ENCM 509 Final Project

Speech Recognition using Gaussian Mixture Models

Jessica Ritchie 30071655

Jade Fjestad UCID

Table of Contents

[Introduction 3](#_Toc130641733)

[Objectives 3](#_Toc130641734)

[Methods 3](#_Toc130641735)

[Conclusion 3](#_Toc130641736)

[Appendix 3](#_Toc130641737)

# Introduction and Objectives

Speaker recognition is the process of matching an individual to a voice sample [1]. In contrast, speech recognition is the challenge of interpreting what a person is saying in an audio clip [1]. Most home smart systems are concerned with speech recognition that react to user input. For example, Amazon’s Alexa can respond to requests to play a certain song, recite your calendar details, or report the weather. In contrast, speaker recognition is mainly used for user authentication and verification. This can be used to replace the need for passwords on a phone or computer. Another application of speaker recognition is to incorporate it with speech recognition to improve the performance of these home smart systems. For example, speaker recognition would allow a member of a family to make a request, and the system would automatically recognize who was speaking and respond with the appropriate playlists, calendars, or emails.

Speaker recognition is also emerging in forensic science [1]. This technology can be used to identify people from their voices in recorded telephone calls for use in evidence [1]. Another application of speaker recognition is to improve live closed captioning. For example, in online conference calls, speaker recognition can identify how many different speakers are present and differentiate them to accurately label subtitles [2].

Speaker recognition is typically preformed with a Gaussian Mixture Model (GMM) [3]. A GMM is an unsupervised clustering method that determines the probability that a sample belongs to each class [4]. GMM’s are suitable for speaker recognition because they can handle vast amounts of numeric data with many features [4]. In addition, speech features are assumed to be normally distributed which permits simpler algorithms such as GMMs [5]. More recently, speaker recognition is moving towards neural networks as research in deep learning models has advanced. However, GMMs were long considered state of the art, and are sufficient for this project.

The objective of this report is to preform speaker recognition verification (Am I whom I claim to be?) using a GMM. The GMM should accurately identify if a test sample is someone known to the database or not. The model will be trained on one individual to resemble a verification system for phone access. The GMM will be used to calculate test scores that will be compared to a threshold. Samples that score above the threshold would be accepted as the real user, and samples that score below the threshold would be considered imposters and denied access.

# Methods

Diagram

Description automatically generated

Fig. 1 Project Flow Block Diagram

## Dataset

## Feature Extraction

## Splitting Data

## Training Model

# Results

illustrations, graphs (such as DET, ROC) and their analysis

# Future Work

# Conclusion

# References

# Appendix

Code