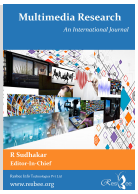
Vol.5 No.3 July 2024

Multimedia Research

Received 10 April, Revised 25 June, Accepted 9 July



Language-Independent Speaker Identification System-based Attendance System

Deepak Chouhan
Dept. of Computer Science & Engineering MIT Art, Design and Technology University, Loni Kalbhor Maharashtra Pune, India

T Aryan Raju
Dept. of Computer Science & Engineering MIT Art, Design and Technology University, Loni Kalbhor Maharashtra Pune, India

Vaidehi Pradeep Gade
Dept. of Computer Science & Engineering MIT Art, Design and Technology University, Loni Kalbhor Maharashtra Pune, India

Arva Bhargava
Dept. of Computer Science & Engineering MIT Art, Design and Technology University, Loni Kalbhor Maharashtra Pune, India

Rajesh S. Prasad
Dept. of Computer Science & Engineering MIT Art, Design and Technology University, Loni Kalbhor Maharashtra Pune, India

Abstract: The most convenient way of communication since ancient times is talking i.e., speaking to each other. Whenever we speak to someone, we convey information in the form of words or voice or to be prudent say speech as this project is related to it. Recognizing the speaker can simplify the task of authenticating or verifying the identity of a speaker as part of a security process as it enables access control of various services by voice. The development of language independent speaker Identification system for student attendance system is beneficial in many ways. It helps the lecturer in administrating the attendance of their student with efficiency. This is because students sometimes cheat with their attendance by signing on behalf of their friends who did not attend class. The proposed system intends to improve the process of taking attendance by enabling the lecturer to efficiently record attendance and prevent students from marking attendance for their friends.

Keywords: Attendance System, Classification, Deep Learning, Language Independent, Speaker Identification.

1. Introduction

Taking attendance during a lecture is very time-consuming, also, many students cheat on their attendance by signing on behalf of their friends who did not attend class [8]. The development of a speaker identification system for student attendance will help the lecturer in administering attendance of their student efficiently & preventing cheating among students as each voice is unique [9]. Speaker identification is a powerful [10], [11], non-invasive, inexpensive bio-metric technique fueled by rapid advances in algorithms [13] and computer technology; security systems based on speaker identification are on the verge of commercial success [12].

2. Literature Review

2.1 Text-Independent Speaker Identification in Multilingual Environments

This paper is based on the benefits of adding short-term intonation and energy information to MFCC features in order to optimize language-independent speaker recognition systems. Although these short-term intonations improve the language robustness in speaker recognition systems, the final results are not satisfactory. This happens due to the great inter-session variability of the prosodic features causing slight performance loss when training and testing even in the same language.

2.2 Language Independent Speaker Identification

This paper talks about the important factors and their uses in language-independent speaker identification. It talks about how features like MFCC and mean pitch information can boost speaker

https://doi.org/10.46253/j.mr.v5i3.a3

identification. Although pitch information is not a very vital factor it helps in downsizing the MFCC output thus giving a more accurate and precise result. The accuracy lies between 81%-87% proving to be very helpful.

2.3 Language and Text-Independent Speaker Identification System using GMM

In this paper there was an experiment with 120 speakers and the features used to identify them individually were MFCC and pitch-based DMFCC. The combination of these two features proved to be the most effective while building a language-independent speaker identification system. Individually MFCC gives an error rate of 5.8% while DMFCC gives an error rate of 2.9%. When both of them are combined the error rate turns out to be only 1.2%.

2.4 Language and Text Independent Speaker Recognition System using Artificial Neural Networks and Fuzzy Logic

This Paper discusses the various application fields of language and text-independent speaker recognition systems like biometrics and forensics. The features used in this are either MFCC or DMFCC. , but still, there are a lot of scopes for new techniques to be discovered. The artificial neural network and fuzzy logic of type I techniques are utilized in this thesis for speaker recognition. The fuzzy logic of type II such as Choquet fuzzy integral etc. can also be alternate for these techniques.

3. Requirement Specification

This section describes the user characteristics and functional requirements.

3.1 User Characteristics

There are two types of users in the application one being an authenticated User model for the teachers and the second user type will be the students. Since it is an attendance-based application the features provided to the authenticated teacher model will be adding students, taking voice-based attendance, and making an excel sheet of the attendance. The student can only interact with the software while he/she gives their attendance.

