Voice Print Report

Project idea and overview:

The main idea of the project is to enter your voice and the system will whether recognize your voice depending on the accuracy of the algorithm applied or not.

Algorithm:

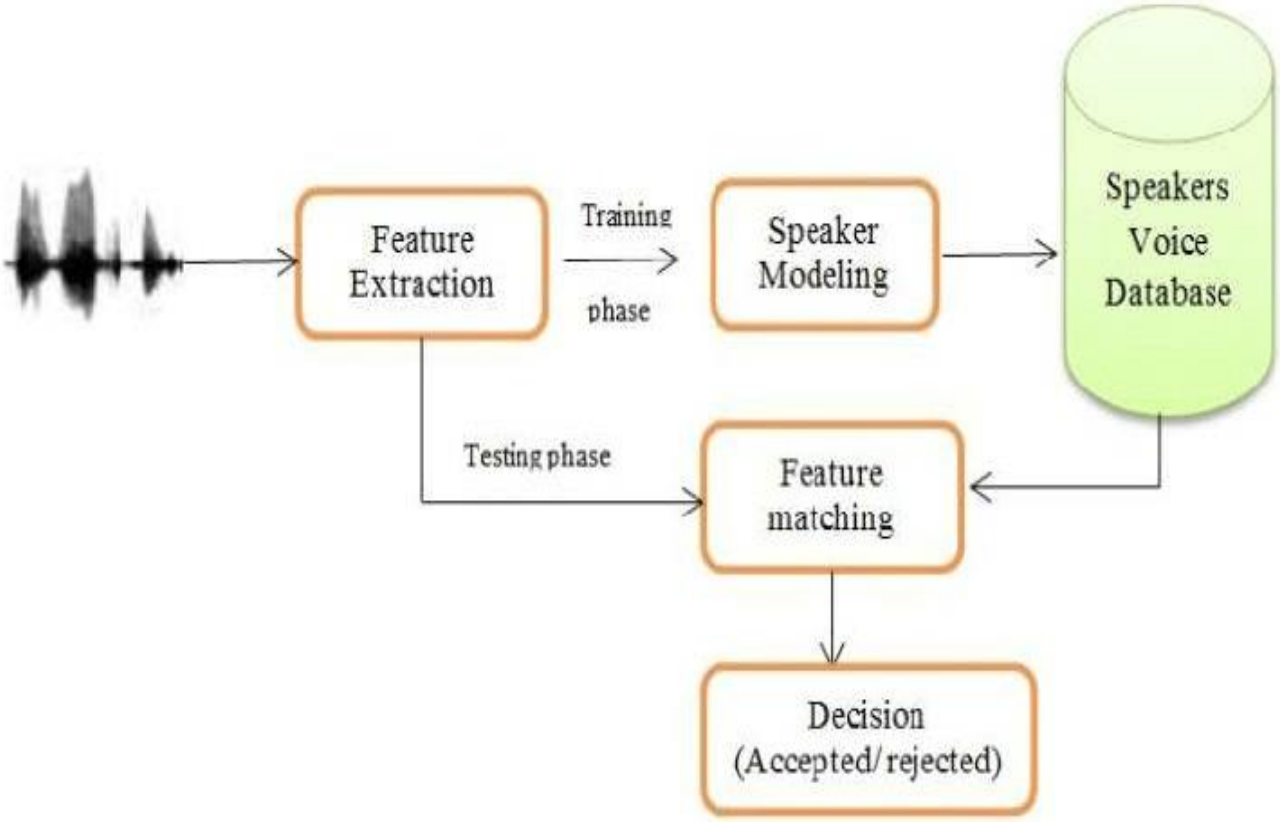
- MFCC

In sound processing, the **Mel-frequency cestrum** (**MFC**) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency.

