NITK-CMU WINTER SCHOOL 2014

VOICE FORENSICS

Abstract

The project aims to determine correlations between physical attributes of a person such as height, weight, age and sex to certain features in the voice of the sample. 2 sets of features were obtained for this purpose, 39 Mel frequency cepstral coefficients (MFCC), and 6000 features (speaker traits) using a software called OpenSmile. Different Classification, and Regression-based machine learning algorithms were applied to the above data, and their results were analyzed.

Objective

Crime is growing at an alarming rate, and we often find ourselves ill-equipped to handle the numerous cases. In instances of terrorist threat calls and false alarms, It is imperative to extract as much information possible from the available audio clippings. With the fast development of technology, it is time we start exploring other methods of identification, apart from fingerprints and images, which would make faster identification of criminals.

Voiceprint of a person is as unique as a fingerprint. There has been little exploration in the area of audio forensics, although we have advanced methods to process voice.

- ✓ Our aim is to build a voiceprint database of the citizens of our country. Law enforcement agencies can make use of this technology to identify criminals based on their voice samples.
- ✓ Producing reliable estimates of physical features and demographics from the voice sample even if the person's voiceprint is unavailable in our database.

Miscreant

Citizen ‘X’

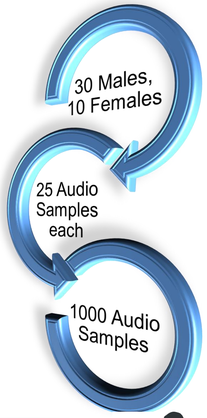
Retrieving Voice recordings

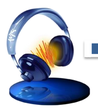
Methodology

Audacity®

Recording/Editing Software

1000 Audio Samples were collected from 40 students of the NITK-CMU Winter School. The students were taken to a quiet room and recordings were made using Audacity. Each person were given a set of phonetically rich sentences from the TIMIT database.



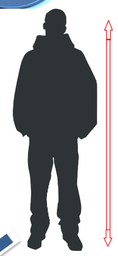


openSMILE: by audEERING™

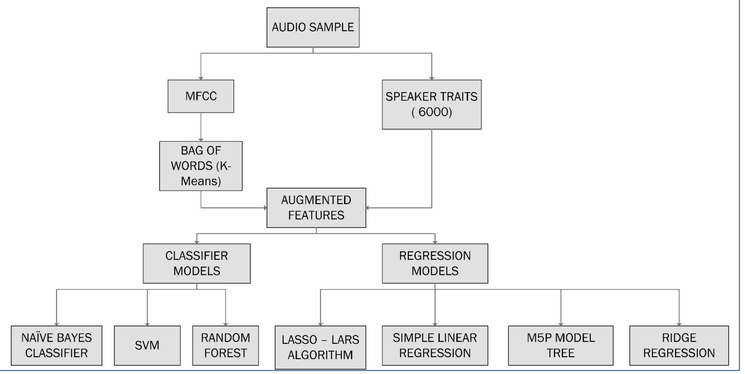
Feature Extraction Software

Machine Learning Algorithm

Reconstructing the Miscreant

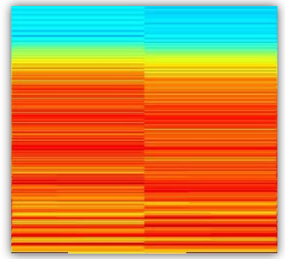


Pipeline

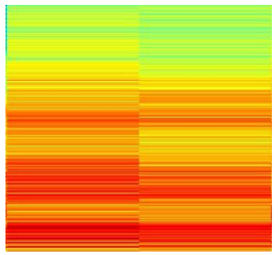


Frequency Spectrogram

X axis: Time, Y axis: Frequency



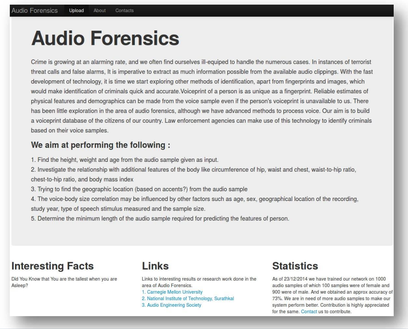
Female Frequency Spectrogram

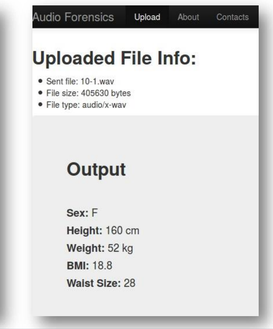


Male Frequency Spectrogram

It can be seen from the spectrogram that the pitch of the female voice is higher than the pitch of the male voice

Website





Results

	Random Forest	Naïve Bayes	SVM chi-square	LASSO	Ridge
Gender 6000	P: 93.3% R: 87.8%	P: 85% R: 86.9%	-	-	-
Gender 39	P: 97.3% R: 65.5%	P: 95.9% R: 95.7%	P:25% R: 20%	-	-
Height 6000	P: 87.5% R: 87.1%	P: 78.2% R: 74.7%	-	Algorithm to be modified	-
Height 39	P: 88.8% R: 67.7%	P: 56.8% R: 66.6%	P: 10.3% R: 12%	Algorithm to be modified	RAE: 88.5% MAE: 6.85

P : Precision, R: Recall

Team Members

Priya Soundararajan
Tejeswini Sundaram
Utkarsh Patange
Sakthivel Sivaraman

References

1. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.
2. Florian Eyben, Felix Weninger, Florian Gross, Björn Schuller: "
3.http://www.csie.ntu.edu.tw/~cjlin/libsvm/Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor", In Proc. ACM Multimedia (MM), Barcelona, Spain, ACM, ISBN 978-1-4503-2404-5, pp. 835-838, October 2013. doi:10.1145/2502081.2502224