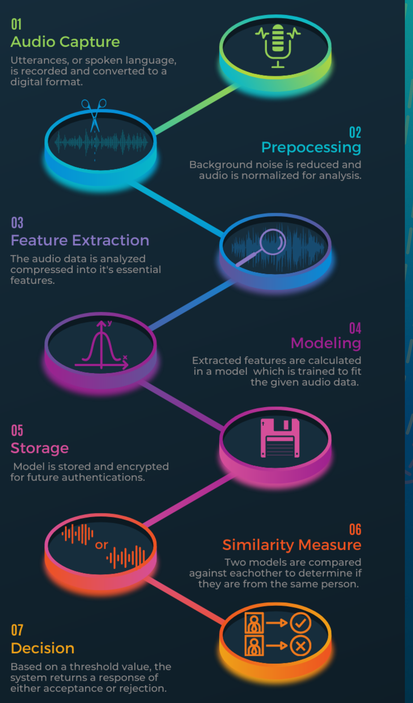
Minimizing Language Dependency in a Voice Authentication System

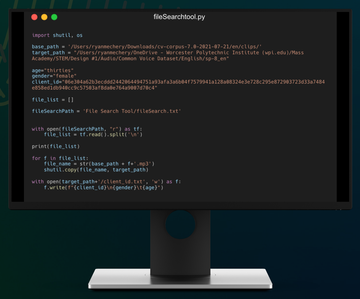
Ryan Mechery
Advisor: Dr. Kevin Crowthers, Ph.D.

Graphical Abstract

Pipeline for Voice Authentication

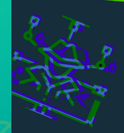


Code Sample

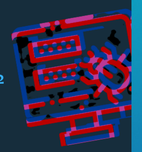


This image above shows a portion of the novel code developed in this project. Python was used in VS Code with Librosa and Sk-Learn serving as the main dependencies.

Background



Just 20 passwords constitute 10.3% of all passwords in use. (1) 01



Two-thirds of people around the world only use **two** passwords for all their accounts. (2) 02

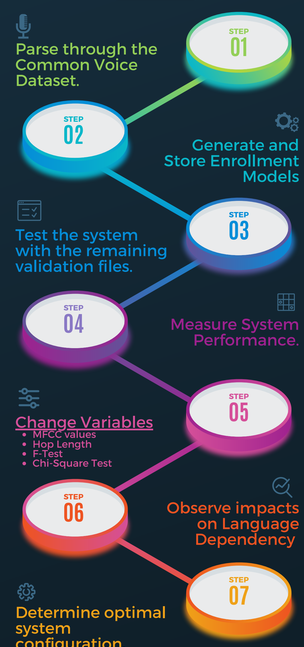
A majority of laptops/tablets around the world do not have any biometric authentication capabilities. 03

Sources

1) https://www.skyhighnetworks.com/cloud-security-blog/you-wont-believe-the-20-most-popular-cloud-service-passwords/
2) https://www.halock.com/passwords-fascinating-facts/

Materials

Text-Independent Speaker Verification System Testing Methodology



Problem Statement

Text Independent Speaker Verification (TISV) systems offer variable accuracies when tested with different world languages due to language dependency.

Engineering Goal

The goal of this project is to engineer a speaker verification system and test it with different world languages. Through analysis of system performance, the aim is to identify hyperparameters that increase the accuracy of the system overall to minimize language dependency.

Results

	Average ROC	Phoneme Count
Mean	0.741186978	37.2
Variance	0.003434493	31.7
Observations	5	5
Pearson Correlation	-0.513933426	
Hypothesized Mean Difference	0	
df	4	
t Stat	-14.40203051	
P(T<=t) one-tail	6.75453E-05	
t Critical one-tail	2.131846786	
P(T<=t) two-tail	**0.000135091**	
t Critical two-tail	2.776445105	

Table 1: Unpaired T test Between Average ROC of the System and Phoneme Count of each Language.

