HIGH LEVEL VOICE BASED WEBSITE AUTHENTICATION SYSTEM

Mr.V Chandrasekhar[1] ,A.Chandramouli[2]
[1]Assistant Professor, Dept of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, AP, India.
[2]PG Scholar, Dept of MCA, Audisankara College of Engineering and Technology

Abstract- Password and authentication systems now in use are insecure and frequently subject to hacking attacks. Secure authentication has become increasingly important as there is an increase in the amount of information available online. Because vocal biometrics, like fingerprints, produce individual identities for each person, biometric-based authentication offers a viable alternative. We analyse the state-of-the-art voice biometric software, suggest a security protocol for a voice authentication system, and offer two distinct implementations of potential voice authentication systems that try to address some of the shortcomings of the opensource alternatives. A review of the benefits and drawbacks of our own solutions is also provided. The best-known commercialised version of vocal biometrics is the Voiceprint Recognition System, commonly referred to as a Speaker Recognition System (SRS). Automated speaker recognition is the process of using voice features to validate a user's stated identification. The speaker recognition systems are created for use with virtually any standard telephone or on public telephone networks, in contrast to other biometric technologies that are primarily image-based and necessitate expensive proprietary hardware like vendor's fingerprint sensor or iris-scanning equipment.

Key Words: Speech recognition, modelling , speech processing, training and assessment

1.INTRODUCTION :

The usage of voice recognition in applications has grown in popularity over the past few years. Applications that can translate audio and comprehend what a user is saying include Google Now and Siri. Some of these programmes can also identify people specifically. Meanwhile, programmes like Shazam and initiatives like Music Brain have developed the ability to recognise music from just two to three seconds of recorded music. Traditional audio processing has been too computationally expensive to be done locally, however this is changing as new processing methods are developed. Innovative new initiatives are also offering inventive methods to use audio for two factor verification. With the help of virtually inaudible noises played through a computer's speakers and a smartphone app that listened to the sound, analysed it, and verified it to the server, Slick Login (bought by Google) made it possible to authenticate users using two factors employing audio. Each user's generated sounds were distinctive and secure so that they couldn't be played back later.

Two distinct methods of authentication have emerged as voice-based

authentication has progressed. The first method involves having someone repeat the same phrase several times while building a very broad template out of the different voice prints. The new voice print created can be compared to the user's previous voice prints to authenticate the use

when they speak in the future. The disadvantage of this method is that the voice print is less specific and hence cannot be authenticated with the same level of precision as voice prints produced with a single password. The second strategy involves recording only one voice print, made up of a single word or phrase.

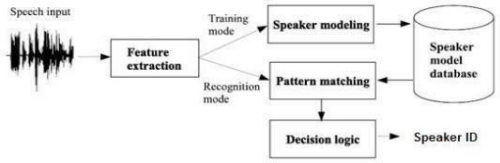


Fig: Speaker Recognition

When this method is applied, a third party may be able to record the authentication attempt and replay it to get access. As part of our examination of the various voice authentication strategies, we looked into both of these procedures. The proposed method we expect will successfully identify people using distinctive voice prints is described in the following sections.

LITERATURE SURVEY :

A biometric is an individual bodily characteristic. The number of applications for conversation systems has grown significantly in recent years, which has contributed to their appeal. However, the field of automated speaker recognition still has a great deal of open issues. Choosing the appropriate speech signal characteristics and machine learning algorithms may be one of them. While looking for the best way to achieve our aim, we tried a number of different blogs, articles, videos, modules, libraries, etc. and encountered failure after failure. Here, we'll highlight a few of their primary arguments. Applications.We have also investigated, compared, and attempted to include many approaches in order to find the best effective model for speaker recognition.

2.SECURITY POLICY AND THREAT MODEL :

We established a security strategy and threat model that detailed the parameters of our project and served as a decision-making tool before analysing opensource alternatives and putting our voice authentication method into use.

