HACKATHON 2021

Presentation by Runtime Terror

12 Nov 2021

AGENDA

01 Problem Statement

02 Idea/Objective

03 How the solution works

04 Tools/Libraries used in the project

05 Plans for future enhancement

Problem Statement

An AI driven solution for user authentication based on voice password and facial recognition. The solution should be two parts:

1. The first part where the user is prompted to enter his/her unique id and then pronounce a pre-defined pass phrase.

2. The second part where the system matches the user's voice pattern and facial image with previously saved data, and accordingly authenticate or reject the user.

Idea/Objective

We propose a Web Based solution where the user can enroll the biometrics i.e., the face and the voice.

It is a 2-Step-Authentication process where the 1st step is the voice authentication, and the 2nd step is the facial recognition mechanism.

Fig 1. High Level Objective

How the solution works

- ✓ We will be using deep learning, to improve the recognition rate of Automatic speech recognition (ASR) systems and map the speech signal into its corresponding text. It is a simplified model-building process and abilities to directly map speech into the text without any predefined alignments.
- ✓ In our project, we will exploring many different models such as CNN-based model for raw speech signal.

Generic framework of automatic speech recognition system:

Preprocessing:

- **Solving the Cocktail Party Problem** : Removing background noise using Fourier analysis.

Feature Extraction :

- Using features from **Mel Frequency Cepstral Coefficient (MFCC) and MEL Spectrogram**

Fig 2. Speaker Recognition Training

How the solution works

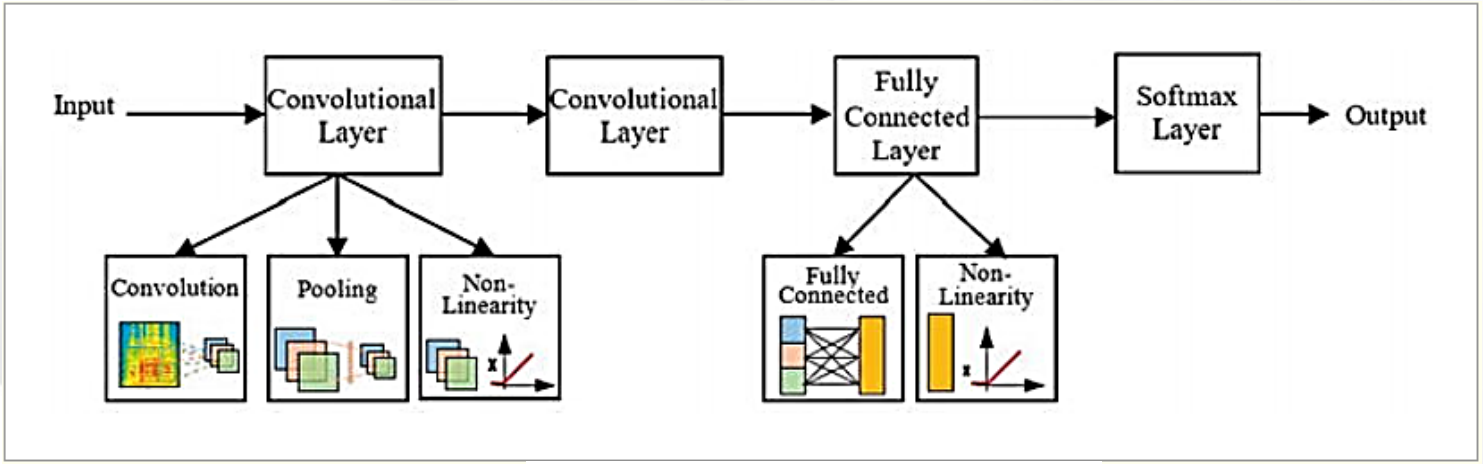


Fig 3. Block diagram of convolutional neural network

CNN Approach:

- Relevant features and classifier both will jointly be learned from the raw speech.
- Raw speech will be processed by first convolutional layer (in CNN model) to learn the feature representation.
- The output of first convolutional layer (in CNN model) , that is, intermediate representation, will be more discriminative and further processed by rest convolutional layers.
- This system will use only few parameters and perform better than traditional cepstral feature-based systems.
- The performance of the system will be evaluated for TIMIT and claimed similar performance as MFCC.