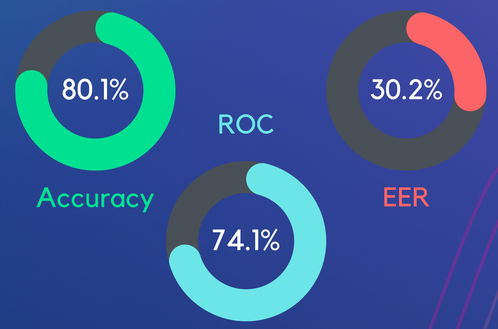
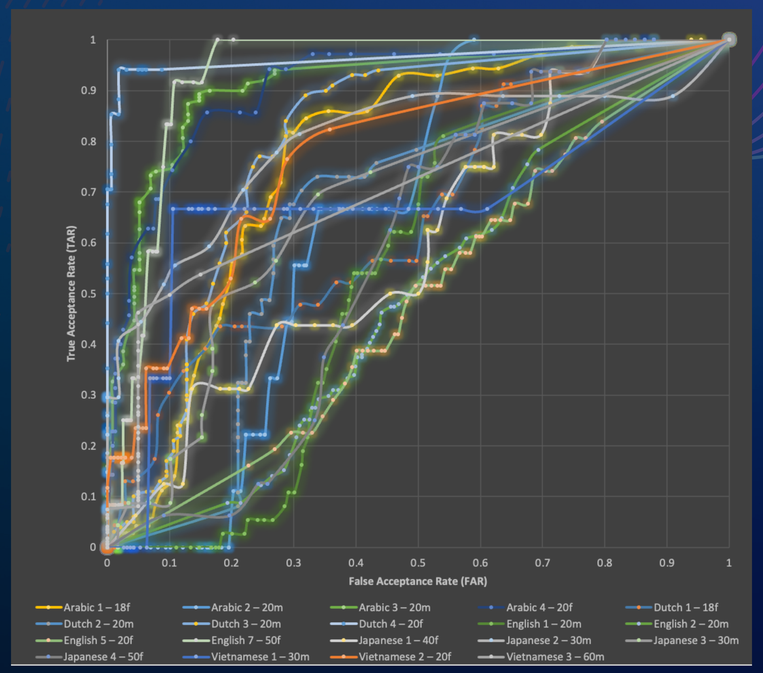


Figure 1: Accuracy (left), Equal Error Rate (Right), and ROC (bottom) of the final system.

By adjusting hyperparameters to minimize language dependency, Voice Authentication can become the world's next universal method of biometric authentication.



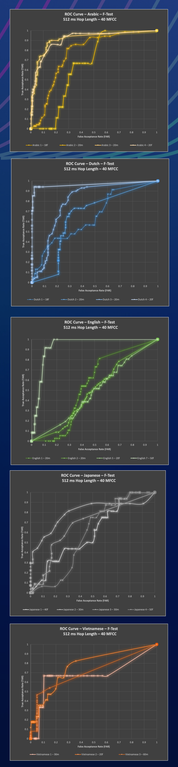


Figure 2: Receiver Operator Characteristics (ROC) Curve Showing System Performance of from every Speaker in 5 languages.
Hyperparameters: 512 ms Hop Length – 40 MFCC – F-Test.

Decision Matrix

	Fingerprints	Hand Geometry	Retina	Iris	Face	Signature	Voice
Ease of use	High (+3)	High (+3)	Low (+1)	Medium (+2)	Medium (+2)	High (+3)	High (+3)
Accuracy	High (+3)	High (+3)	Very High (+4)	Very High (+4)	High (+3)	High (+3)	High (+3)
Cost of Req. Tech	Sensor (+1)	Sensor (+1)	Sensor (+1)	Sensor (+1)	Camera (+2)	Camera (+2)	Mic (+3)
Total	7	7	6	7	7	8	9

Analysis

- A two-sample unpaired t-test showed a strong significance (***p<0.001) between the phoneme count and the system's performance with each language.
- The system's peak performance was with an ROC of 74.124% which translated into an authentication accuracy of **80.100%** an Equal Error Rate (EER) of **30.200%**.

Discussion

- Although the system performed with an acceptable ROC of 74.124% for a binary classification system, for an authentication system, it performed with underwhelming authentication and error rates.
- Results from preliminary trials show that adjusting hyperparameters related to audio processing was largely ineffective.
- Further testing revealed that using F-Test as compared to the Likelihood Ratio Test in nested model detection improved the overall ROC performance by 1.96% which then minimized language dependency.
  - Using an F-Test for nested model detection in TISV-GMM systems has not been written about in academia for this purpose.
- The system itself is not complete, the threshold testing with the ROC curves shows inconsistent threshold values and as such, this system needs more refinement before it is fully functioning.

Conclusion & Extensions

- Although statistical modeling for TISV was found to be ineffective, if this experiment were to be extended, it would use Neural Network or I-Vector Based methods for Speaker Verification for better performance.
  - This system would still be tested in exactly the same way with the same recordings to minimize language dependency.
- Additionally, using recordings from the same microphone would benefit hyperparameter testing by eliminating the possibility that unexpected system performance was due to microphone quality.
- The current command line application could be converted into a GUI application to demonstrate its ability as a new method of Multi-Factor Authentication (MFA).