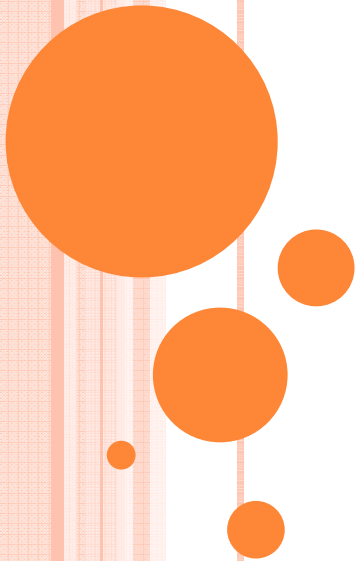


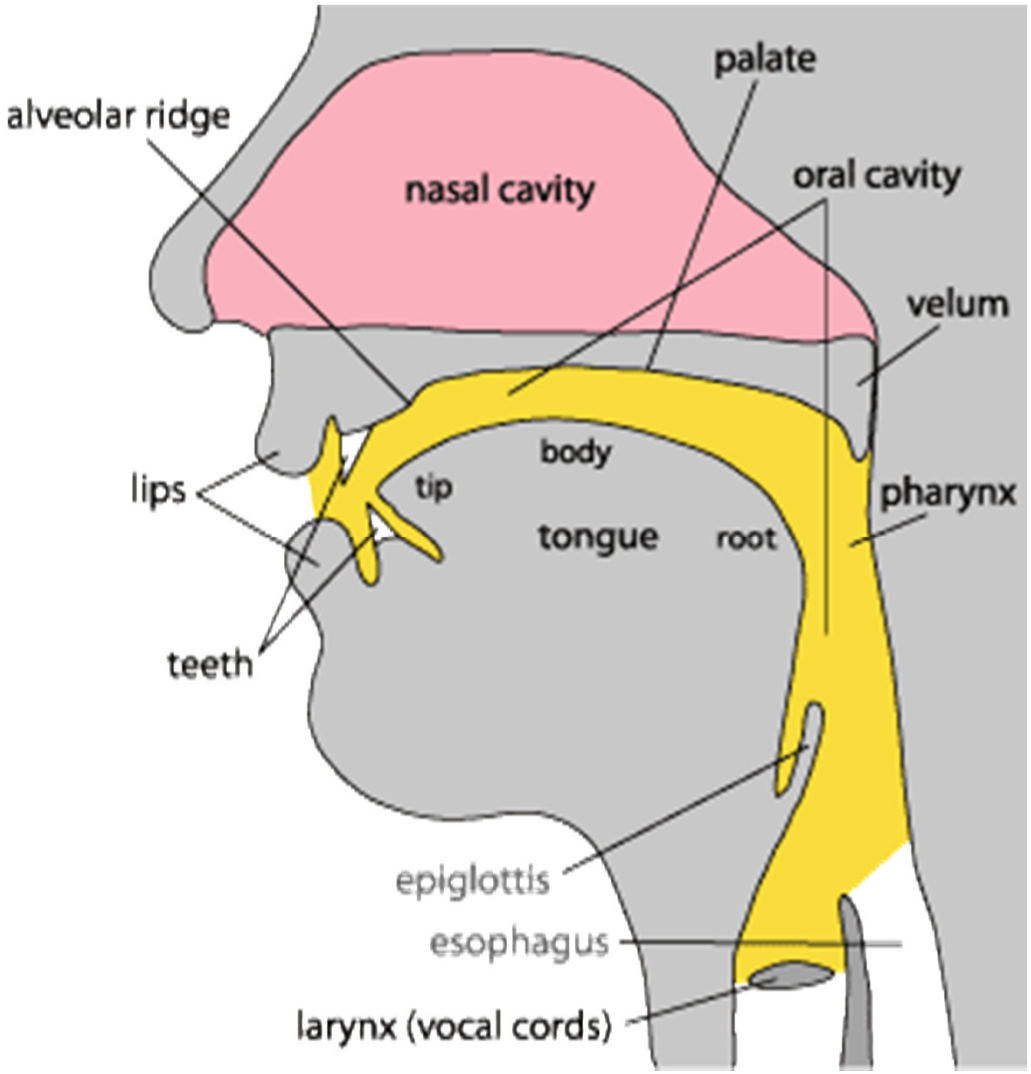
Speech Analysis

Jon Gudnason

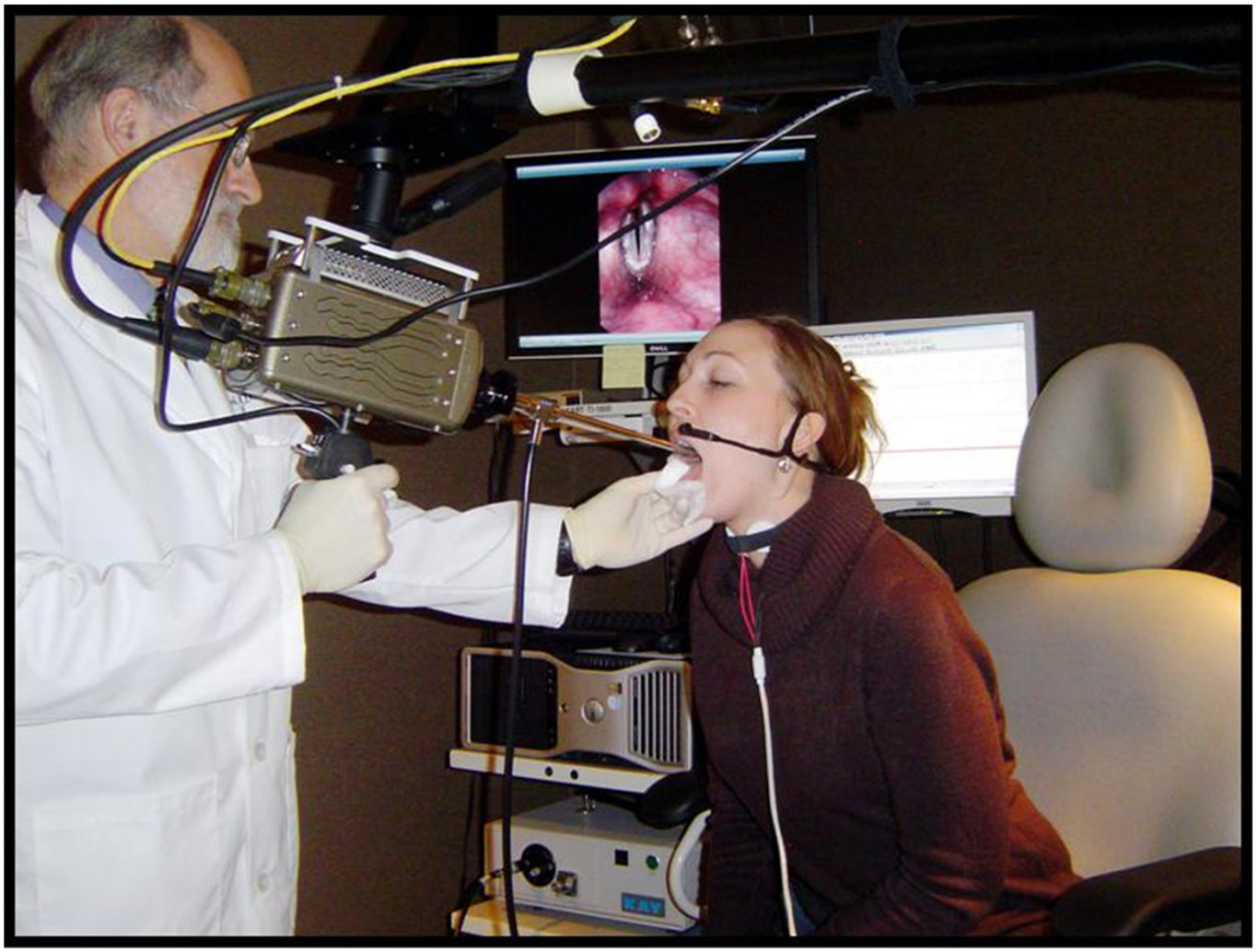
Overview

- Acoustic Speech Processing
- Speech Feature Extraction
- Speech Recognition with HMMs
- Hearing and Human Speech Recognition

