INTELLIGENT CHATBOT FOR LAB SECURITY AND AUTOMATION

V Akshatha Prasad [1], R Ranjith [2]
Department of Electrical and Electronics Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India
vallampati246@gmail.com[1], r_ranjith@cb.amrita.edu[2]

***Abstract*—Artificial intelligence based communicative artefacts are called chatbots. The purpose of a chatterbot or chatbot is to render an interaction between a human and a robot in the form of speech or text. They offer the best services in a variety of areas, such as education, healthcare, transportation, etc. As per the research, nearly 85% of product offerings will be automated by 2020. The proposed system is a voice-based chatbot that helps to improve the security and automation of a lab. The system makes use of Automatic Speaker Recognition (ASR) algorithm in order to recognize a person and allow him/her inside the lab. This allows only authorized person to access the lab facilities. The same algorithm is used for generating the list of available components in the lab based on the person's keyword input thus automating the lab components dispatch.**

***Keywords*—chatbot, automatic speaker recognition system, MFCC, vector quantization, speech to text**

I. INTRODUCTION

A chatbot is a software program [12] which communicates orally or textually with humans. These systems are designed to model a conversational partner like an individual. Chatbots are typically used in dialog systems [13] for a range of practical purposes, including customer support or knowledge retrieval. Many chatbots use advanced natural language processing technologies, but simpler ones search keywords within the content and then take the response with the most appropriate keywords or the most similar text sequence from a repository. There are two categories [14] of bots, rule-based chatbot and a smarter machine learning based chatbot. In the rule-based approach the programmer will code the rules for the system. In the machine learning approach, massive amount of steaming data is needed for the algorithm to train on its own. Developers therefore need to better define the parameters of machine learning. Google, IBM, Amazon have come with their voice assistants enabling users to interact with their device by speaking to them, but there arises lot of security risks, such as, anyone can communicate with them through voice. It doesn't verify who the user it is speaking to. Therefore, it is necessary to build such a system that authenticates the user first and then performs the activities by following voice commands of the user.

The approach used in this paper is rule-based which is simple and efficient. Fig. 1 illustrates the overview of system implementation which is structured and follows a logical flow of information, where voice of the user is taken as the input to the system. Voice signal, a continuous analog signal is converted to digital signal for performing various operations on the signal. On the converted digital voice signal, an appropriate speaker recognition algorithm is applied to identify the speaker.

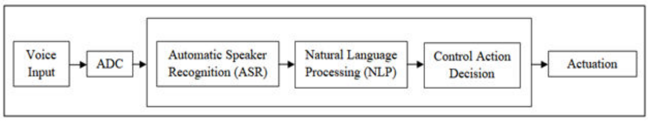


Fig. 1. System Overview

Only when the speaker is recognized, the bot will further interact with the user. Based on the keywords uttered by the speaker such as "FAN 1 ON", "FAN 1 OFF", "PROJECTOR ON", and "NAVIGATE TO SLIDE 2" etc, necessary action will be taken to automate the lab.

This paper is organized into six sections, where the remaining sections provides the previous work that has been done in this area, methodology adopted to develop the proposed system, design approach of simulation, results obtained and future scope.

II. PREVIOUS WORK

Alan Turing and Joseph Weizenbaum [1] invented machines behaving like people and in 1950; it had the potential of creating a test to see if an individual could differentiate a human from a machine: the Turing Test. However, this software is used to fit keywords and a small context that lacks sufficient information to keep the conversation going. The present technology is capable of assembling and gathering massive data that generates knowledge that can almost act to mimic human beings. The new innovative method of collecting data includes an artificial intelligence capable of interacting with applications such as face detection, speech sentimental interpretation, text mining and natural language processing [2], which could include features that add new meaning to marketing and sales.

