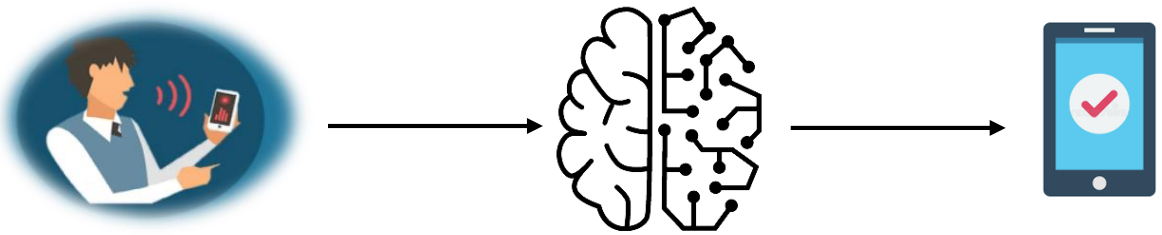
Voice Biometric

Solution for SBI Innovate 2022

Overview

"Biometrics" is the use of unique physical markers of an individual like fingerprints and iris.

Voice biometrics works according to the fact that human voice features are unique and correlate strongly to the physiological and behavioural qualities of how an individual creates speech. The result is a "*Voiceprint*" analogous to a fingerprint which can be used for authentication.



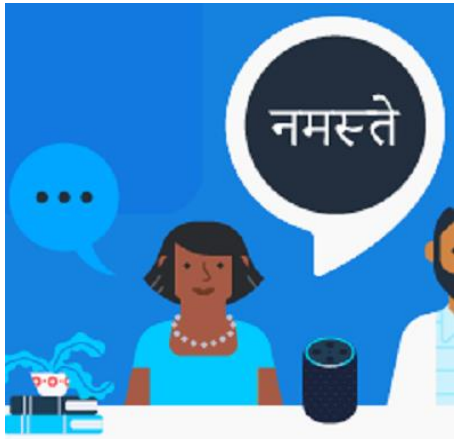
In our project, we are attempting the use of Voice Biometrics to verify the identity of a user simply by his/ her voice.

For this we have chosen certain models and mathematical computations to increase the accuracy of this process, implemented using Python and its library support.

Also, since an individual's biometric is highly sensitive information, we are using encryption methods to store and process the biometric information.

Advantages of our approach:

Signal enhancement and modulation for better Voiceprint



Language Independence - No dependence on “What is spoken"

Secure storage of sensitive data

Voiceprint accuracy continually increasing

Process Flow

Enrollment Phase

Bank Customer

Voice Sample

Voice Pre-processing:
- Background noise Suppression
- Silence Removal

Feature extraction:
- Using mel-frequency cepstral coefficients (MFCCs)

User modelling:
- Using Gaussian Mixture Model (GMM)

User Model Database

Authentication Phase

Model Enhancement, if user is authenticated

Voice Pre-processing:
- Background noise Suppression
- Silence Removal

Feature extraction:
- Using mel-frequency cepstral coefficients (MFCCs)

Similarity Scoring:
- Using Log-likelihood

Score generated

User is authenticated, if score is greater than the threshold decided

Unknown Speaker

Voice Sample

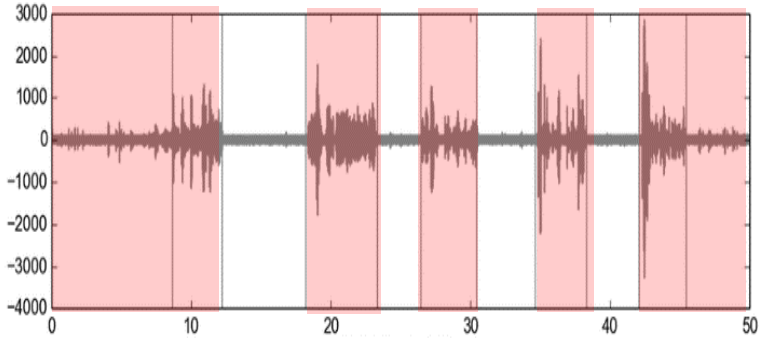
Pre-requisite: Voice clip provided by the user of a minimum of 10 secs