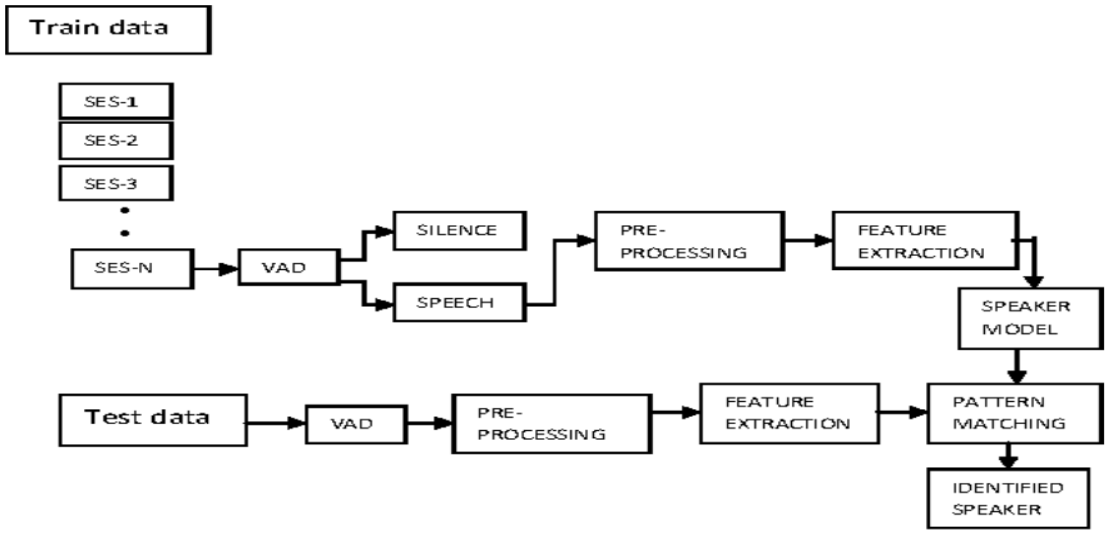
Mel-frequency cepstral coefficients (**MFCCs**) are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). The difference between the cestrum and the Mel-frequency cestrum is that in the MFC, the frequency bands are equally spaced on the Mel scale, which approximates the human auditory system's response more closely than the linearly spaced frequency bands used in the normal spectrum. This frequency warping can allow for better representation of sound, for example, in audio compression.

MFCCs are commonly derived as follows:

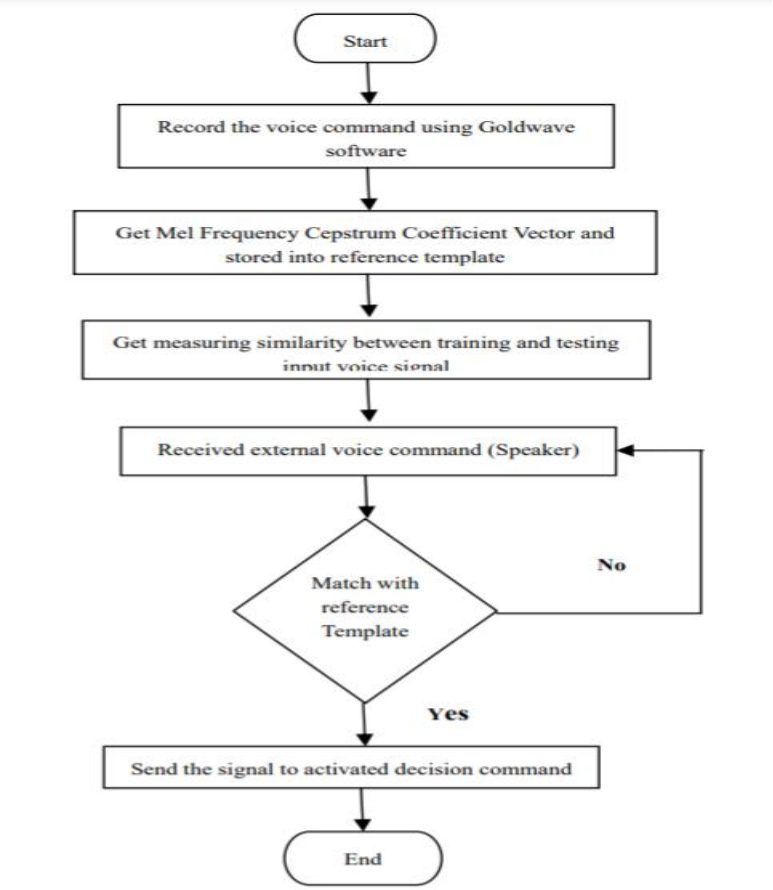
1. Take the Fourier transform of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained above onto the Mel scale, using Triangular overlapping or alternatively.
3. Take the logs of the powers at each of the Mel frequencies.
4. Take the discrete cosine transform of the list of Mel log powers, as if it were a signal.
5. The MFCCs are the amplitudes of the resulting spectrum.