2.1. Voice Authentication System :

A client-side application and a server make up the authentication mechanism. On a party's device, whether it be a tablet, laptop, or mobile device, the client side programme will be installed. The system must have access to the client-side recording mechanism on the device and be linked to the Internet. The client-side application will transfer the recorded audio files across the Internet while making API calls to the server to authenticate users. The server will listen to the audio and fingerprint it to determine if the user is the one they are attempting to authenticate as. The technique is meant to be utilised as a second factor in two-factor authentication due to flaws in voice fingerprinting. The

technology is designed to be used as a second factor in two-factor authentication because voice fingerprinting has flaws. Through an enrollment process where many audio samples from one user are used to extract features and build a general voice fingerprint for a user, the server maintains the speech features connected to a user's account.

2.2. Person of Interest :

An individual who has an account in the voice authentication system or will do so is the subject of interest. In the event that the subject of your inquiry has a voice authentication account, you can use voice authentication to access it. In the solutions we have examined and in ou implementation, we presum that the subject of the investigation has successfully supplied the first factor of two-factor authentication. The only person who can access their account is the subject of your investigation. The device they are logging in with does not have to belong to them. If the subject of interest doesn't already have an account in the system, they can sign up for it there. Only the person of interest should be permitted to add their vocal features to the voice authentication system. By using certain words, a person of interest can register for the voice authentication system. The sentences are then captured, and the system analyses the audio's properties. By registering with the system, the individual of interest is also given access to an account with those characteristics. These features are kept in the user's previously created account when using two-factor authentication. The user will try to verify themselves by reciting a brand-new phrase that has been supplied to them in order to log into the system. When the voice file's characteristics match the ones that were recorded in the system upon enrolment, a person has successfully authenticated themself.

3.3.Attacker :

The account of the person of interest will be targeted by an attacker. We state that the attacker won't be able to access the account using our system since the voice used must match what was entered during enrolment and has been logged in the system. Additionally, the hacker shouldn't be able to access the system by playing back a recorded version of the target's voice without that person being there.

We presume that the attacker cannot get beyond the voice authentication method and access the model used for authentication as our system design is not meant for the construction of a secure server.

4.ANALYSIS OF CURRENT AVAILABLE OPTIONS :

We investigated Microsoft's Speaker Identification API Bob, Dejavu, and three more audio fingerprinting and speaker recognition alternatives. In this part, we examine Dejavu's security and Microsoft's API and Bob.

4.1. Microsoft Speaker Recognition API :

The Microsoft Speaker Recognition API was one of the libraries we looked into. This is one of the publicly accessible APIs that promotes identification and verification. The web application's developer must register for an account with the API in order

to receive a private key, which they must include in the request header. A user must speak for at least 60 seconds without any silences in order to be added to the system. The system listens to any audio of a user speaking in order to verify their identity. The user can work in a noisy environment thanks to this API's high accuracy and ability to remove the majority of background noise. The user doesn't have to wait too long to hear back about their authentication because to the application's quick processing speed. The API does contain some backups, though. The enrollment terms are all common and provided on the website, which allows an attacker to foresee which Enrollment phrases a user could use and pre-record those phrases. As there is no unique audio passphrase required for authentication, the words can be played back if an attacker tries to hack into an account. In this case, the service will authenticate the attacker. Since the registration words are predetermined, a hacker might potentially record the enrollment phrase and use it to establish a different account with the same voice fingerprint as the user, which may make authentication more challenging. Second, since the request is not hashed to sign it and the private key is presented in the header as plaintext, it is possible to intercept the API call and change the contents. The attacker can then add their own recordings for registration and authentication. The API can optionally accept a pre-recorded sound file instead of requiring real-time passphrase recording.

4.2. Bob :

Additionally, we made an effort to use Bob, a signal-processing and machine learning toolbox created by the Biometrics team at the Idiap Research Institute. We had hoped that using this library, which has been utilised in other speaker recognition systems, would aid in the development of our application. However, we discovered that there were numerous incompatible dependen- cies and that the dependen cies that could be installed consumed a significant amount of memory on our machines. If we had used the voice corpus that was provided, we would have had to apply numerous filters to any user's recording in order to convert it to the desired format. Therefore, it would be computationally more time-consuming than desired to undertake any speaker identification trials. These limitations led us to choose not to use Bob for our.