Physiology of Speech

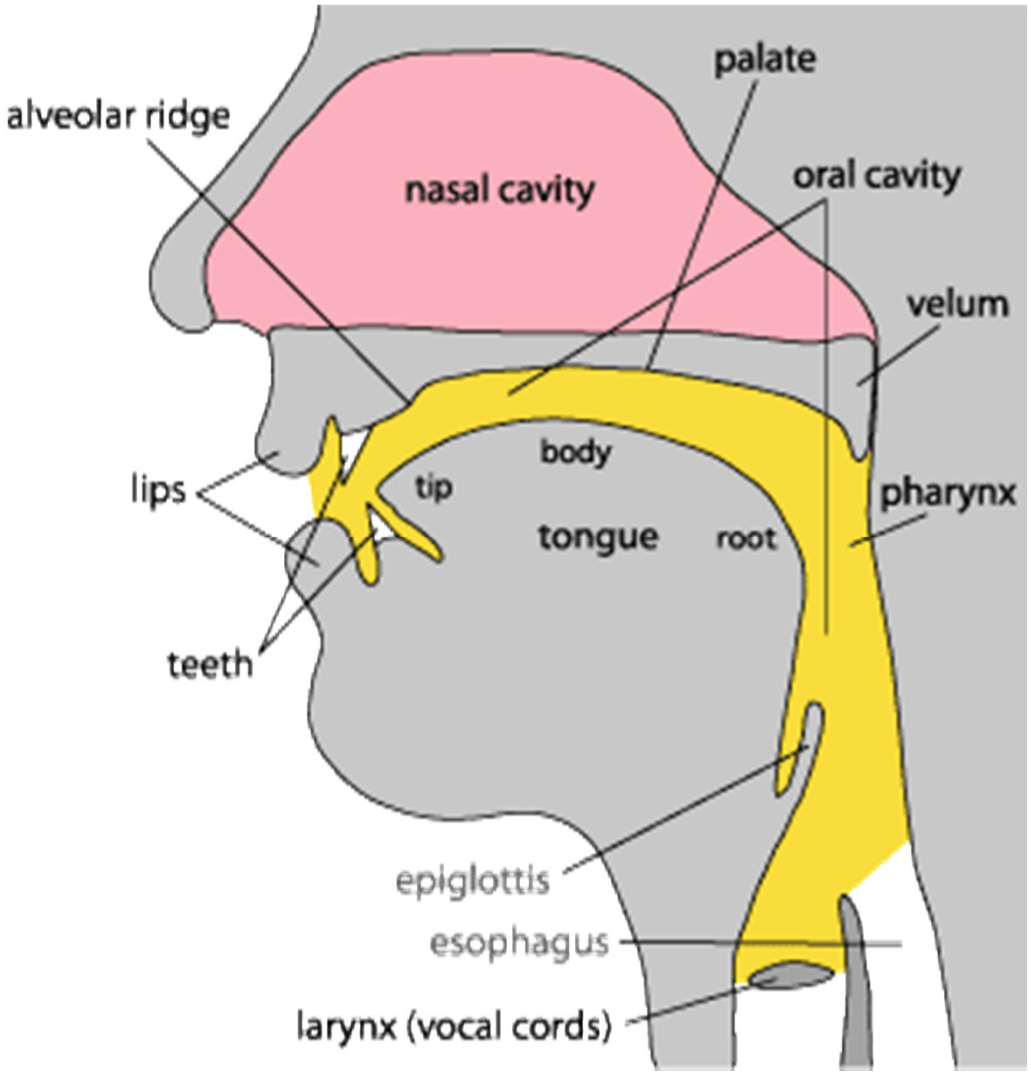


Video Endoscopy

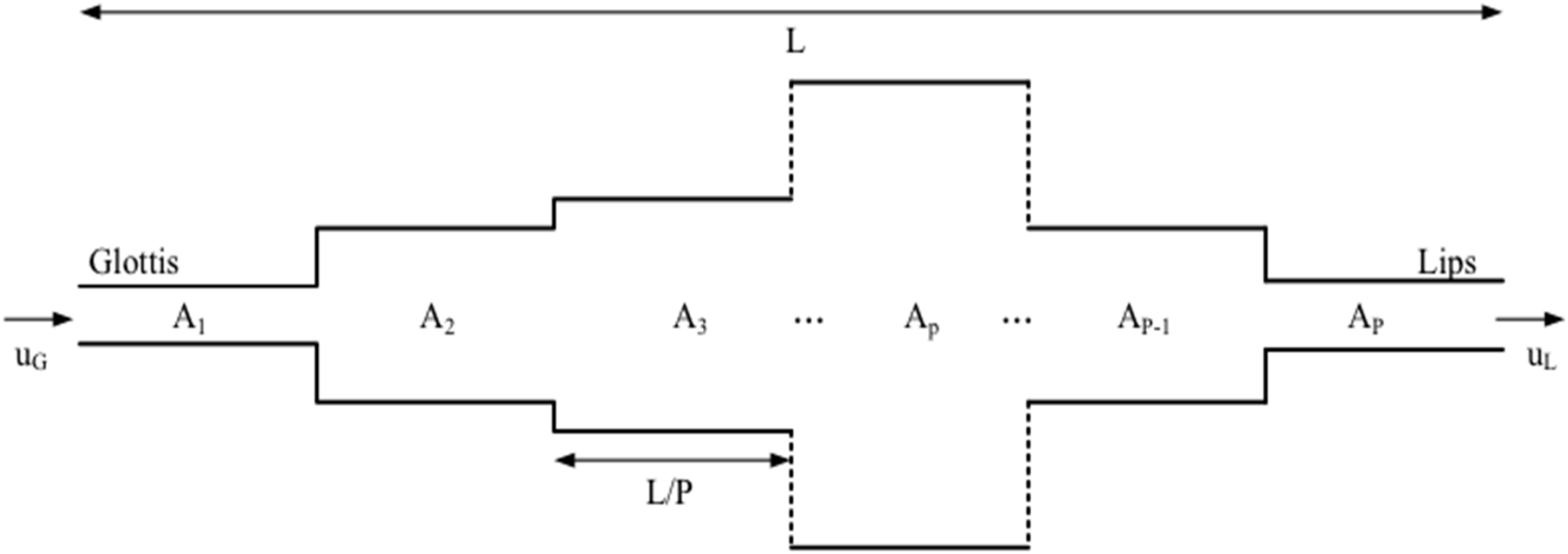


Voice Source – Video Endoscopy

Physiology of Speech

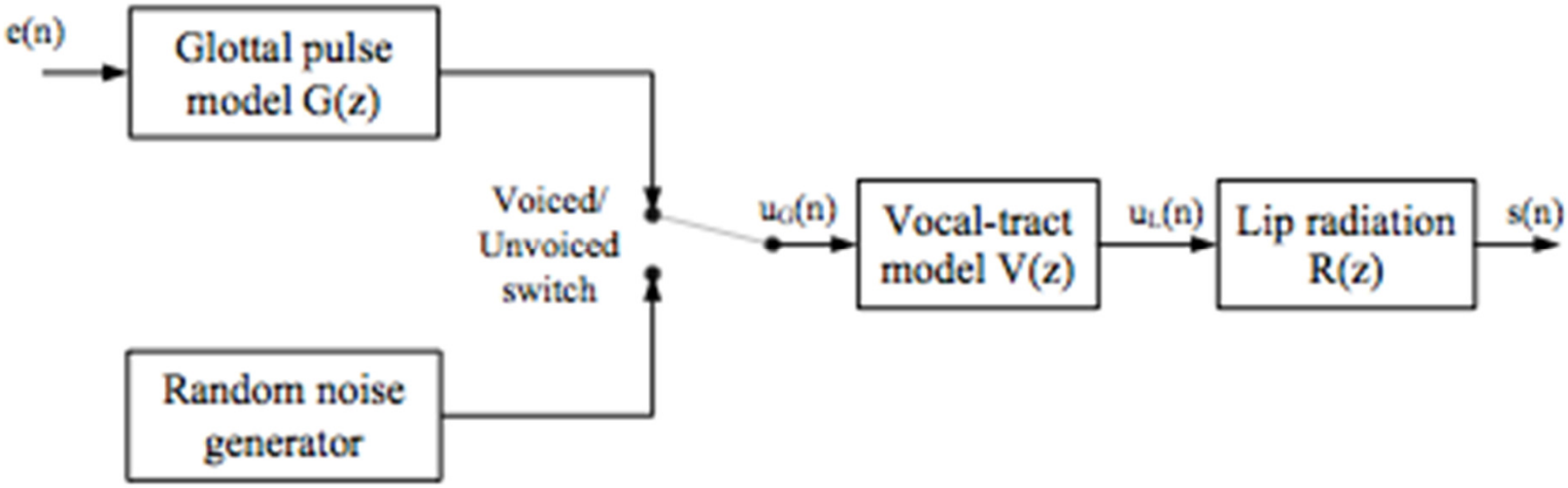


Lossless Tube Model



$$V(z) = \frac{z^{-P/2} K_V}{1 - \sum_{p=1}^{P} a_p z^{-p}}$$

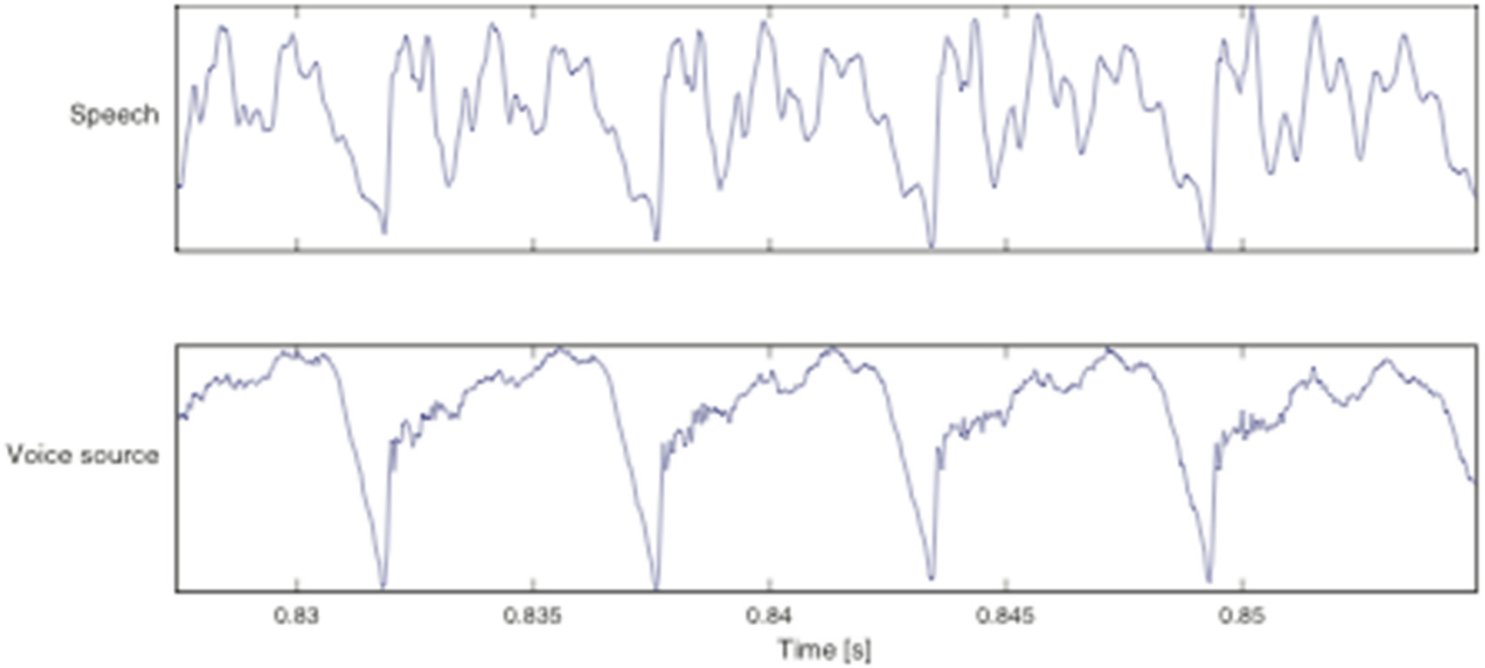
