Week 2

Thursday, September 18, 2025 3:59 PM

Agenda

Overview of what has been done so far

Papers read:

- In pursuit of Visemes
  - First introduction to the viseme side of things and the short comings involved
    - Multiple phonemes mapped to same viseme (McGuirk Effect)
- Simon King Speech Processing
  - Slides have been gone through
  - Introduction to ASR techniques
    - Dynamic Time Warping
    - Markov
- Speech Analysis and Perception
- SpecAugment
- MixSpeech

Data Augmentation Ideas

- Phonemes –
  - block out certain parts of spectrogram to hinder ability to distinguish phonemes
    - Frequency or time domain or both
  - Use spectrograms from audio with added noise – will make for less clear spectrograms making it harder to identify phonemes
  - Slow down audio to elongate speech – different phoneme lengths – dynamic time warping tested
  - Random silence insertion or false sentence start replication or dealing with
  - Skip and repeat audio
  - Time Warp (as seen in SpecAugment) -> not that useful according to paper
  - MixSpeech – concatenate two audio features and a singular visual feature
- Visemes –
  - alter the fps of visual side of things
  - Blacked out squares of mouth region to not rely on certain parts of mouth and jaw
  - Mouth only vs full face
  - Rotations and tilts
  - Skip and repeat frames

Help with setting up AV-Hubert in Colab

Would really appreciate some help setting up AV-Hubert in colab

What's next?

Set up AV-Hubert

Read rest of the reading you sent me (briefly looked at the "Data Augmentation" section of links yet)

Look into how to augment data in the ways outlined above

- I've done flips and rotates and blur before in 4C16