Shabina et al. [3] developed the android based chatbot for the hotel booking. In this process, the user enters the specifics of the reservation that the chatbot asks. The booking will be rendered once the details have been furnished. Fernando A. Mikic Fonte et al. [4] have proposed a chatbot in which the android app functions in the learning process as a student's assistant. Students can ask for information on their courses, exam questions and answers, and the model will give answers to their request. Tarun Lalwani et al. [5] have developed a chatbot incorporating artificial intelligence (AI) and natural language processing (NLP) algorithms for college purpose that helps students to find details easily. It has an easy UI that answers to the questions relevant to the test cell, registration, educational, user participation and grade point average, placement cell and others. Imran Ahmed and Shikha Singh have structured a web-based chatbot using python and AIML language [6]. It consists of NLP and speech recognition techniques that outputs the text information in terms of voice.

Many researchers have developed chatbots that allows any user to interact with the system, but have not taken into consideration the concern of security. It is therefore important to create a system that first authenticates the user and then executes the tasks by following the user's voice commands. In this paper, MFCC technique is used to authenticate the user and Google's Speech to text API is used to perform user's required actions.

III. PROPOSED METHODOLOGY

One way of securing the lab is to allow only a certain group of people to enter the premises by means of voice identification as a biometric at the entrance. In this paper, Automatic Speaker Recognition (ASR) is used for voice authentication by the chatbot. It consists of two phases such as, Registration and Recognition. During the registration process, voice recordings of different speakers are taken; the attributes are extracted and added to the database. In the recognition stage, when a text-independent phrase is spoken by the individual, the features are extracted and updated to the database to classify the speaker. Based on the identity of the speaker, the appropriate control action shall be taken. The system design is illustrated using a flow chart in the below figure Fig. 2. The system consists of five stages,

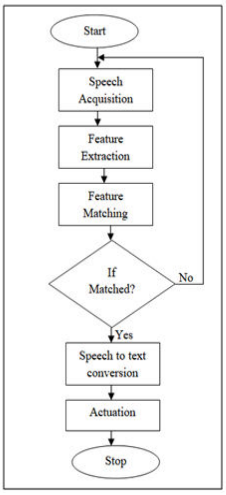


Fig. 2. System Flow

A. Speech Acquisition

A person can enter the lab after the voice verification. For that, the user must speak to the system so that the audio signal is recorded. All speech signals are analog in nature. In order to perform signal processing operations on the signal, analog signal to digital signal conversion is mandatory. So the input to the system is the sampled signal.

B. Feature Extraction

Every individual has a specific way of speaking because each person can be characterized by the different sizes of his or her larynx, vocal tract and other sections of an auditory system for the production of sound. Each speaker has a particular accent, intonation style, pacing, speech pattern, vocabulary, and much more. Feature extraction is a method for extracting the feature vectors from the speech signal by the speaker-dependent information and vector quantizing them to obtain the specific codebook for each speaker. The feature that is extracted is Mel Frequency Cepstral Coefficient (MFCC) [7].

In MFCC, filters are linearly distributed for low-frequency signals and logarithmic for high-frequency signals. This is used to identify the phonetic characteristic of voice. The approach is focused on the human being's listening behavior. All signals in this procedure are shown in the MEL scale. Linear frequency spacing is given for a signal less than 1000 Hz, while logarithmic spacing is provided for a signal greater than 1000 Hz. Anatomical experiments have shown that human perception of frequency identification does not work effectively on a linear scale. Vector quantization (VQ) maps the vectors inside the space from a broad vector space to a finite number of regions [7]. Each region is called a cluster and may be represented using its centroid. All cluster centroids are being combined to form a common speaker codebook.

As part of registration stage, voice signals are recorded from the user and are stored in the database. MFCC features are extracted, mapped into its equivalent VQ code vectors and are stored in the train database as shown in Fig. 3. After all the information are stored in the database, a message pops out saying that, "Sound added to database".