Model of Speech Production



Voice Source

- Extracted voice source using inverse filtering

$$s(n) = u_{\mathrm{d}}(n) + \sum_{k=0}^{p} a_k s(n-k)$$



Speech Spectrum

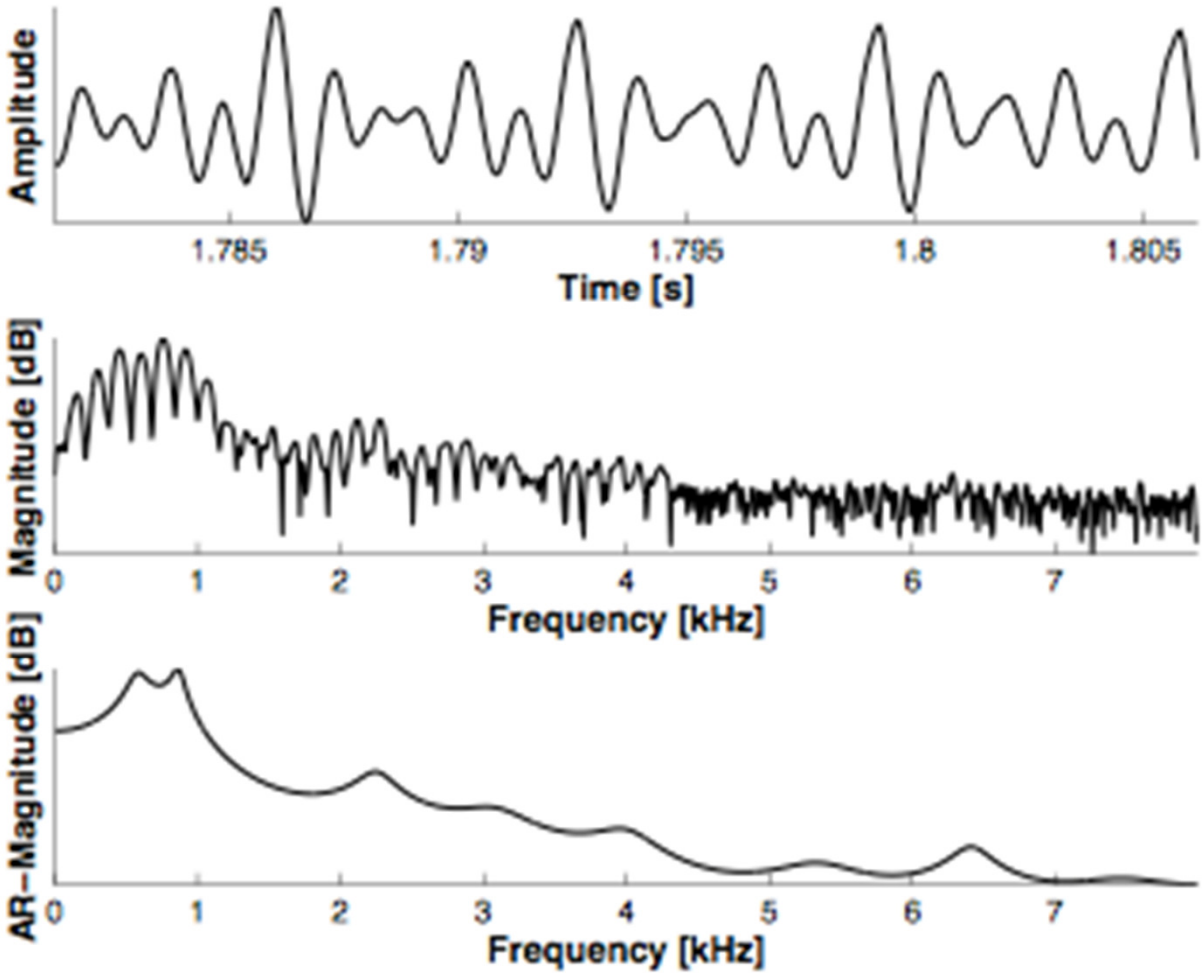
- Fourier Analysis applied to the speech signal:

$$S(k) = \sum_{n=1}^{N-1} s[n] e^{\frac{-j2\pi}{N} kn}, \quad k = 0, 1, \ldots, N-1$$