How the solution works – Enrollment Process

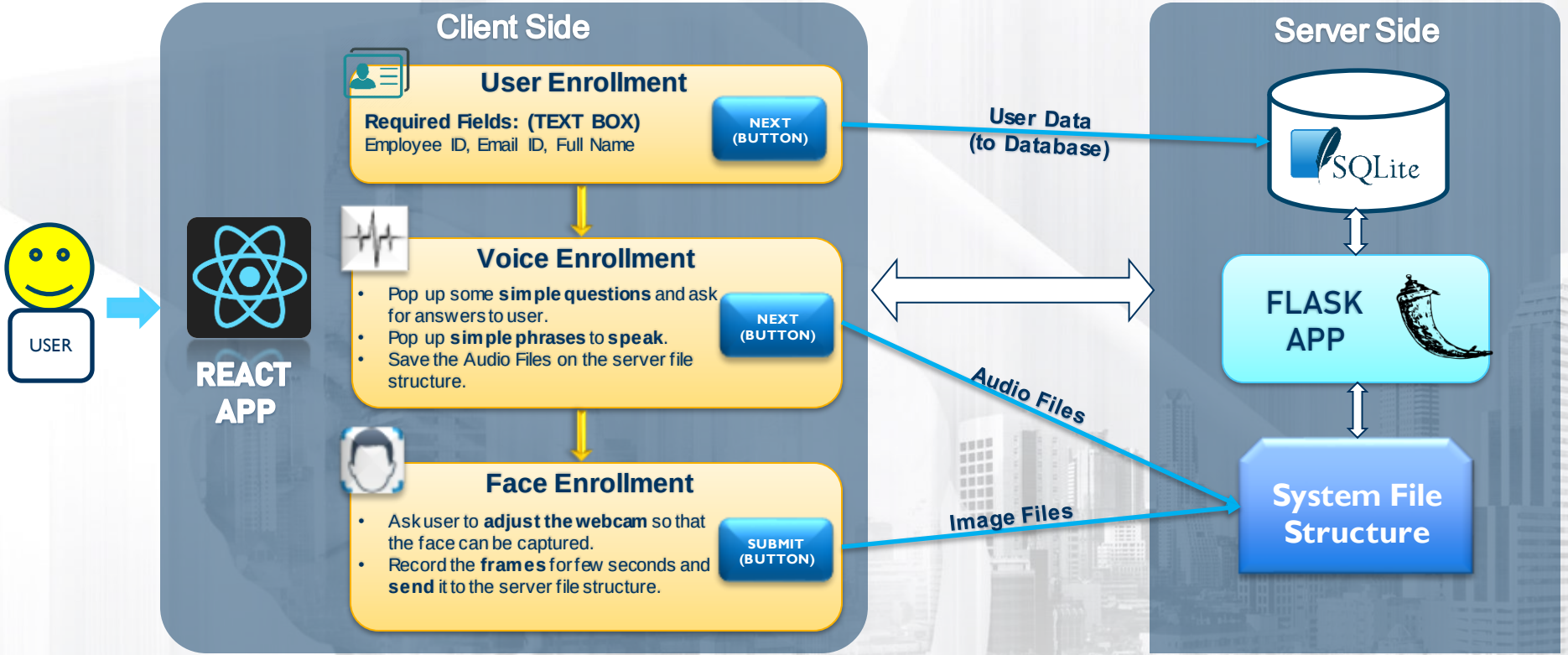


Fig 4. Enrollment Process Flow Diagram

How the solution works – Authentication Process

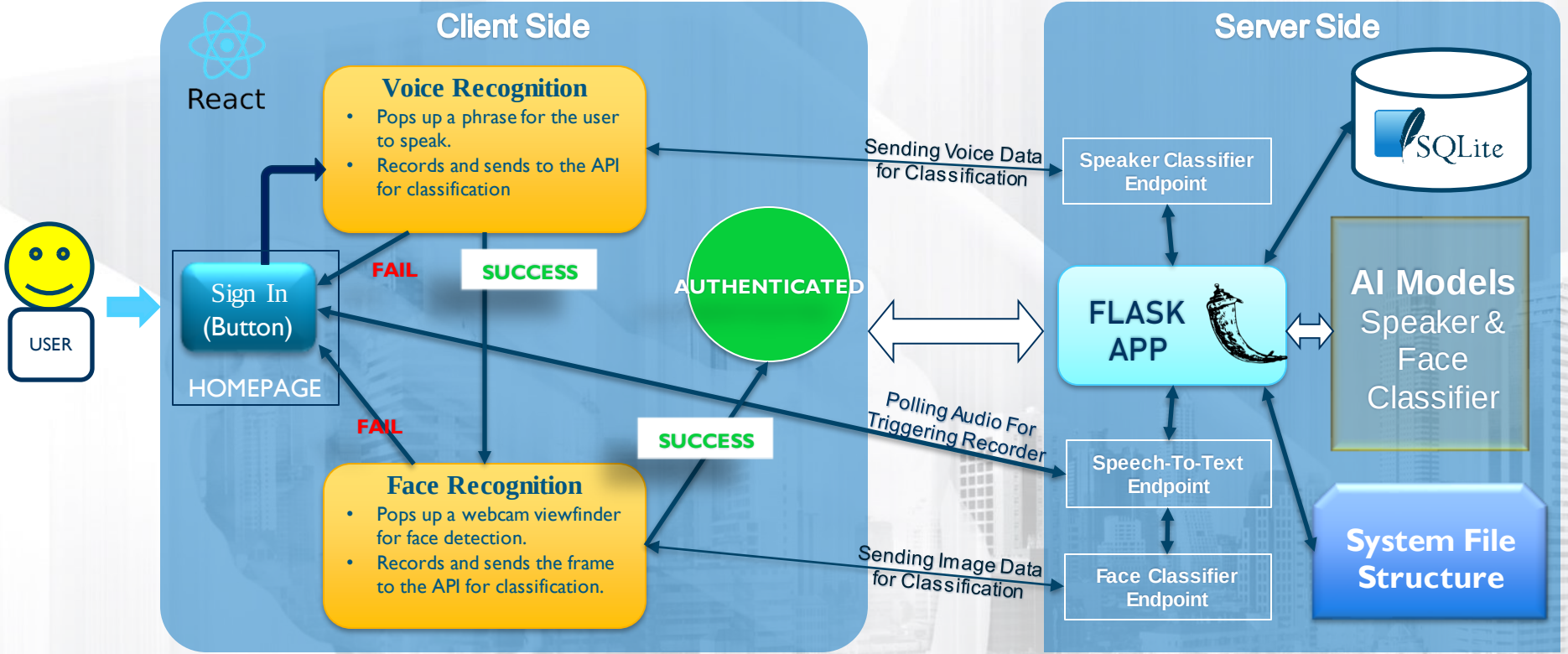
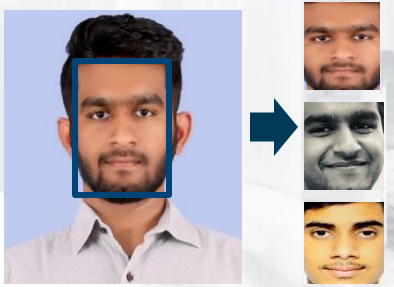


Fig 5. Authentication Process Flow Diagram

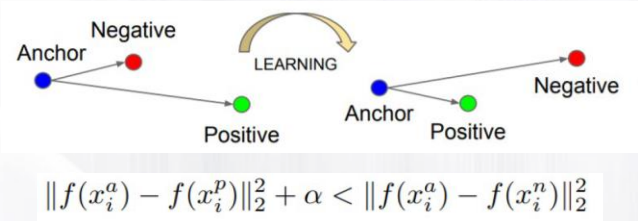
How the solution works – Face Recognition (FR)

The model



Detection via MTCNN

160x160 px



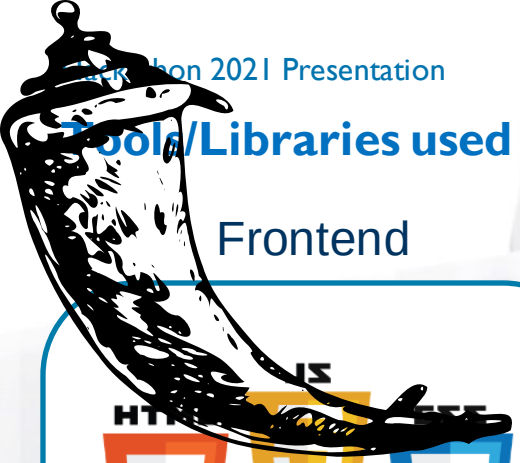
$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2$$

f(x)

Fig 6. Face Recognition Training

Actual deployment

Tools/Libraries used in the project



Frontend

Backend

Flask

React

Bootstrap

API

NumPy

pandas

TensorFlow

Keras

python

librosa

SQLite

scikit learn

jupyter

Fig 7. Technology Stack

HARMAN
A SAMSUNG COMPANY

Plans for future enhancement

- We can explore other ASR algorithms to improve accuracy
- Multilingual support for Speaker recognition
- Improve the training algorithm used in deep learning for face recognition
- Improvement of face recognition under different environmental conditions such as dim/bright light, different eyes (Korean, Japanese)
- Caching access information while the server is down.
- Improvements in Speech-To-Text such as semantic corrections, etc.
- Try out different approaches to bring down the cost such as server less deployment,etc.