3.2 Functional Requirements

(i) Functional requirements 1
Actor**:** User
Input**:** Student Details and Voice Recordings
Description**:** The API processes audio and converts it into base64 format and sends it and other student details to the back-end where the audio is decoded back and stored into an audio file which is later stored as a model including the student details. The model is then stored in the database.
(ii)Functional requirement 2
Actor: User
Input: Student Attendance Voice
Description: The API processes audio and converts it into base64 format and sends it to the back-end where it is converted again to mp3 format and saved with the roll no of the student. This is used later by the model to mark the student's attendance.
(iii)Functional requirement 3
Actor: ML Model
Input: Student Attendance Voice Object
Description: The API processes audio and converts it into base64 format and sends it to the back-end where it is checked with the other trained data of students. The voice which is mathematically most similar to the attendance voice is given the attendance.

3.3 Dependencies

There are two main web frameworks used in the construction of the web application: -
Django:- A python-based web framework used to make the back end and the API
Apart from the frameworks, the main front-end tools used for the website are HTML, CSS, and JS
There are a bunch of libraries also used in the website mainly for tasks related to machine learning, natural language processing, and data manipulation like: -

- NumPy
  - TensorFlow
  - Django-rest framework
  - Pandas
  - Scikit learn
  - Librosa
  - Matplotlib
  - Keras

3.4 Hardware Requirements

(i) Client Side

To run the web application user needs to have a functional device like mobile\desktop\laptop\tablet etc. with an internet connection and an upgraded web browser like
Desktop browser: Safari, Chrome, Firefox, Opera, etc.
Mobile browser: Android, Chrome mobile, iOS safari, etc.
An additional feature requires a cardboard VR but even without that user can take care of all other requirements of the website.

(ii) Server Side

- Any operating system
- At least 2 GB of RAM
- Python compiler installed
- Django installed
- Python libraries installed are mentioned in the dependencies

4 . Proposed System

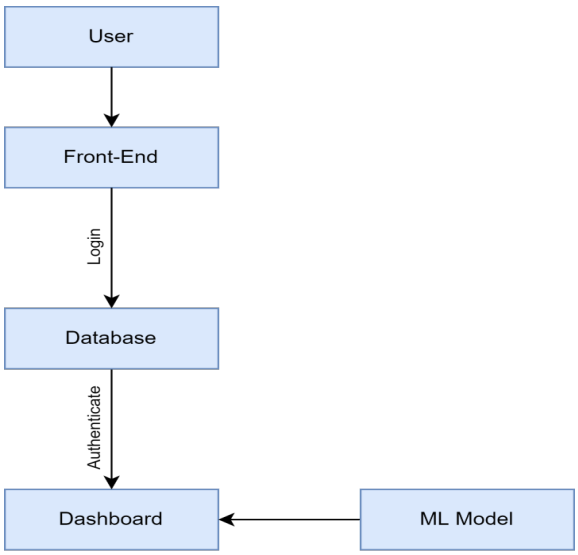


Fig.1. *Working on the website*

The above diagram shows the brief working of our website, this includes Front-end, Back-end, database, and machine, learning models.