Pre-processing Phase

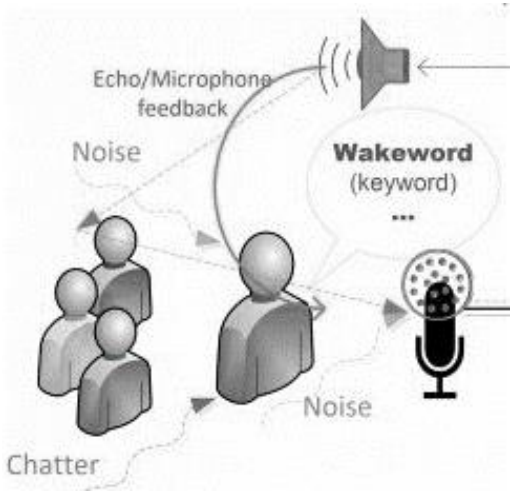
Input:

.wav

Voice sample is passed through pre-processing phase



Unnecessary silences will be removed so that **only the sections containing speech is processed further**



Sample audio will be passed for noise reduction **which will improve the quality of unique features extracted from the voice sample**

Result:

Enhanced voice sample

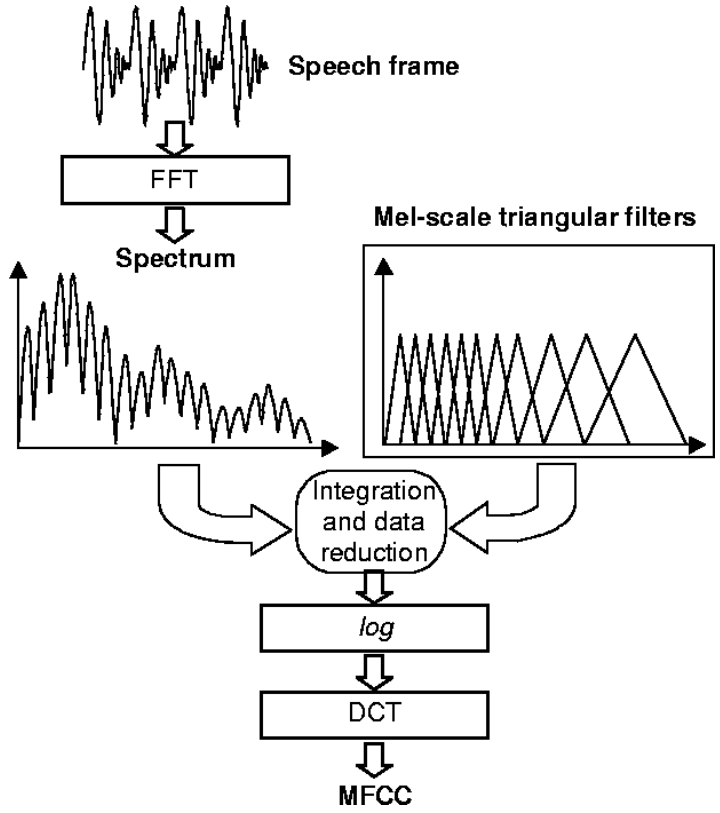
Feature extraction

Mel scale relates the perceived frequency of a tone to the actual measured frequency. It scales the frequency in order to match **more closely to what the human ear can hear.**

The sound generated by humans is determined by the shape of their vocal tract. **More accurate the prediction of the shape, better the representation of the audio signal.**

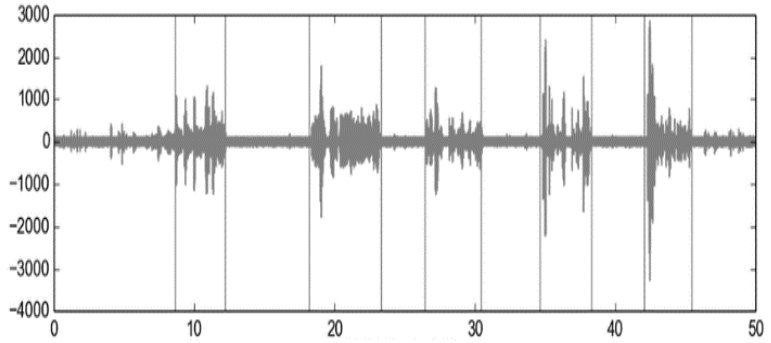
The envelope of the time power spectrum of the speech signal is representative of the shape vocal tract. The **Mel-frequency cepstral coefficients (MFCCs)** accurately represents this envelope.