5.LOGGING INTO THE SYSTEM :

The python audio fingerprinting library, Dejavu appeared to be the best existing library to further investigate due to its light weight, simple usability, and accuracy after examining a range of prospective voice au- thentication libraries. Audio fingerprinting is a method used by Dejavu to match audio samples. Creating a 2D array spectrogram, whose amplitude is a function of time and frequency, using fast Fourier transforms in overlapping windows over the course of an audio sample is known as audio fingerprinting.

The corresponding frequencies and timings of the amplitude peaks are retrieved from this spectrogram. The peak frequencies and the time intervals between the peaks are then hashed. A short window of the sample is used to represent a distinct audio fingerprint using this hash. As a result, each sample is represented by various audio fingerprints.

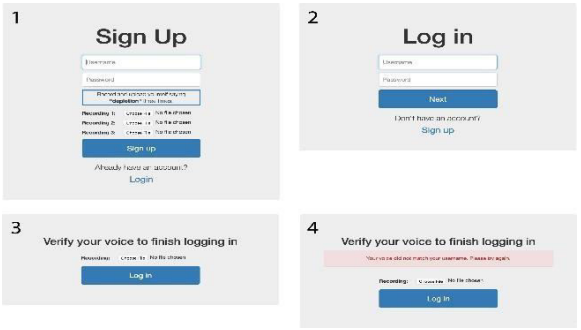
These fingerprints for an audio sample are then compared to fingerprints already present in a database to determine which sample already exists is most likely the source of the test sample.

In music detection apps like Shazam and SoundHound, the method of audio fingerprinting is frequently applied. Dejavu's speech recognition capabilities have a few clear advantages in light of this. Despite different degrees of background noise, Dejavu is incredibly precise in matching audio samples

These benefits, however, are not without drawbacks. Dejavu is accurate because it matches audio samples, not voice samples, specifically. The user's enrollment and authentication phrases must be the same as a result. This leaves our system open to replay attacks, in which a malicious party may capture a user's activity throughout the registration or authentication process, play back the video, and then utilise it to successfully log into the system. Additionally, because each user's vocal fingerprint is so distinct, a change in voice quality, such as a cold, might prevent a user from properly logging into the system.

Dejavu offered audio fingerprinting rather than voice fingerprinting, but we choose to start there since it allowed us to build a first prototype of our system. We created a straightforward Web API wrapper for Dejavu that allowed users to sign up and then authenticate themselves. In a two-factor authentication system, we intended for our solution to operate as the backup authentication mechanism. So, using Dejavu API that we developed, we made a straightforward web application that required a password and a voice sample to log in.

Fig : Login into the System



6.GMM IMPLEMENTATION :

We made the decision to put our own speech authentication system into practise as a proof of concept, using a Gaussian Mixture Model (GMM) to cluster based on characteristics collected from audio spectrums.

The procedure for establishing an account in our implementation is shown.

1) Register the user using random phras2) Speech recognition to ensure that the user's words and the produced words are same

3) Try voice activity detection

Rationale Behind GM Rationale we use comparable signal processing techniques, the procedure for authenticating an account is similar.

1) Request the username of the account the client wants to use to log in; 2) Request that the client say a series of randomly generated sentences.

3) Speech recognition to ensure that the user's words and the produced words are same

4) Try voice activity detection

5) Extract and normalise MFCC characteris

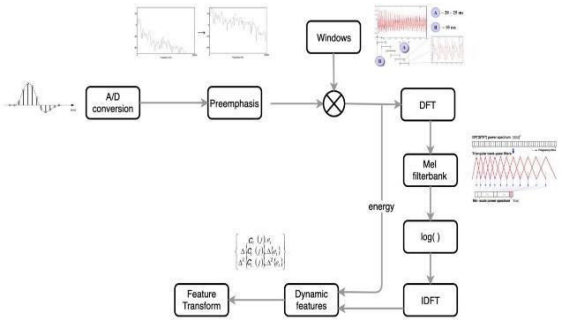


Fig: GMM

The MFCC attempts to understand the vocal passage in order to get the optimal outcomes, as seen in the above figure, which demonstrates how it works with voice input.