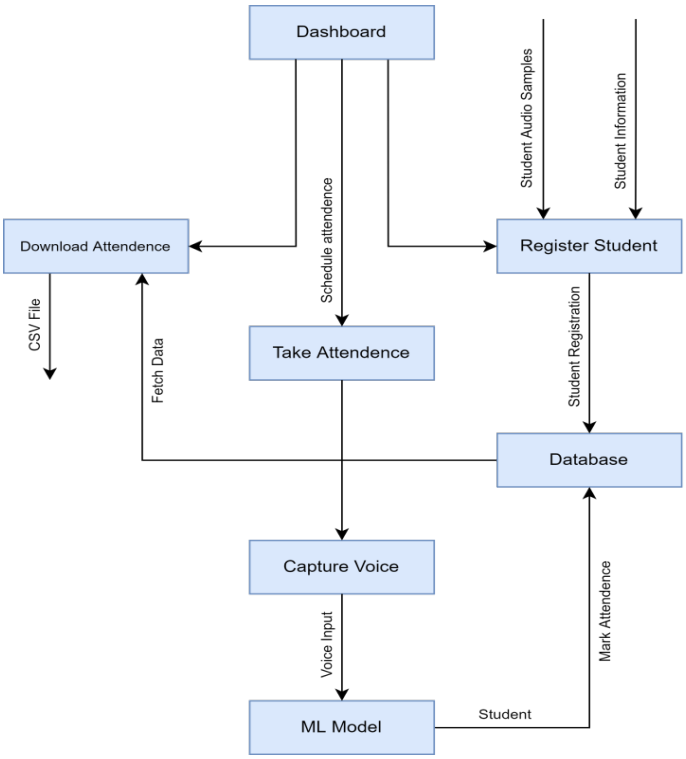


Fig.2. *In-depth work on the server*

In order to use the website, the faculty first have to log in to the system. Once logged in they will be redirected to the dashboard where they can Download the attendance, Register Students, or Schedule the attendance for a lecture.

4.1 Register Students

Students must first be registered in the database before their attendance may be recorded. In order to train the ML Model to recognize their voice, students' data such as their Roll Number, Name, Batch, Division, and 3-4sec Sound Samples are captured.

4.2 Schedule Attendance

Now Students must record a 3-second sound sample in every lecture in order to document their attendance. This will be fed into the ML model in order to recognize the students, and then the attendance of recognized students will be recorded.

4.3 Downloading Attendance

To retrieve lecture attendance information, class specifics such as batch, course, and date will be needed. The lecturer will be able to obtain a CSV file having all of the details of the lecture.

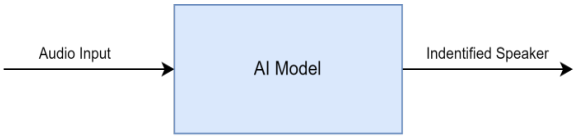


Fig.3. *Speaker recognition machine learning model*

The ML Model takes the Student's Audio as Input Predicts the student and marks the attendance of that student.

The Process for predicting the student based on Audio:

1. The Audio is Recorded in Front-end as an audio blob and converted into a base64 string then the base64 string sends to the server using an AJAX request.
2. The base64 string is received at the back end and it is converted into Audio File.
3. Then the Audio file is given to the ML Model as Input to predict the student.
4. Then the Audio file is loaded using the Librosa library and audio vector and its sample rate is extracted from the audio file.
5. Using the Audio Vector features like MFCC, chroma_stft, RMSE, spec_cent, spec_bw, roll-off, zcr are extracted and put into a python list.
6. The python list is then converted into a NumPy array and reshaped to shape like (1, features. shape[0])
7. Then the array is scaled using the standard scaler.
8. And scaled features are used to predict the student roll number using which we can extract the details of the student from the database.
9. Once the student's detail is found using the roll number than his/her attendance is marked.

The Machine learning model takes the speaker's voice as input and tells us who is the speaker. This model is created using a Deep neural network, we use the Sequential model to create a neural network of our model to predict emotion in the audio file. The Sequential model is a way of creating deep learning models where an instance of the Sequential class is created and model layers are created and added to it.

5. End-user Interaction

5.1 Register Students

Faculty can use this feature to register new students. This student's academic details and audio samples are used to train the model, and the information is maintained in an SQLite database. Once a student has been registered, he or she can quickly record their attendance.

5.2 Record Attendance

This feature allows faculty to record student attendance; attendance can be scheduled from the database with the appropriate Batch, Class, and Teacher selected; students can come in and record one audio sample, which will be passed to the ML model, where the student will be identified and the attendance marked; and students can come in and record one audio sample, which will be passed to the ML model, where the student will be identified and the attendance marked; and students can come in and record their one audio sample, which will be passed to the ML model, where

5.3 Download Attendance

This program allows lecturers to download attendance for any class they want. They must choose a Batch, a Course, and a Date. After the information is filled out, it is retrieved from the database and transcribed into a CSV file that the lecturer can access.

6. Results

These sections show the result and accuracy of the different machine-learning models in this project.

6.1 Students Classifier

The Students classifier is a machine learning model in which the audio file is supplied as a parameter, and the ML model predicts the student based on the audio file.

