Uniquely Determining Identity Using Computer-based Analysis Of Human Speech

Alan Chen[1], Robert Cheung[2]
Henry M. Gunn High School[1], Fin[2]

INTRODUCTION

In this research, we are exploring the area of voice recognition, the identification of a person from the characteristics of his or her voice. Voice recognition in this research project is distinguished from speech recognition in that it is the identification of who is speaking rather than what is speaking. This topic of voice recognition/speaker recognition has been studied for about four and has employed the acoustic features of speech that are found to be different between individuals.

MATERIALS & METHODS

1. We first reached out to volunteers interested in participating in our research and collected voice samples of them reading a predetermined block of text (same for each individual).
   - These voice samples were divided into three parts:
   - The first part, we woul

1. These voice samples were divided into three parts:
   - The first part, we would use to enroll the individual, which would allow our computer to build its model of the speaker's voice.
   - The next part, we would use to further train our model
   - Third part would be used as test data that is not used for building the model so as to make sure identification based on the model truly works.

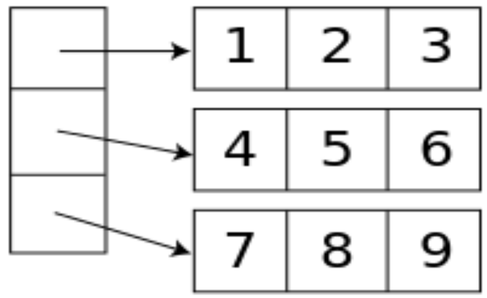
1. After we collected the voice samples, we "enrolled" the individual into our research
   - Extracted the audio features of each individual's voice samples using YAAFE
   - Chose to extract a total of five audio features for analysis, extracted into numerical CSV files

Project Materials are listed below:

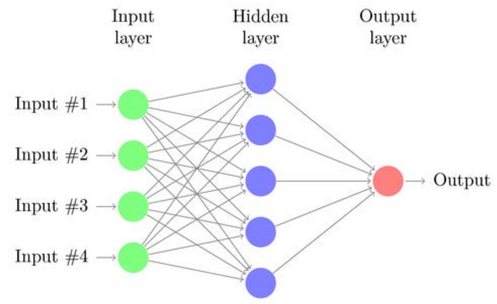
- Voice samples from online sources
- Computer
- Tensorflow (https://www.tensorflow.org/)
- YAAFE Audio Extraction library (http://yaafe.sourceforge.net/)

MATERIALS & METHODS

4. We then interpreted the data
   - Put the contents of the CSV files into one multidimensional vector data structure
     - each of the audio features took up one vector of values in the multidimensional vectors.



5. Finally, we analyzed the data
   - Used Tensorflow functions to implement Artificial Neural Networks (ANN)
   - Create models for each speaker based solely on the numerical values that we extracted



SUMMARY / CONCLUSIONS

- We were not successful in our goal of uniquely determining a speaker's identity using machine learning algorithms on the audio features we extracted from the data
- We did not have the time necessary to program the parts of the project necessary to analyze the test data after extracting the audio files and producing models for the audio feature data for each individual

DISCUSSION/ FUTURE STEPS

- There were many audio features that we extracted that were not effectively related to our goal of creating a model to identify speakers. Thus, many of these features were eliminated from consideration in our machine learning algorithm.
- Audio features tested in the end
  - AmplitudeModulation
  - Loudness
  - PerceptualSharpness
  - PerceptualSpread
  - SpectralFlatness
- Steps that can be taken to improve upon the existing research are as follows
  - Finish analyzing the test data with computer model and testing to see the accuracy of the model
  - Expanding the data pool to include more people and more vocal samples to compare and test
  - Refining the techniques of basic ANN by researching deeper learning algorithms and procedures

REFERENCES

- Machine Learning for Audio, Image and Video Analysis by Francesco Camastra and Alessandro Vinciarelli
- Google's Open Source machine learning library, Tensor Flow
- YAAFE Open Source Audio Extraction Library Documentation

ACKNOWLEDGEMENTS

A huge thank you to mentor Robert Cheung for his help and guidance throughout the duration of this research. Also, thanks to Dr. Choe for helping with organization and procedures of making this research project and for guidance throughout the year.