# Kaldi ASR Team Sprint 05 Demo

Adam G., Milan H., Tabitha O., David S., Tahmina T.

#### Introduction

- Develop acoustic ASR model capable of transcribing live ATC transmissions in real-time using the Kaldi ASR Toolkit and the ATC02 corpus.
  - Integration with RTube web application in the future
    - Training tool for student pilots.
    - Designed to make learning ATC phraseology easier.



#### Design Constraints

- Minimum storage size of 12.5 GB
- Minimum video memory (VRAM) of 12 GB
  - NVidia GeForce RTX 4080
- All inputs as WAV files
  - Convert non-WAV files using FFMPEG
- General American English
  - "Color" vs. "Colour"
- No punctuation or grammatical marks
  - Commas, colons, etc.
- Federal Aviation Administration (FAA)



#### Assumptions and Dependencies

- Clear and direct communication
- Low interference and background noise
- General American English
- Sufficiently large data set
- Sufficient storage
- Sufficient video memory
- Operating system
  - o Linux (e.g., Ubuntu, Debian, etc.)
  - Windows Subsystem for Linux (WSL)
  - Virtual machine for MacOS



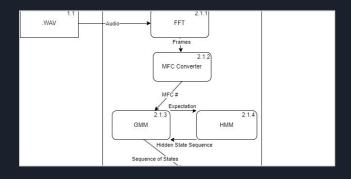
## System Architecture (Preprocessing)

- ATC0 Corpus:
  - Linguistic Data Consortium (LDC)
  - Contains 30-hour ATC dataset with audio files and text transcriptions
- ATCO2 Corpus:
  - Repository for ASR and NLP research
  - Provides preprocessing script
    written in Bash designed for Kaldi
    ASR Toolkit

```
| Comparison of the Comparison
```

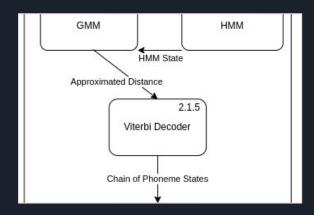
#### System Architecture (Preparation)

- Setting up audio for decoding
- Fast Fourier Transform (FFT)
- Discrete Cosine Transform (DCT)
- Mel-Frequency Cepstrum (MFC)
  - Mel-Frequency Cepstral Coefficients (MFCCs)
- Gaussian Mixture Model (GMM)
- Hidden Markov Model (HMM)



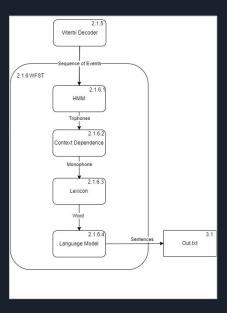
### System Architecture (Decoding)

- Finding sequence of events
- Viterbi decoder



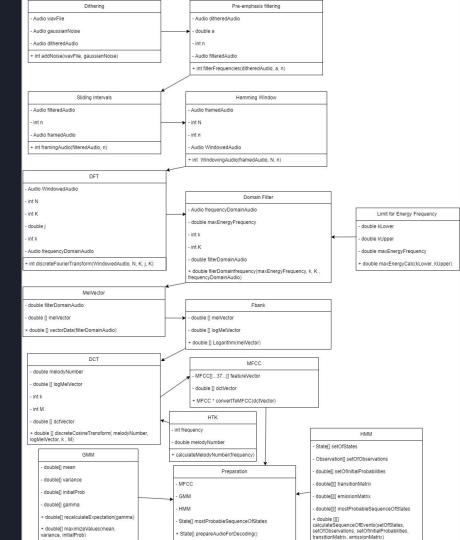
### System Architecture (Output)

- Converting to sentences
- Weighted Finite State Transducers (WFSTs)
  - o HMM
  - Context Dependence
  - Lexicon
  - Language Model



## Sub-System Design

- Data Preparation Phase
  - Mel-Frequency Cepstrum Coefficients (MFCC)
  - Gaussian Mixture Model (GMM)
  - Hidden Markov Model (HMM)
  - MFCC is the culmination of a dozen steps as shown
  - Aside from Data Preparation there are also the Decoding and Output Phases



#### Project Timeline

- Sprint 05:
  - System Requirements Specification
  - System Design Document
  - System Research & Test Plan
  - Use case diagram
  - Audio preprocessing script
  - User Manual Guide
- Sprint 06 Plans:
  - Integration of preprocessing script with Mini-LibriSpeech
  - Creation of script to run Mini-LibriSpeech
  - MacOS installation



#### Lessons Learned

- Importance of establishing guidelines with product owner.
- Importance of establishing product timeline with product owner and course instructor.
- Laborious installation processes
  - Mini-LibriSpeech
  - o ATCO2 corpus
  - MacOS and Oracle VirtualBox
  - MacOS through Embry-Riddle Cisco VPN



Q&A