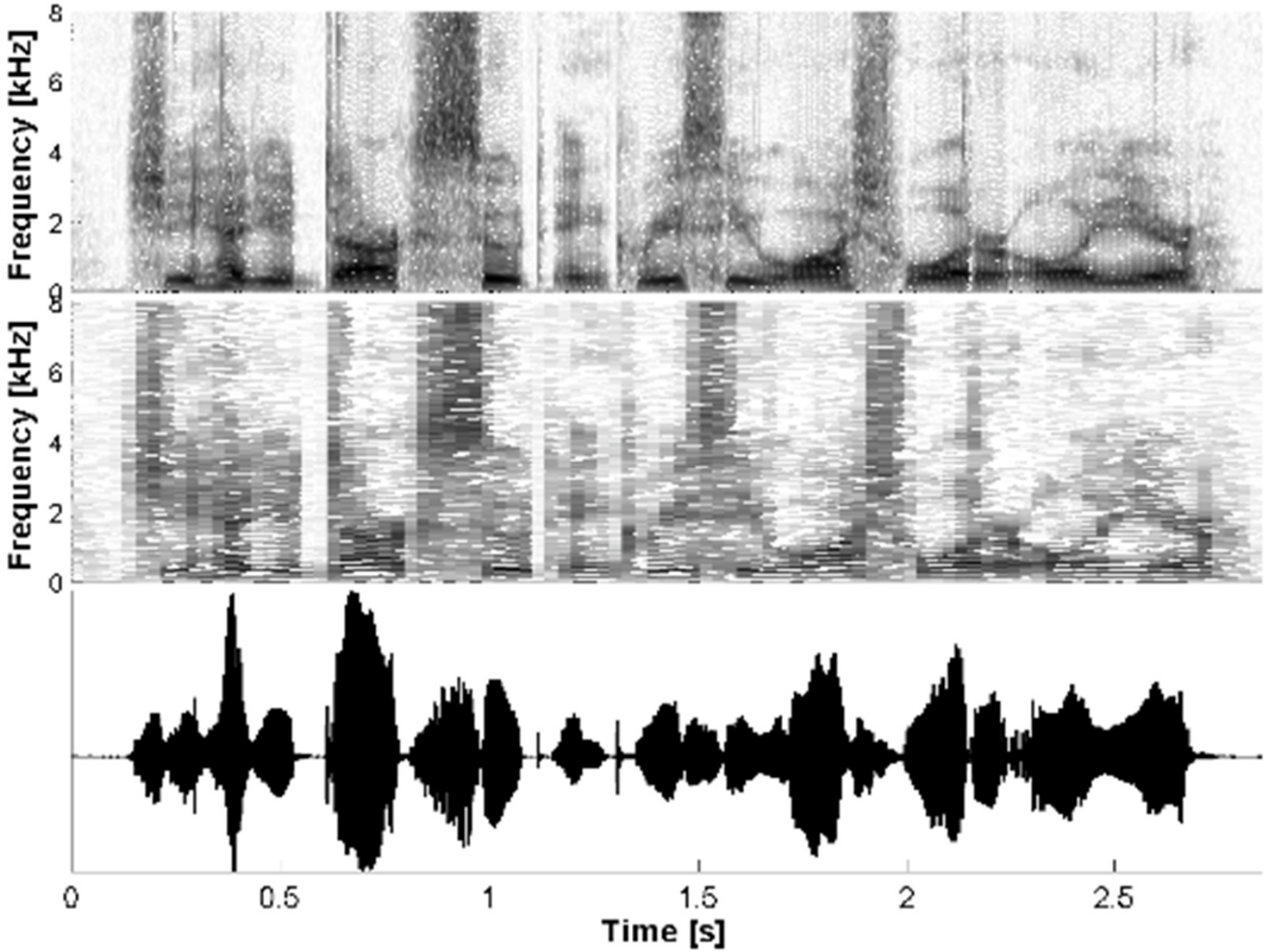
- Compare this to the envelope of V(z)