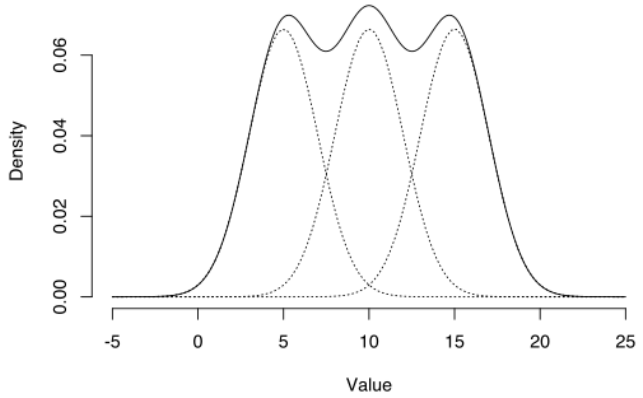
Hence, we have used Mel-frequency cepstral coefficients (MFCCs) to extract features from the voice sample.



User Modelling

The **Gaussian Mixture Model (GMM)** is a probabilistic model that assumes all the data points are generated from a mix of Gaussian distributions with unknown parameters.





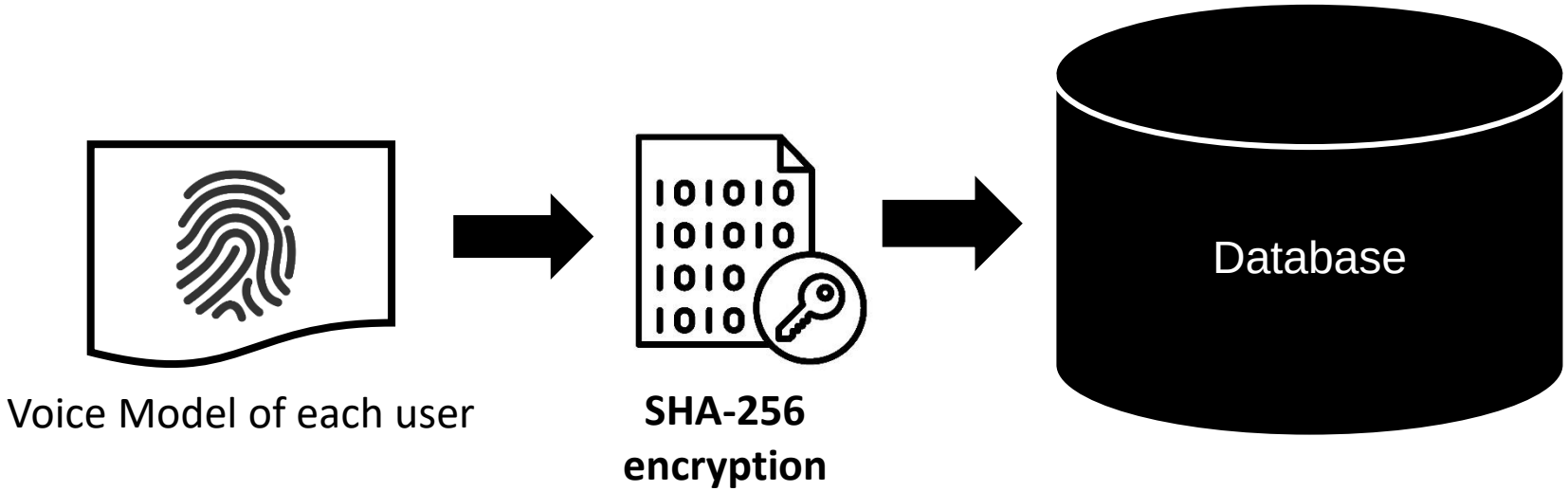
As audio signals are not uniform, there are multiple peaks in one audio signal. Intuition behind choosing GMM for our approach is *representing an audio signal as multiple gaussian distributions and generating an envelope of the waveform ,* the envelope will be generated for each user using GMM technique and stored as voice model in the database.

User Model Storage

Users’ unique Voiceprints are stored after modelling here.

For ensuring security of voiceprint, each voice model is stored in encrypted form in the database, which will be decrypted in Authentication Phase.

We are using **SHA-256 encryption** mechanism as it doesn't have any known vulnerabilities that make it insecure and it has not been broken.



Response: Similarity Score

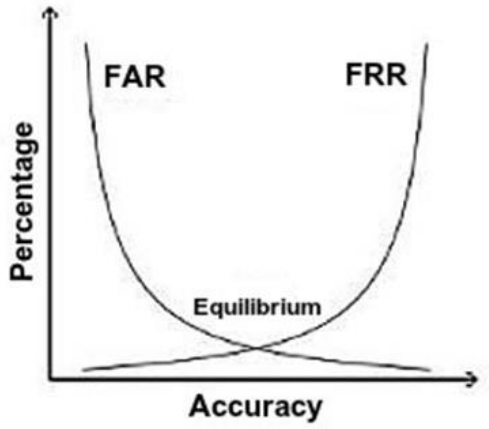
We generate a **percentage score** to express the similarity between the voice sample being verified and existing biometric template of the user.