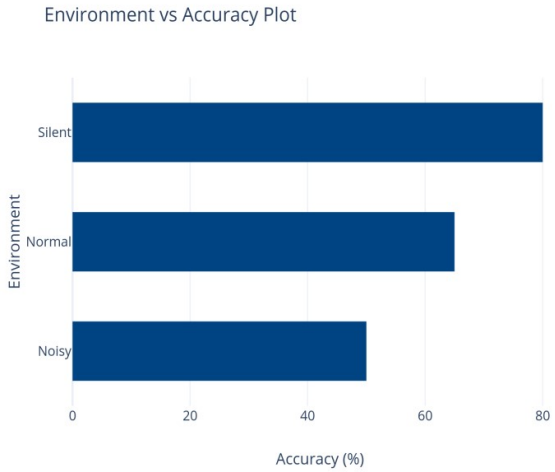


Fig.4. *Model Accuracy graph with respect to different environments*

The Student Classifier Model performs best in quiet environments with no background noise, with an accuracy of around 80% or higher; however, when the environment becomes slightly noisy, the accuracy drops below 50%, and when the environment has too much background noise, the accuracy drops below 50%.

7. Future Improvements

Some of the possible improvements which can be done in this project are:

7.1 Improving ML Model

The model is now under-trained due to the usage of fewer audio samples; however, in the future, a concept known as "Data Augmentation" could be implemented to help solve the problem of under-trained models.

7.2 Noise Reduction

Noise in audio files is another factor for the ML model's low accuracy. Noise Reduction, which removes excess noise from the background, can be used to significantly increase the accuracy of the ML model.

8. Summary

The goal of introducing this automated attendance system was to reduce the errors that occur in the traditional attendance-taking system. We can solve the problem of bogus attendance while simultaneously saving time and increasing efficiency by implementing this technology. Aside from attendance, voice-based speaker identification offers a wide range of applications, including smart homes, digital assistants, and more.

Compliance with Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

[1] Luengo, Iker & Navas, Eva & Sainz, Iñaki & Saratxaga, Ibon & Sánchez, Jon & Odriozola, Igor & Hernáez, Inmaculada. Text Independent Speaker Identification in Multilingual Environments. 2008.
[2] Chougule, S. & Rege, Priti. Language Independent Speaker Identification. Pp 364 – 368, 2007.

[3] Bose, Smarajit & Pal, Amita & Mukherjee, Anish & Das, Debasmita. Improved Language-Independent Speaker Identification in a Non-contemporaneous Setup. International Journal of Machine Learning and Computing. Vol. 10. Pp. 630-636. 2020.
[4] Nidhyananthan, Selva & Selva Kumari, R.Shantha. Language and Text-Independent Speaker Identification System Using GMM. Language and Text-Independent Speaker Identification System Using GMM. Vol. 9. 2013.
[5] Reynolds, D.A. & Rose, Richard. Robust text-independent speaker identification using Gaussian Mixture speaker models. Speech and Audio Processing, IEEE Transactions on. Vol. 3. Pp. 72 - 83. 1995.
[6] Singh, Satyanand & Rajan, E.. Vector Quantization Approach for Speaker Recognition using MFCC and Inverted MFCC. International Journal of Computer Applications. vol. 17. 2011.
[7] Samudre N. A, Text-Independent Speaker Identification using Vector Quantization, International Journal Of Engineering Research & Technology (IJERT) Volume 02, Issue 08 (August 2013). 2013.
[8] Leu, Fang-Yie, and Guan-Liang Lin. "An MFCC-based speaker identification system." In 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), pp. 1055-1062. IEEE, 2017.
[9] Reynolds, Douglas A. "Speaker identification and verification using Gaussian mixture speaker models." Speech communication 17, no. 1-2 (1995): 91-108.
[10] Saeed, Khalid, and Mohammad Kheir Nammous. "A speech-and-speaker identification system: Feature extraction, description, and classification of speech-signal image." IEEE transactions on industrial electronics 54, no. 2 (2007): 887-897.
[11] Wu, Jian-Da, and Yi-Jang Tsai. "Speaker identification system using empirical mode decomposition and an artificial neural network." Expert Systems with Applications 38, no. 5 (2011): 6112-6117.
[12] Islam, Md Atiqul, Wissam A. Jassim, Ng Siew Cheok, and Muhammad Shamsul Arefeen Zilany. "A robust speaker identification system using the responses from a model of the auditory periphery." PloS one 11, no. 7 (2016): e0158520.
[13] Kumar, Chandar, Faizan Ur Rehman, Shubash Kumar, Atif Mehmood, and Ghulam Shabir. "Analysis of MFCC and BFCC in a speaker identification system." In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), pp. 1-5. IEEE, 2018.