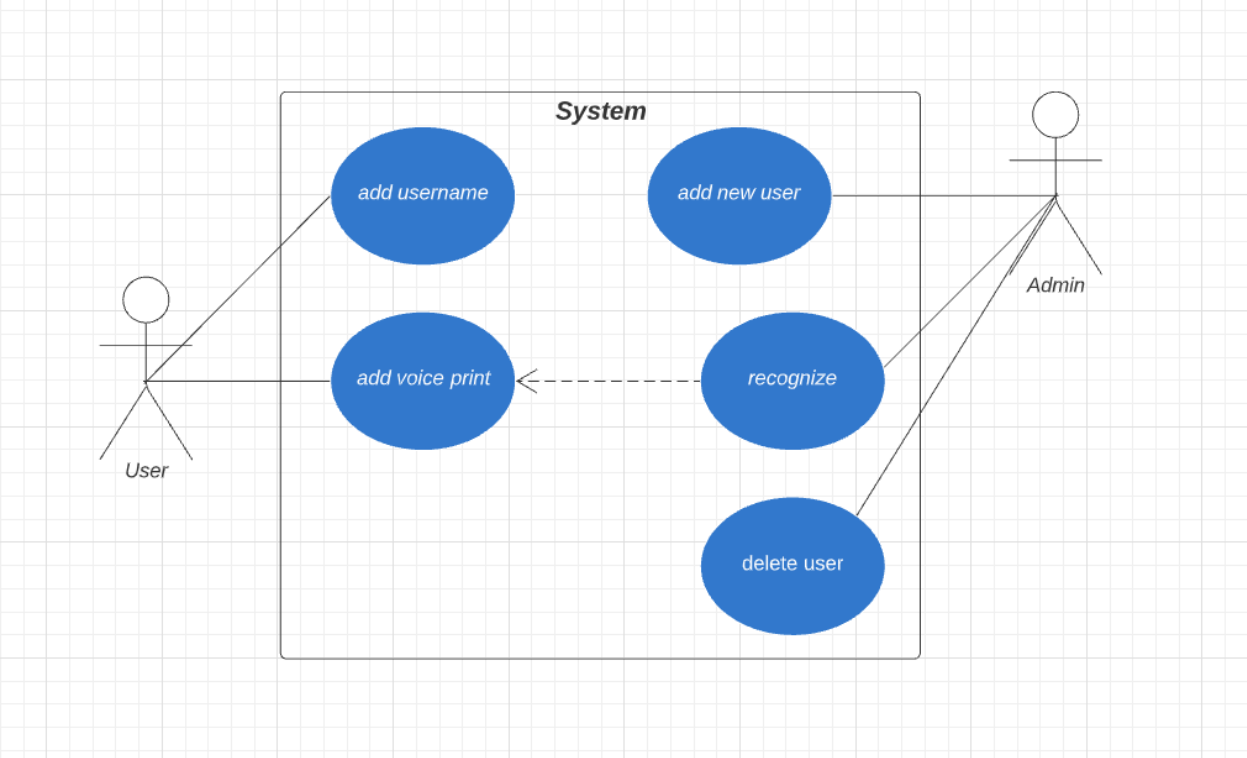
Data Testing and Training Sequence:



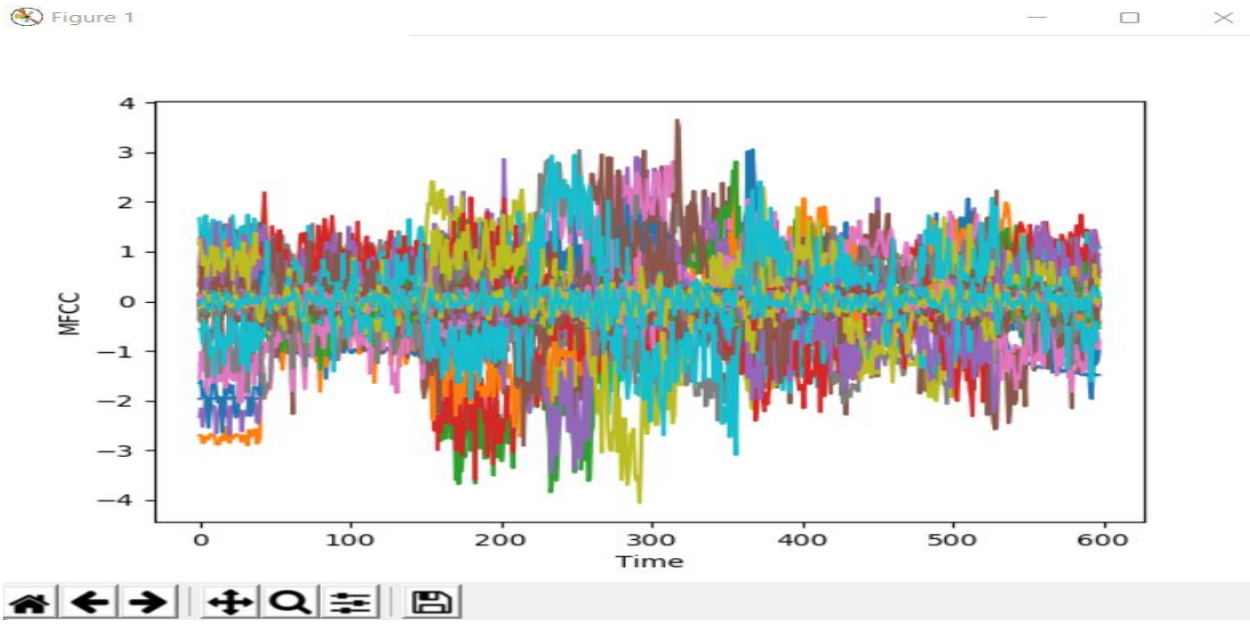
Flow chart Diagram: to describe how the algorithm work in sequences



Use Case Diagram: to describe what privileges could either user or admin do.



MFCC Sample Output:



Insights:

State of Voiceprint Technology Today While voiceprint is still maturing, its adoption is growing every day. Recently, one of Australia's largest banks enrolled 120,000 of its customers in a voiceprint-based authentication service. The Australian Taxation Office has registered the voiceprints of over 15% of the country's citizens. There are still challenges to overcome, including accuracy, voiceprint algorithm training data, privacy, and potential discrimination. However, it is a promising area of research, with advancements by companies such as Huawei, Google, etc. to push the envelope in this space.

Advantages:

- The voice biometrics market is growing at an impressive pace, expected to reach $4.9 billion by 2027, compared to less than $1 billion in 2019. This growth is propelled by the following benefits: A person's voiceprint is extremely hard to spoof – In an identity theft case, a fraudster can get hold of a customer's date of birth, address, and unique information like their mother's maiden name or the name of their first pet. However, every individual has a distinct voiceprint, which is far harder to obtain or mimic Its convenience helps the quality of CX.

- Voice-biometrics-based authentication is more convenient for the customer than physically entering a password, remembering answers to secret questions, etc. There is no dependence on memory or recall, as your voice itself acts as the identifier It is an excellent candidate for automation.

- Voice biometrics can be used to automatically obtain a person's approval after verifying their identity. For instance, a live agent doesn't need to ask a customer for consent before recording a call – an automated IVR can request, while voice biometrics verifies that it was indeed the customer who provided their consent as a result, there is a lot of interest and investment around voice biometrics, with nearly every technology company opening a patent in this field.