6.1. Gaussian Mixture Model :

We build a Gaussian Mixture Model (GMM) for every user in order to authenticate them. A GMM can combine many Gaussians to create a distribution that reflects the possibility of a feature vector landing in a certain area, with the guarantee that the probability will never be zero. In a model of K Gaussians, the probability that a feature vector x belongs to the model is given by the formula: K p(x wi, I I = wi (x, I I (1) i=1, where (x, I I is a normal distribution with mean I and variance I [15][21]. The total of the wis must equal one in order to determine how much influence each Gaussian has on the model.

A user can try to log in after we've constructed the model. Following feature extraction, the log-likelihood is calculated for the recording's features using score samples in Sci-kit Learn's GMM implementation. The per-sample likelihood of the data using the model is returned by score samples [17]. The model with the highest log-likelihood is chosen. We indicate that the user has successfully signed in and may access their account if this model matches the account that they were trying to enter into. If not, we indicate that the user is attempting to log into a non-personal account.

As demonstrated in the work done by Reynolds et, we chose a GMM over other machine learning methods because it is particularly helpful in speaker recognition. Assuming that feature vectors are independent, the model incorporates several Gaussian distributions. By using characteristics from the speaker's introductory sentences, a model is generated for each speaker. Features are again derived from a sentence given to the speaker when they try to log in. We determine whether the extracted features are likely to be part of the user trying to log in model or whether there are other models that more closely match these characteristics using an equation for log likelihood. If the characteristics match the speaker's model that is attempting to log in, whether there are any alternative models that fit these features more effectively. We may infer that the speaker is successfully login into their own account if the features match the model of the speaker who is attempting to log in.

6.2. MFCC :

The sound's immediate power spectrum is represented by the Mel-Frequency Cepstral Coefficient (MFCC). It serves as a method to extract characteristics from audio in automatic speech recognition.

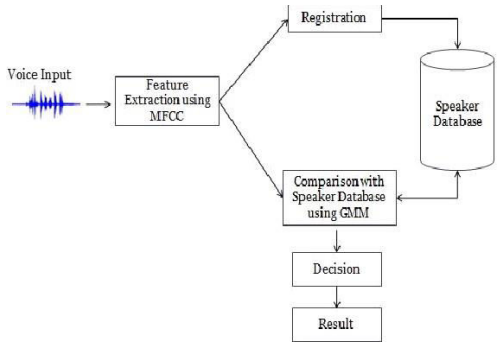


Fig: MFCC

In addition to being included in voice fingerprinting implementations, MFCCs are the most often employed characteristics in speech recognition features. Initial brief time frame division of the signal. We compute the discrete fourier transform for each of these panes. The powers of this spectrum are translated onto the Mel scale, a logarithmic curve that simulates pitches that are normally perceived as being equally distant from one another. In order to do a discrete cosine transform, we first obtain the log of the powers at each of the mel frequencies. The spectral coefficients that we obtain from the cosine transform are what we refer to as the features we extract, or MFCCs

Using speech characteristics from Python, we extract these features. Following normalisation using the mean and inverse standard deviations, we employ these features in our Gaussian Mixture Model.

7.FUTURE WORK :

While we have developed a functional voice authentication system, there are a number of ways that it and our voice authentication security policy might be strengthened so that it can be utilised with confidence in two-factor authentication.

7.1. GMM Improvements :

The log likelihood of acquiring each model presently stored on the system for each characteristic collected from the audio recording is computed when a user tries to log in. The processing time for this might increase to the point where the user's logging in latency becomes unacceptably high for a system with a big number of users. Potential assaults might also result if a hacker establishes a large number of fictitious identities to slow down the system's computations for everyone attempting to log in. We'd want to consider altering our GMM model to incorporate the idea of K-means in order to be able to account for clusters as a potential remedy. We might then only study a subset of the GMMs, decreasing computing time and avoiding any latency attacks, after we know which models will most likely have a high log probability for the characteristics we are looking at, based on clustering of the GMMs. As a result, our system will be more scalable.