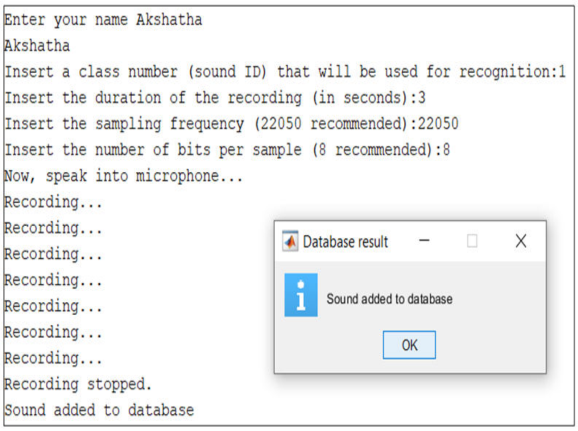


Fig. 3. Adding student 1 sound to trained folder

C. Feature Matching

In feature matching, the overall VQ variance between these MFCCs and the codebooks contained in the stored collection are tested while comparing the input voice vector of an anonymous speaker. The interval from a vector to the closest codeword of a codebook is called the VQ distortion. VQ distortion is like the Euclidean distance between the two classifiers [8] and is shown by the equation (1).

$$d(x, y) = \sqrt{\sum_{i=1}^{dim}(xi - yi)^2} \qquad (1)$$

On the grounds of this distortion of the VQ [9], whether the user is a real individual or an impersonator is determined [10]. i.e. if the distortion of the VQ is less than the threshold value, then the speaker is a real individual else, the speaker is concluded as the imposter.

Fig. 4 shows the identification phase, where the user is identified and allowed inside the lab if he/she is authenticated. If the person is not verified, then the message pops out saying “No matching user” as shown in Fig. 5 and is not allowed inside the lab.

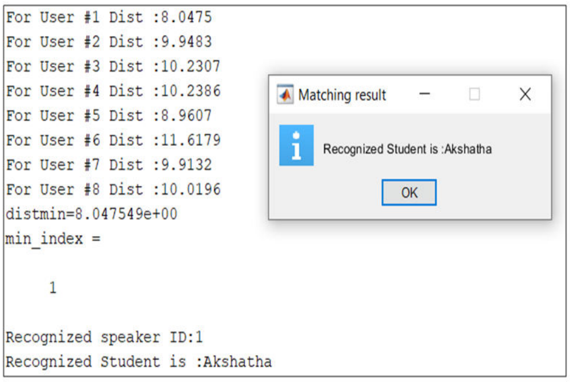


Fig. 4. Calculation of distances and matching student 1

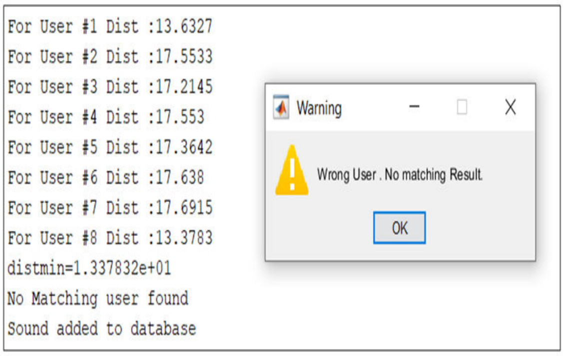


Fig.5. Wrong user Result

D. Speech to Text Conversion

After the user is authenticated as a real individual, the person is allowed inside the lab. If the person wants to switch ON the lights or fans he/she can simply chat with the LAB-BOT. The LAB-BOT uses Google’s cloud service, Speech to text API to translate whatever the person is uttering and convert them into text as shown in Fig. 6, and take the necessary actuation.

Transcript	Confidence
"turn to slide number two"	0.79229

Fig. 6. Google’s Speech to Text transcript

E. Actuation

Inside the lab whatever actuation that the student or laboratory assistant wants they can perform by just speaking to the BOT. BOT is trained to automate the lab such as finding the components placement, switching ON/OFF the lights, fans, projector, go to the particular slide in the PPT etc;.