Disadvantages:

- one should remember that voice biometrics isn't a mature technology. It relies on sophisticated AI algorithms, trained on comprehensive datasets, and tested in real-world scenarios – which can be difficult to achieve. This

can lead to the following pitfalls in biometrics implementation: There is a risk of discrimination and racial bias.

- As mentioned, voice biometrics algorithms (especially those used for identification and not verification), must be trained on comprehensive datasets, comprising a diverse range of human voices. But studies suggest that the dataset used for training most mainstream voice technologies might be racially skewed, which makes the AI better at recognizing some demographics than other Companies must be extremely sensitive about privacy and consent.

- Voice is innately personal and not every customer will be comfortable with sharing their voice data. Enforcing data privacy laws like GDPR can be problematic, as customer voice samples are relatively easy to collect. Already, the Chinese government has come under scrutiny for potentially breaching privacy rights and collecting voice pattern samples to build a national database Voice deep fakes are possible – Finally, audio deep fakes are becoming increasingly common and may be able to fool the Ai into believing the audio sample's veracity.

Development platform:

- Tools: Visual Studio Code

- Programming Languages: Python

- Libs: absl-py,astor,backcall,decorator,gast,google-pasta,grpcio,h5py,ipython,ipython-genutils,jedi,Keras,Keras-Applications,Keras-Preprocessing,Markdown,numpy,opencv-python,parso,pexpect,pickleshare,prompt-toolkit,protobuf,ptyprocess,PyAudio,Pygments,python-speech-features,PyYAML,scikit-learn,Scipy,six,tensorboard,tensorflow,tensorflow-estimator,termcolor,traitlets,Wave,wcwidth,Werkzeug,wrapt.

Applications are like the voice print identity:

- Siri from Apple.

How Siri Works: Upon receiving your request, Siri records the frequencies and sound waves from your voice and translates them into a code. Siri then breaks down the code to identify patterns, phrases, and keywords. This data gets input into an algorithm that sifts through thousands of combinations of sentences to determine what the inputted phrase means. This algorithm is complex enough that it can work around idioms, homophones, and other literary expressions to determine the context of a sentence. Once Siri determines its request, it begins to assess what tasks needs to be carried out, determining whether the information needed can be accessed from within the phone's data banks or from online servers. Siri is then able to craft complete and cohesive sentences relevant to the type of question or command requested.

Future modifications of MFCC:

A pitch normalization algorithm is proposed for addressing the pitch mismatch between adults' and children's speech for children's automatic speech recognition (ASR). Motivated by the appearance of pitch-dependent distortions in the smoothed Mel spectral envelope for high-pitched children's speech, the algorithm modifies the multitrack during MFCC feature extraction to improve AS performance. Relative improvements of 16 % and 9 % are obtained over the corresponding baseline in children's mismatched ASR performance on a connected-digit recognition task and a continuous speech recognition task. The improvements obtained in ASR performance with the proposed pitch normalization algorithm are also found tube additive to that obtained with existing speaker normalization techniques, VTLN and CMLLR.

Resources:

[1] A. Anjos et al. "Bob: a free signal processing and machine learning toolbox for researchers". In: *20th*
ACM Conference on Multimedia Systems (ACMMM), Nara, Japan. ACM Press, Oct. 2012. url:
http://publications.idiap.ch/downloads/papers/2012/Anjos_Bob_ACMMM12.pdf.
Gaussian mixture and hidden Markov models". In: *International Computer Science Institute* 4.510 (1998),

p. 126.
[2] George E Dahl et al. “Context-dependent pre-trained deep neural networks for large-vocabulary speech
recognition”. In: *Audio, Speech, and Language Processing, IEEE Transactions on* 20.1 (2012), pp. 30–42.
[3] John Godfrey, David Graff, and Alvin Martin. “Public databases for speaker recognition and verification”.
In: *Automatic Speaker Recognition, Identification and Verification*. 1994.
[4] Patrick Kenny. “Joint factor analysis of speaker and session variability: Theory and algorithms”.
[5]A Review on Noisy Password, Voiceprint Biometric and One-Time-Password
url: https://www.sciencedirect.com/science/article/pii/S1877050916000806.