Voice recognition

Tom Rochette <tom.rochette@coreteks.org>

July 24, 2025 — daae079c

0.1 Context

0.2 Learned in this study

0.3 Things to explore

- MFC/MFCC
- Recognize speakers using per speaker models
  - Start with a single model for all speakers, and slowly figure out when a given speaker speaks, then retrain their individual model to become more and more accurate

0.4 To try

- Determine vocal tract size

1 Overview

The goal of this project is to recognize a person based on a record of his/her voice.

- Record audio
- Convert a certain window size (e.g., 20ms long) of the signal into the frequency domain using a fast Fourier transform

2 From the Deep Learning Book

- Factors of variation
  - Age
  - Sex
  - Accent
  - Words spoken

3 From the Bay Area Deep Learning School

- Application matters
  - Styles of speech
    - Read
    - Conversational
    - Spontaneous
    - Command/control
  - Issues
    - Disfluency/Stuttering
    - Noise

    * Microphone quality/Number of channels
    * Far field
    * Reverb/Echo
    * Lombard effect
    * Speaker accents

4 From PHO121 - Speech Analysis

- Vowels can be classified by their two first formants (F1 and F2)
  - Resonant frequencies that can roughly be associated with the size of specific cavities in the vocal tract
  - F1: Pharyngeal cavity
  - F2: Front cavity

5 See also

6 References

- https://www.youtube.com/watch?v=9dXiAecyJrY&feature=youtu.be&t=13874
- https://www.youtube.com/watch?v=MyNrmiJQ4dI