Kaldi-based Robust ASR System for ATC Applications

Problem Statement

Automatic speech recognition (ASR) can be used to transcribe live ATC transmissions into text in real time.

- Helpful learning resource for student pilots to practice their ATC communication skills
- Helping teaching resource for pilot instructors to train their students.

Data Sources

Training Data:

- Air Traffic Control Complete: audio recordings and the corresponding transcriptions from Dallas-Fort Worth, Logan International, and Washington National airports.
- ATCO2: speech corpus containing audio communications and transcriptions from Prague and Brno airports
- Air Traffic Control Simulation: audio recordings and controller transcriptions rom EUROCONTROL Experimental Centre
- ZCU CZ ATC Corpus: audio and transcripts from the Czech airspace

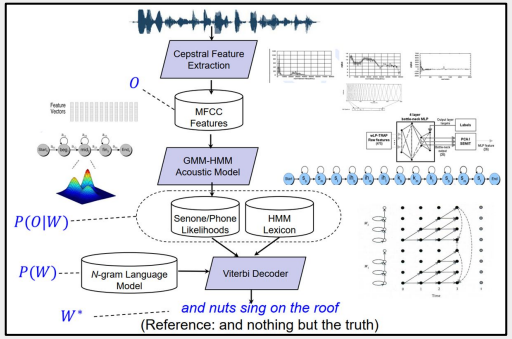
Input:

- LiveATC: live audio communications from ATC and pilots
  - Live Demos
  - Student Pilots

Research Description

How can ASR modeling be utilized by student pilots at ERAU as a resource to improve their ATC communication skills?

ASR modeling can be used to transcribe live ATC transmissions into text in real time, which reduces the difficulty of learning ATC vernacular for student pilots.



Outline

Decoding:

- Partition audio into data segments
- Convert segments from time-domain to frequency-domain
- Generate a 39-dimensional MFCC feature vector
- Find the distance to the closest phoneme state
- Construct a chain of phonemes states
- Build triphones from phoneme states
- Build monophones from triphones
- Build words from monophones
- Compile and predict next words into sentences

Training:

- The ASR model shall train itself to recognize speech and sound through changes in frequency by using the training data
- The ASR model shall attempt to match the labels and continues training until its improvement plateaus