Voice-based Login System

Vedant Baviskar[1], Shubham Bhagwat[2], Deepjyoti Barman[3], Gokul Viswanath[4],

Mrs. R.T. Waghmode [5]

[1,2,3,4,5] *B.Tech student, Computer Science, Sinhgad Institute of Technology and Science, Pune, India*

---***---

Abstract - *This project on Voice Authentication analyses audio to determine the identity of a user and authenticate them into a system while using Gaussian Mixture Models. This has a wide set of uses, from a normal web application to a high-security facility.*

Understanding how complex and high dimensional audio can be recognized is one a challenge. Audio data contains information that can be used effectively for information or can be filtered out (in the case of our project) using Audio Content Analysis. This information includes background noise, crowd chatter and other noise like animals or cars. These reasons make audio recognition and classification a very challenging research topic. High dimensional audio can be exceedingly large in size and can pose problems to the speed of authentication and the performance impact on the devices used. So to solve this problem, we use audio formats such as MPEG to maintain the high level of detail in the audio clip, while also keeping the file size small.

***Key Words*: High dimensional audio, Audio Content Analysis, Moving Pictures Expert Group, Gaussian Mixture Model**

1. INTRODUCTION

Voice Biometrics is a technology that relies on the recognition of voice patterns to verify the identity of individuals. This is possible as each person's vocal tract is unique. Physical features, both phonetic and morphological are unique to each individual making them virtually fraud proof. It enables fast, frictionless and a very secure way to authenticate users.

This process of authentication works in two steps, the first being the classification of the audio into speech and non-speech audio, the latter being music, noise and environment sounds.

The second part starts by creating a list of characteristics that the trained model perceives from the voice, which is stored in an array. Then the user's characteristics that were saved during registering will be downloaded and used to compare both arrays. If the distance between both the arrays is below a specific threshold value, then the user is authenticated.

1.1 RELEVANCE

There are many areas that can use Biometric identification (in this case, Voice Authentication). The main areas this can be used are the ones where security is prioritized, ranging from a company login for admin tools to a high-security military checkpoint to watch border crossings at different frontiers.

Higher ranked military forces can use this to regulate entry to particular top-level facilities.

Although the military is one main sector, even normal social media logins can be done using voice authentication, if they wish to.

When a person is driving a vehicle, they can use systems that use Voice Authentication and recognition to use services while also keeping their eyes on the road, thus contributing to road safety.

People with dyslexia would experience difficulties reading items on a screen and thus can use Voice Authentication to log in and use a service without much difficulty.

1.2 MOTIVATION

As stated before, analysing and understanding multidimensional audio data is a big challenge. Using data from that analysis and implementing it in a real-time system is even more so.

Such analyses and implementations can provide useful information to us about audio, and how they can be used in a wide range to applications that we use every day as part of our lifestyle. The voice is something you are, and uniquely defines you and your characteristics. This can help people and organizations introduce more such platforms that use Voice Authentication, and thus make the users' life easier while maintaining (or even strengthening) the level of security.

2. REQUIREMENTS

Following are the requirements for the Voice Authentication system.

1. Server side
   - Language – Python
   - Libraries – FastAPI, Fastdtw, OCI, Scipy, Ffmpeg-python, Numpy, Python-jose
2. Client-Side
   - Language - HTML, CSS, JavaScript
   - Libraries – React, Next, Axios, Recorder.js

3. Methodology

In order to authenticate the voice, we first get the users voice sample.

We take a base sample from the user, the first voice they upload, and we pass it through our feature extraction pipeline. After that we take at-least 4 more samples of the users voice and pass them through the same pipeline. This pipeline, at the end, returns a value that is between 0 and 1.

This value indicates that the distance (or angle) between the two voices is most if the value is 0 and least if the value is 1. In other words, the larger the value, the closer two voice samples are.

Once we take the users samples, we do a mean of all the values and determine an optimum value out of it. This value is what we use in order to compare the login voice sample.

When a user tries to login, they pass a voice sample. This voice sample is also passed through the same pipeline and it is compared to the optimum value that we store. Depending on the result of the above comparison we either let the user login or not.

The pipeline that we have built, does a few things. First step is to reduce the noise from the audio. We do this by utilizing a Python library that determines noise based on the frequency.

Once we pass the audio through this, we pass it through another algorithm that returns a spectrogram of the audio.