7.2. Replay Attacks :

Even though we produce a list of random words each time a user signs up or attempts to log in, we are only using terms from a small corpus of English nouns, and certain words have been in passphrases several times. A malevolent actor might easily record a user repeating these phrases and eventually amass a sufficient corpus to be able to log in as that user by playing back the recorded audio pieced together. There are a few options for fixing this First, we may increase the number of words in our corpus, making it more difficult for an attacker to capture every word that is said. Second, if we intentionally add a noise to the signal that is captured that is exclusive to that session, the recording will be

dismissed as a replay attack if the noise is found there.

7.3. Voice Generation Attacks :

We speculate that future methods for generating a person's voice from a predetermined set of samples may exist, similar to replay attacks. Because the challenge text would be difficult for the machine to understand, this would allow anyone to create audio from it. Even if this isn't a problem in the actual world right now, we admit that it might in the near future, making it a fault that future generations of the system should try to address.

7.4. Securing Voice Print Data :

Biometric information, as we just established, is both distinctive and difficult to change. The need for protecting voice print data and voice models is therefore considerably more pressing. This would be applicable to all programmes that use speech biometric data since, if the samples from one programme were to be compromised, the samples from other programmes would also be exposed to assaults.

8. CONCLUSION :

We analyse the various voice authentication techniques now in use and offer a security strategy, along with two distinct implementations of biometric two-factor authentication. Our solutions offer quick and easy ways to add two-factor authentication to already-existing apps while also outlining potential directions for further research. Our approach lays the path for further development in open source biometric authentication systems by providing a speech biometric based authentication system.

9. REFERENCES :

[1] “bob.bio.spear 2.0.4–Tools for running speaker recognition experiments,” https://pypi.python.org/pypi/bob.bio.spear/ 2.0.4

[2] J. Bonneau et. al., “The Quest To Replace Passwords,” IEEE Symposium on Security and Privacy, 2012, pp. 553-567

[3] P. Cano et. al., “A Review of Audio Fingerprinting,” Journal of VLSI Signal Processing, 2005, pp. 271-284

[4] “Dejavu: Audio Fingerprinting and Recognition in Python,” https://github.com/worldveil/dejavu

[5] “Fernet (symmetric encryption),” https://cryptography.io/en/ latest/fernet/

[6] “FuzzyWuzzy,” https://github.com/seatgeek/fuzzywuzzy

[7] “Levenshtein Distance,” https://en.wikipedia.org/wiki/ Levenshtein distance

[8] “Man-in-the-middle Attack,” https://en.wikipedia.org/wiki/ Man-in-the-middle attack

[9] “Mel-Frequency Cepstrum,” https://en.wikipedia.org/wiki/ Mel-frequency cepstrum

[10] “Microsoft Cognitive Services: Speaker Recognition API,” https://www.microsoft.com/cognitive-services/en-us/ speaker-recognition-api

[11] R. G. Hautamaki, T. Kinnunen, V. Hautamaki, and A.-M. Laukkanen, *Speech Communication* 72, 13 **(2015)**.

[12] Available: http://www.biometricsinstitute.org/pages/types-of-bio-metrics.html, Biometrics Institute, Australia **(2016).**

Author's Profile:

Mr.V.CHANDRASEKHAR has received him MCA degree from Sri Venkateswara University in 2001,Tirupati respectively. He is dedicated to teaching field from the last 21years. He has guided P.G students. At present he is working as Associate Professor in Audisankara College of Engineering and Technology, Gudur, Tirupati(Dt), Andhra Pradesh, India.

A.CHANDRAMOULI has Pursuing his MCA from Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, affiliated to JNTUA in 2022. Andhra Pradesh, India.