What does speech “look” like to an Automatic Speech Recognition System?

Jiang Wu
jiang.wu@binghamton.edu
Electrical Engineering Department
What are “speech features?”

- Say if you are only allowed to use 39 values to represent a speech segment of 1 sec long...

- What are good features?
  - Discriminative
  - “Curse of Dimensionality”
How do we extract features from speech?

From both time and frequency domain...
For each frame of pre-recorded speech, we try to extract the feature as to compress its spectrum.
Most recent research has shown that spectral trajectories, over time, also play an important role in ASR.

Thus, we also want to let computers see what happens over time, about the center of each static feature.
So finally to the ASR system, the speech features look like...
Other Features for Future Study...

- **Pitch Contour:**
  - Strongly related to “tones”
  - Very popular feature type for tonal languages: Mandarin, Cantonese, Some of Korean dialects, etc.

- “Perceptual Features:
  - Analyze speech signal as to how human’s auditory system “perceptually” process sound
  - Frequency resolution and time resolution both depend on frequency and time.
Our Current Projects

- Project 1: To create an open source multi-language audio database for spoken language processing applications.
- Project 2: To understand tonal languages.

Our Speech Lab

- Dr. Montri Karnjanadecha (Force Alignment)
- Chandra Vootkuri (Ph.D.) (Landmark Theory)
- Brian Wong (M.S.) (Freq. Non-linearity)
- Andrew Hwang (M.S.) (Feature Transform)
Background you need..

- Signal processing (A/D)
- Probability theory, pattern recognition and machine learning
- Understanding of human auditory system/linguistic/musicality will be a bonus!
Speech Research is interesting!

Tons of interesting applications!!

- Pronunciation therapy
- Hearing aids
- Singing voice processing

[Image of hearing aid and singing voice processing]
Talk to a public computer in your real life...

- Not just the Microsoft speech-text software on your PC..
Thank you!

Questions?