This is done by computing '***log-likelihood'* that the features of the sample clip lie under the model of the user being verified.**

The higher the score is, the higher is the similarity between them.

Access to the system should be granted only **if the score is higher than a certain threshold.**

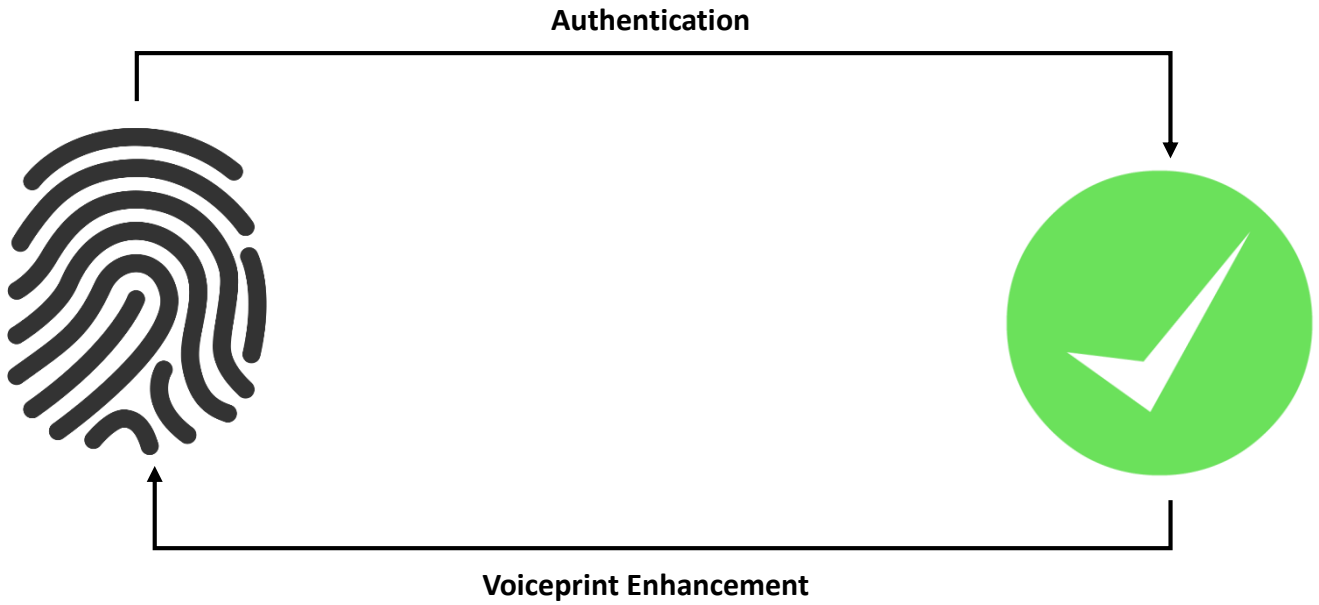
We will be choosing a threshold level such that the
False Acceptance Rate (FAR)
and False Rejection Rate (FRR) are both minimum:



Note: Threshold can be adjusted according to bank's policies.

Training: Enhancing Voiceprint

In the authentication phase if the voice sample being verified is accepted, the feature vectors of the this new sample are again used to **train the existing model of the user to enhance the Voiceprint of the user**.
This means, that the accuracy should keep on increasing.



Future Work/ Enhancements

Use of Deep Learning by fitting a Convolutional Neural Network to spectrograms of short audio segments of speaker utterances to perform end-to-end feature extraction and speaker recognition.

Use of Blockchain-based Biometric Storage to enhance security of biometric information through blockchain-based distributed management. The template can be fragmented and stored on different blockchain nodes so that there is no single point of vulnerability.

Thank You

Version: 2.00

Team: Ingenious

Members:

Name:	Shivani Tyagi	Divya Gupta
Designation:	Asst. Manager (Systems)	Asst. Manager (Systems)
Company:	SBI – GITC	SBI – GITC
Mob:	9810807077	8360670689
Email:	st.tyagishivani@gmail.com	diyyaguptauiet@gmail.com