$$|V(k)| = \left| \frac{1}{1 - \sum_{p=1}^{P} a_p e^{-j2\pi kp/N}} \right|$$

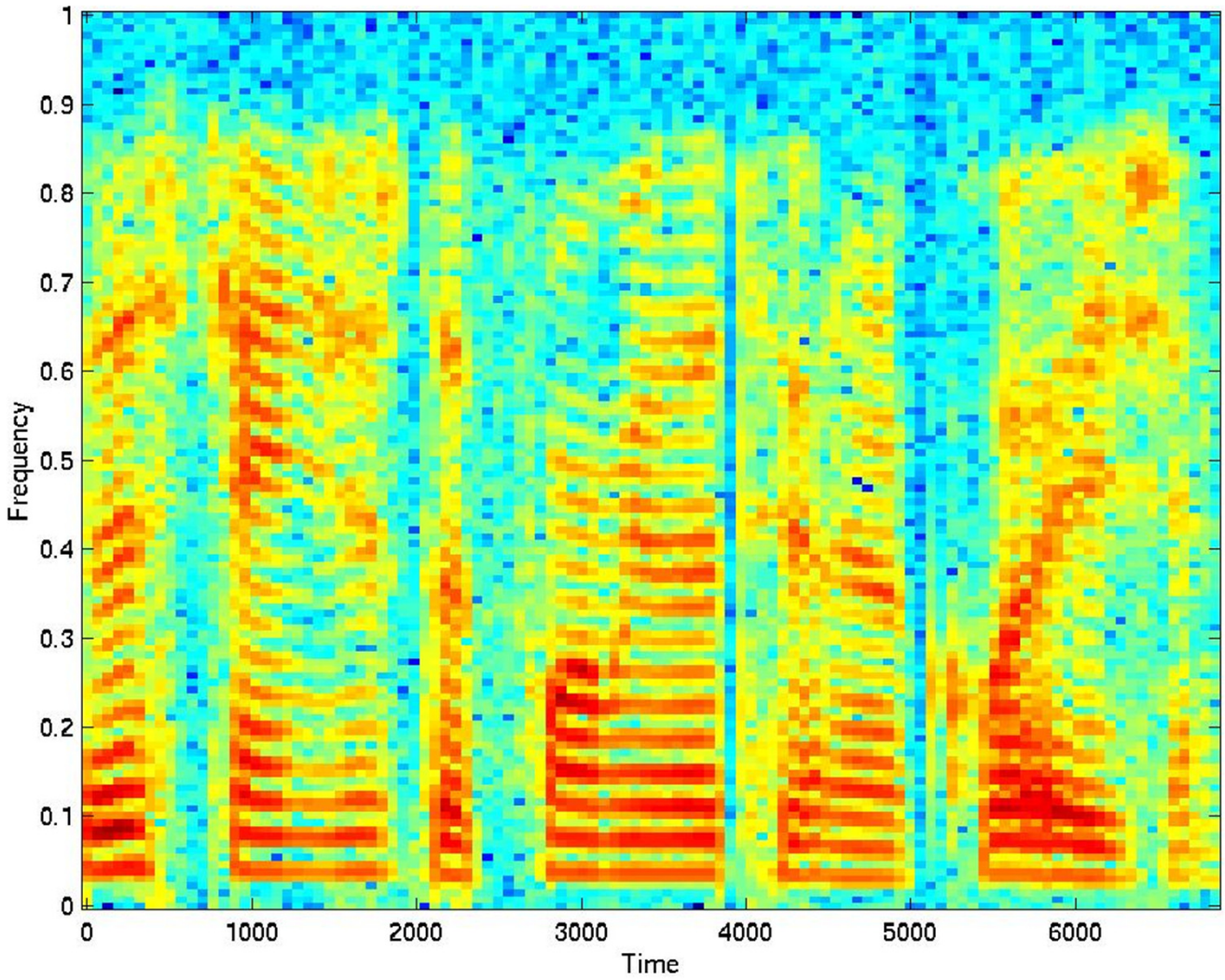
Speech Spectrum



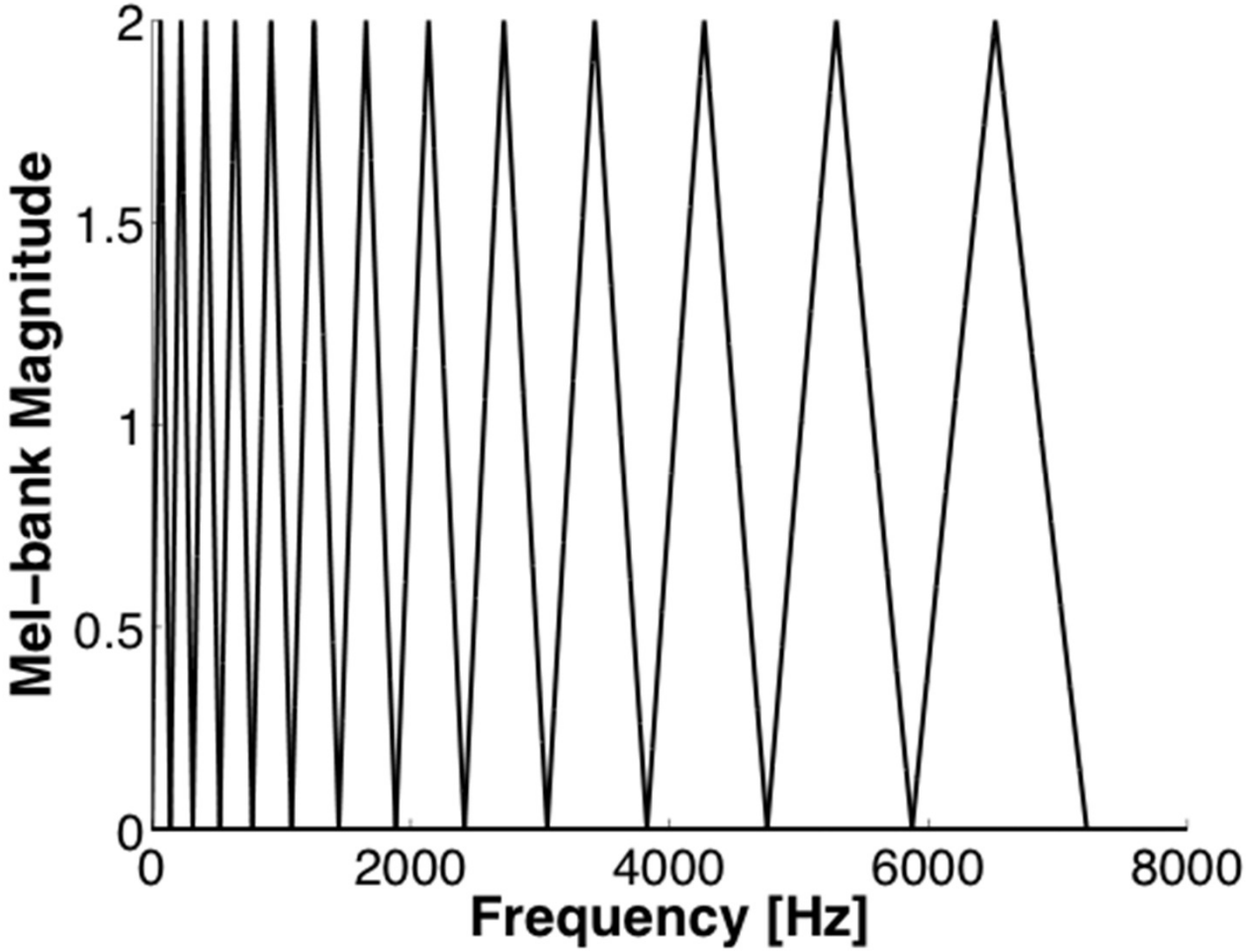
SPECTROGRAM



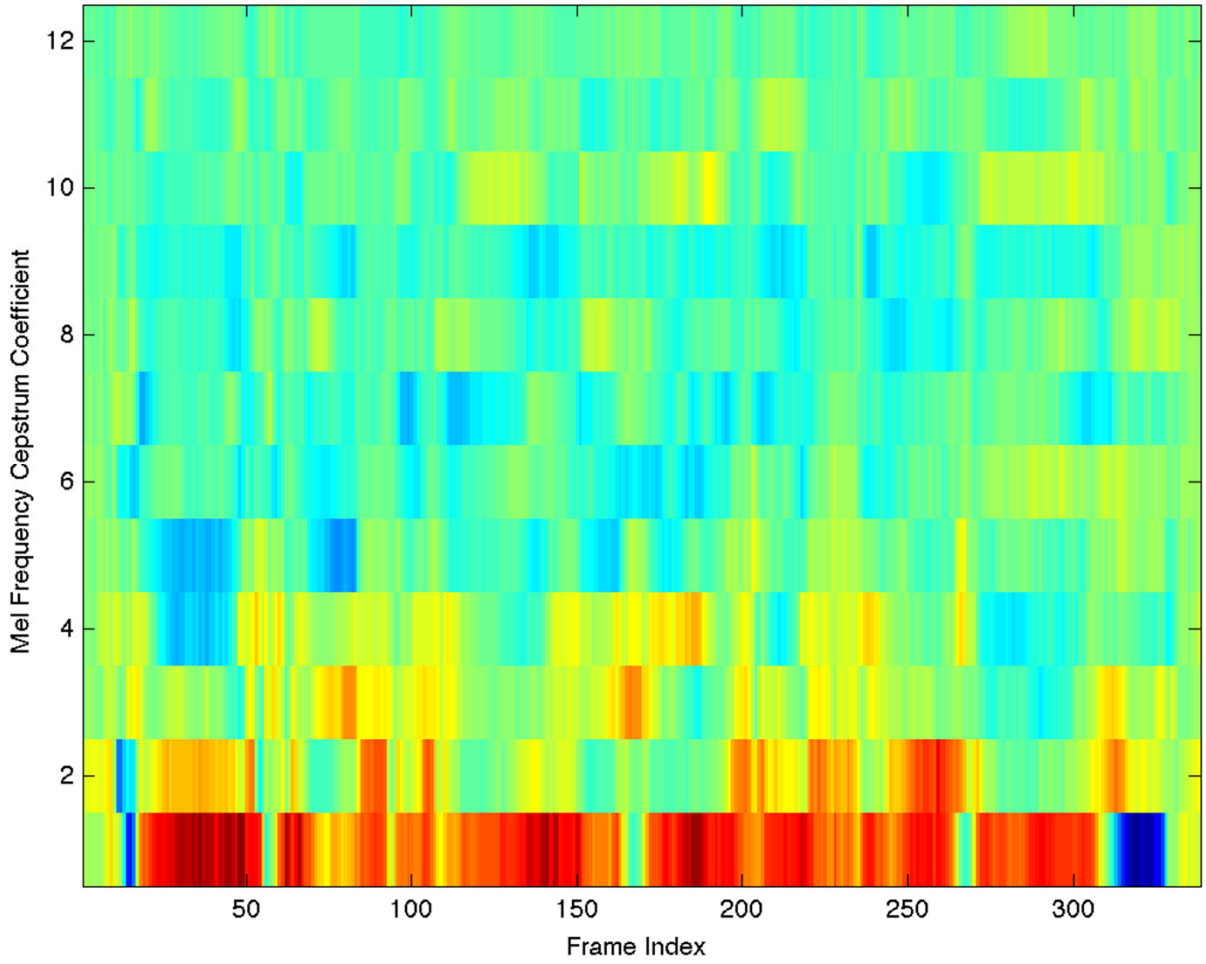
Spectrogram of Speech



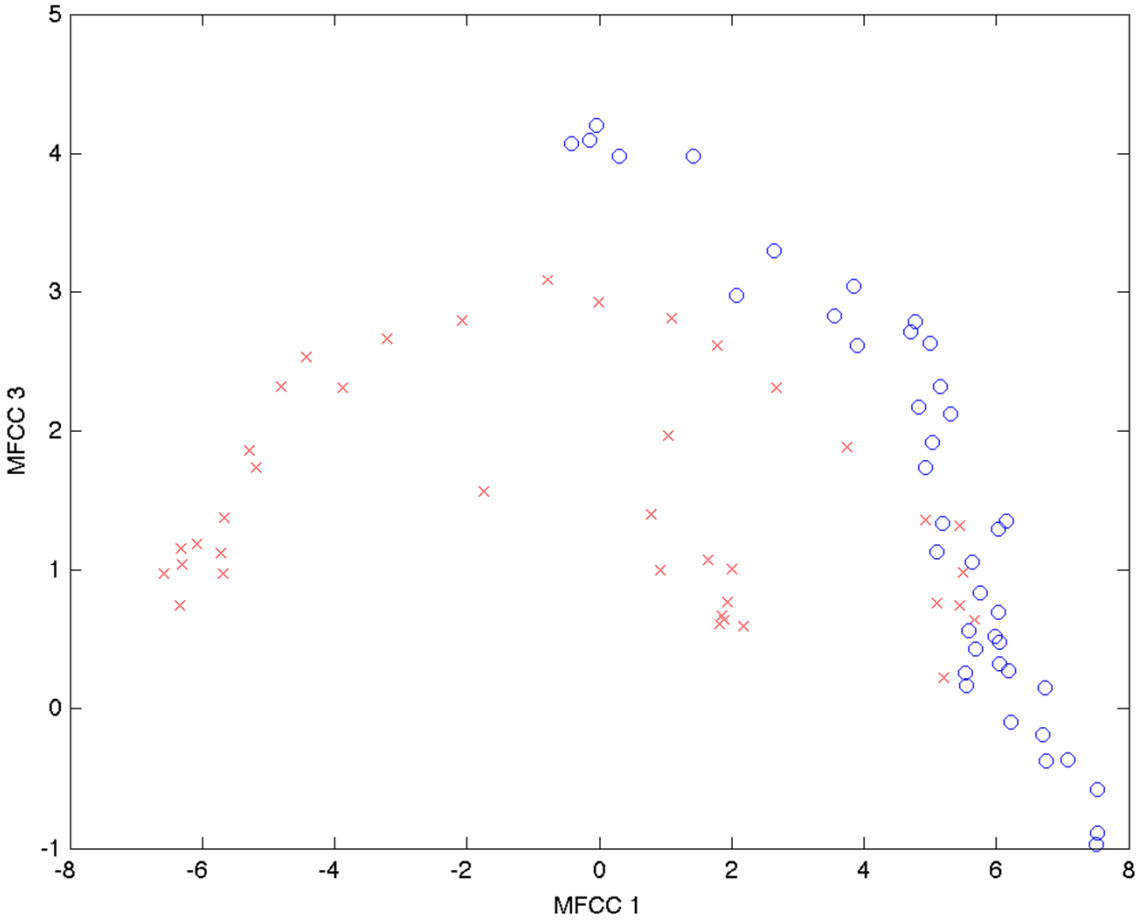
