604,2013

Ramachandran, R. P., Farrell, K. R., Ramachandran, R., & Mammone, R. J. (2002). Speaker recognition—general classifier approaches and data fusion methods. Pattern Recognition, 35(12), 2801-2821.