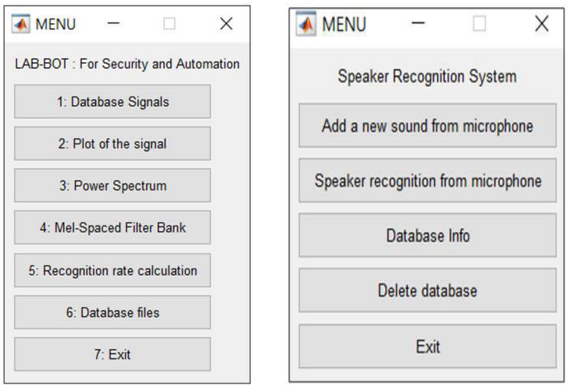
IV. SOFTWARE DESIGN APPROACH

The entire system is implemented using MATLAB software because of its simplicity in implementation. Using an in-built audio record and playback tools, train-data is collected from the users, MFCC features are extracted and VQ code words are stored in the database (in train_database.mat file). Later while testing, the MFCC features of voice during the test are collected and are compared with that of the VQ code words present in trained database. Euclidean distances between the trained and tested code words are calculated. Among various distances calculated, the one with minimum distance whose value is less than the threshold is the matched user. If the distance is greater than the threshold, then the user is treated as invalid user.

After the user is treated as valid user, he is allowed to enter inside the lab and whatever the voice commands the person speaks out, those will be captured by the system and is converted to text using Google’s Speech to Text API.

V. RESULTS

This section consists of the results obtained for developing the system in MATLAB.



(a) Main User Menu (b) Speaker Recognition Menu Options

Fig. 7. User Menu Options

Fig. 7(a) shows the main menu where the number of signals present in the test and train database can be seen along with their plots in time domain and power spectrum. Also recognition rate is calculated i.e. efficiency with which the system can detect the registered users. From the main menu options, if option 6 is selected then the second menu option Fig. 7(b) is popped up where the user can add, test and delete the signals from the database.

Fig. 8, Fig. 9 and Fig. 10 show the plot of the signal, power spectrum and Mel filter bank.

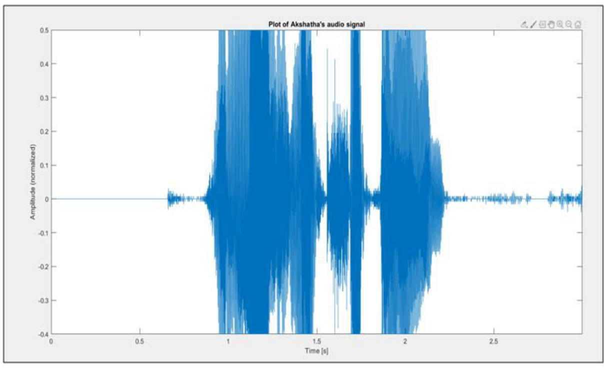


Fig. 8. Speech signal plot of student 1

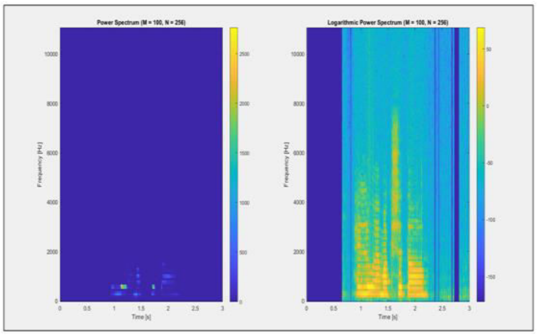


Fig. 9. Power Spectrum Plot of student 1

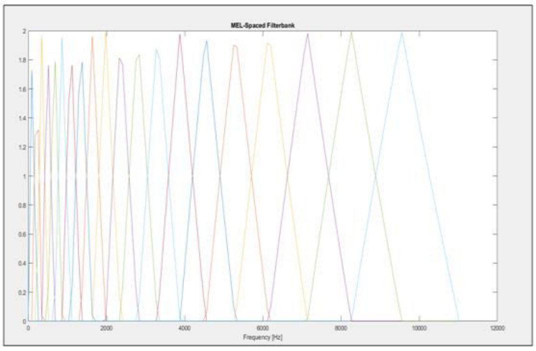


Fig. 10. Plot of MEL Spaced Filter

As it can be seen from Table I, out of 8 sounds, the system is detecting 6 sounds correctly and works with efficiency more than 71.42% as shown in Fig. 11.