Speech Feature Extraction



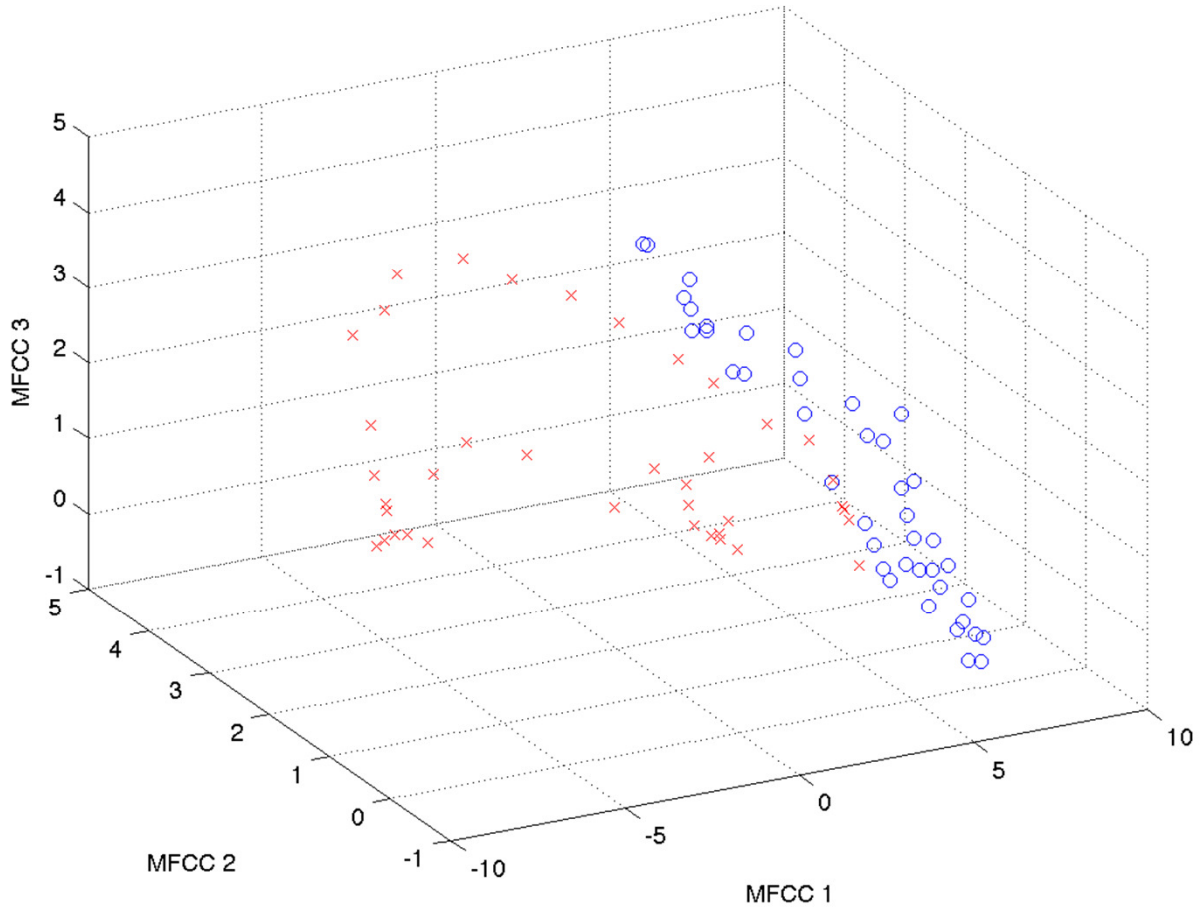
Mel-Frequency Cepstrum Coefficients



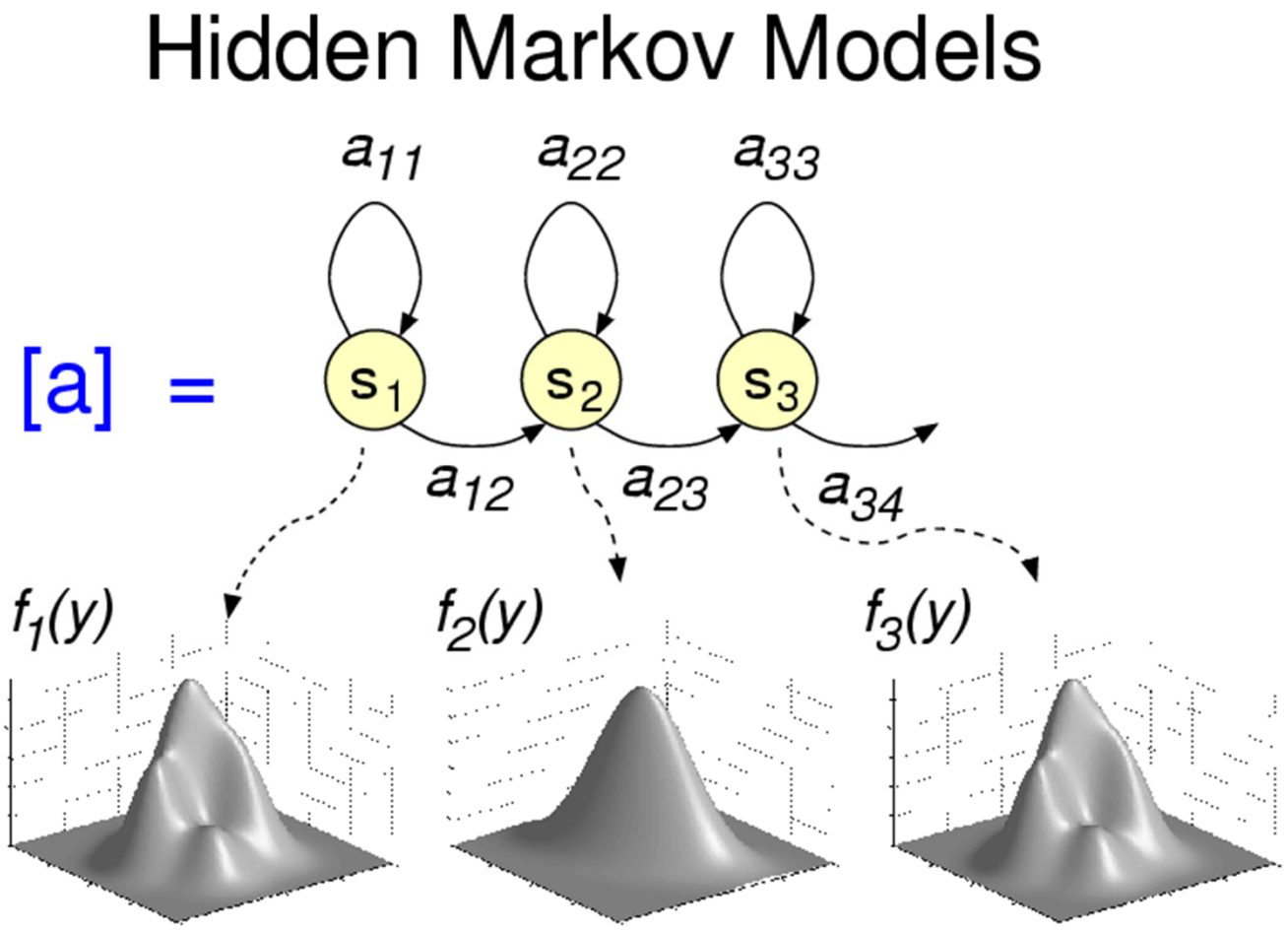
Features Scatter Plot



Features Scatter Plot



Connection to Hidden Markov Models



Three Tasks of HMMs

- Evaluation
  - What is the probability that this observation was generated by the model?
- Decoding
  - What is the most likely state sequence in the model that produced the observation
    - Viterbi, A* algorithm
- Learning
  - How should we adjust our model parameters given observation sequences
    - Baum Welch algorithm (special case of Expectation maximization, EM)

Hearing and Human Speech Recognition

- If we understand how hearing works we can:
  - Compress speech more efficiently
  - Imitate the functionality of hearing in speech processing
  - Examples: The Mel- and Bark scales
- Human speech recognition
  - Telecommunications
  - Psychoacoustics