After this, we pass the second audio sample through the same above steps and we end up with a spectrogram as well.

Once we have two spectrograms, we find the cosine similarity of the two spectrograms in various plots. The value of cosine similarity is between 0 and 1. Once we have the cosine similarity of various locations in the spectrogram, we find the mean of those values and get a single value.

This value is the overall cosine similarity which is also between 0 and 1. This value indicates how close the two audio samples are.

As explained above, depending on the above value, we either let the user log-in or show an error.

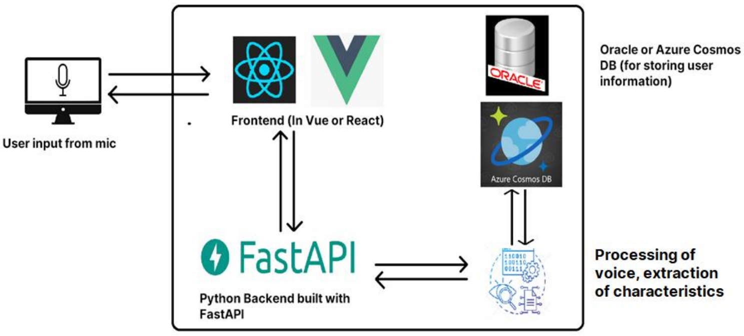
User input from mic
Frontend (In Vue or React)
Oracle or Azure Cosmos DB (for storing user information)
ORACLE
Azure Cosmos DB
FastAPI
Python Backend built with FastAPI
Processing of voice, extraction of characteristics

Fig -1: Project Architecture

3. CONCLUSIONS

During this project and the experiments we conducted during this project, we uncovered a few limitations. They are:

- Recognizing a user's voice if it changes due to sickness, accent changes or speech patterns is a challenge.
- Authentication speed and processing impact on the host device, while being low, is still high compared to other techniques.
- Because voice technology is easier to spoof as compared to other forms of biometric authentication, liveness detection tests needs to be used to verify if the user trying to log in is actually the user and not someone else trying to spoof the system.

These limitations can definitely be solved in the future with advances in technology and the increase in the speed and power of processors, but currently hold back Voice Authentication in being a widespread authentication technology.

As the years go by, the number of users of digital platforms for government services as well as the private sector services is increasing rapidly. Such a secure and foolproof authentication technique is a needed to ensure that the login process is smooth and easy while also maintaining the same high level of security.

As Voice authentication can uniquely identify an individual by their different traits i.e. identifying you by "who you are" and not by "what you have", it is becoming a widely used technique. Its low chance of failure and easy usage makes it an attractive option in systems around the world.

ACKNOWLEDGEMENT

We would like to thank our Head of Department Mrs. G. S. Navale, for giving us all the help and important suggestions all over the Seminar Work along with our internal guide Mrs. R.T. Waghmode, for giving us guidance and support throughout the Seminar. Lastly, special thanks to our Principal Dr. S. D. Markande for providing ambience in the college, which motivate us to work.

REFERENCES

[1] Deep Learning for Audio Signal Processing - Hendrik Purwins∗, Bo Li∗, Tuomas Virtanen∗, Jan Schlüter∗, Shuo-yiin Chang, Tara Sainath, M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[2] Unsupervised feature learning for audio classification using convolutional deep belief networks - Honglak Lee, Yan Largman, Peter Pham, Andrew Y. Ng

[3] An Introduction to Audio Content Analysis Applications in Signal Processing and Music Informatics – Alexander Lerch

[4] librosa: Audio and Music Signal Analysis in Python - Brian McFee, Colin Raffel, Dawen Liang, Daniel P.W. Ellis, Matt McVicar, Eric Battenberg, Oriol Nieto

[5] Biometric Audio Security – Lloyd Trammell, Lawrence Schwartz

[6] Audio Data Analysis Using Deep Learning with Python (Part 1) - Nagesh Singh Chauhan

BIOGRAPHIES

Mrs. R.T. Waghmode is currently working as a Professor in Sinhgad Institute of Technology and Science, Narhe.

Vedant Baviskar is a final-year B.E. student in Savitribai Phule Pune University.

Shubham Bhagwat is a final-year B.E. student in Savitribai Phule Pune University.

Deepjyoti Barman is a final-year B.E. student in Savitribai Phule Pune University.

Gokul Viswanath is a final-year B.E. student in Savitribai Phule Pune University.