TABLE I. RESULTS OF SPEAKER RECOGNITION

S.No	Speaker	Gender	Recognition Result
1	Speaker 1	Female	Success
2	Speaker 2	Male	Success
3	Speaker 3	Female	Failure
4	Speaker 4	Female	Success
5	Speaker 5	Male	Failure
6	Speaker 6	Male	Success
7	Speaker 7	Female	Success
8	Speaker 8	Female	Success

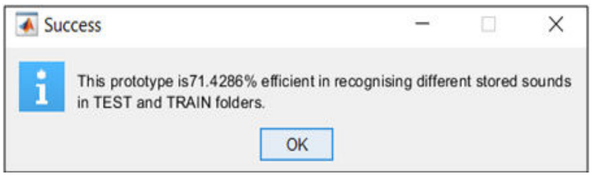


Fig. 11. Recognition Rate for testing 8 speakers

VI. CONCLUSIONS AND FUTURE SCOPE

As the prevalence of virtual assistants such as Amazon Alexa and Google Assistant is growing, voice-based chatbots have become a valuable feature in improving the user experience of specific users. Undoubtedly, speech recognition technology developed by basic and custom software is improving and will continue to improve in the coming years. Recent study forecasts that 80% of companies are planning to introduce chatbots by 2020 [11]. Also by 2021, 85% of customer interactions will be handled without human agents. Therefore there is a need of a reliable system that allows the users to interact and get responses very accurately and precisely. The proposed system in this paper is a rule-based voice-based structured chatbot that is developed to enhance security and automation to the lab which provides the recognition rate of greater than 71.42%.

REFERENCES

[1] Joseph Weizenbaum, "ELIZA A Computer Program For the Study of Natural Language Communication Between Man And Machine", Communications of the ACM, Massachusetts Institute of Technology, Cambridge, Volume 9, Number 1, pp. 36-45, January 1966.

[2] Darius Zumstein and Sophie Hundertmark, "Chatbots – An Interactive Technology for Personalized Communication, Transactions and Services", IADIS International Journal on WWW/Internet, Volume 15, Number 1, pp. 96-109, November 2017.

[3] Shabina Sayed, Rushabh Jain, Burhanuddin Lokhandwala, Fakhruddin Barodawala and Murtuza Rajkotwala, "Android based Chat-Bot", International Journal of Computer Applications, Volume 137, Number 10, pp. 28-32, March 2016.

[4] Fernando A. Mikic Fonte, Martin Llamas Nistal, Juan C. Burguillo Rial, and Manuel Caeiro Rodr´ıguez, "NLAST: A natural language assistant for students", IEEE Global Engineering Education Conference (EDUCON, pp. 709-713), April 2016.

[5] Tarun Lalwani, Shashank Bhalotia, Ashish Pal, Shreya Bisen, Vasundhara Rathod, "Implementation of a Chatbot System using AI and NLP", International Journal of Innovative Research in Computer Science & Technology (IJIRCST), Volume 6, Number 3, pp. 26-30, May 2018.

[6] Imran Ahmed and Shikha Singh, "AIML Based Voice Enabled Artificial Intelligent Chatterbot", International Journal of u- and e-Service, Science and Technology (IJUNESST), Volume 8, Number 2, pp. 375-384, March 2015.

[7] Shreya Narang, Ms. Divya Gupta, "Speech Feature Extraction Techniques: A Review", International Journal of Computer Science and Mobile Computing (IJCSMC), Volume 4, Issue 3, pp. 107-114, March 2015.

[8] Akanksha Singh Thakur, Namrata Sahayam, "Speech Recognition Using Euclidean Distance", International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 3, pp. 587-590, March 2013.

[9] M. Akumar, Dr. Soman K. P., "Amrita_CEN at SemEval-2016 : Complex Word identification using word embeddings", in International Workshop on Semantic Evaluation (SemEval 2016), 2016, 2016.

[10] R. Vinayakumar, K. P. Soman and P. Poornachandran, "Long short-term memory based operation log anomaly detection," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 236-242, 2017.

[11] [Online] Available at: https://www.invespcro.com/blog/chatbots-customer-service/

[12] [Online] Available at: https://expertsystem.com/chatbot/

[13] [Online] Available at: https://en.wikipedia.org/wiki/Chatbot

[14] [Online] Available at: https://marutitech.com/heres-need-know-chatbots/