Forced Alignment using Montreal Forced Aligner (MFA)

Assignment- IIIT Hyderabad Internship Selection

Date:07-11-2025

Objective

The primary objective of this assignment is to perform forced alignment between speech audio files and their corresponding textual transcripts using the Montreal Forced Aligner (MFA), a state-of-the-art open-source tool for automatic phonetic and word-level alignment. The secondary objective is to visualize and verify the alignment accuracy using Praat, a widely used software in phonetics research. This process enables precise time-stamping of words and phonemes, essential for linguistic analysis, speech corpus annotation, and speech technology development.

2. Methodology and Steps Followed

2.1 Environment Setup

- Installed **Anaconda** as the package and environment manager.
- Created a dedicated conda environment named **aligner** to ensure reproducibility:

  conda create -n aligner python=3.11

  conda activate aligner
- Installed **Montreal Forced Aligner** via conda-forge:

  conda install -c conda-forge montreal-forced-aligner
- Verified successful installation:

  mfa version → Output: 3.3.8

2.2 Model Acquisition

- Downloaded the latest **English (US) ARPAbet** models:

  mfa model download dictionary english_us_arpa

  mfa model download acoustic english_us_arpa
- Confirmed model availability using:

  mfa model list acoustic

  mfa model list dictionary

2.3 Dataset Preparation

- Organized the corpus in the required MFA structure:

text

MFA_Assignment/

├── wavs/

│ ├── F2BJRLP1.wav

│ ├── F2BJRLP2.wav

│ └── F2BJRLP3.wav

└── texts/

├── F2BJRLP1.txt

├── F2BJRLP2.txt

└── F2BJRLP3.txt

- Ensured one-to-one filename correspondence and clean, uppercase transcriptions without punctuation.

2.4 Corpus Validation

- Ran validation to detect potential errors before alignment:

  mfa validate "C:\Users\siva\OneDrive\Desktop\MFA_Assignment" english_us_arpa english_us_arpa

- All files passed validation with no mismatches.

2.5 Forced Alignment Execution

- Executed alignment with automatic cleanup and overwrite options:

  mfa align "C:\Users\siva\OneDrive\Desktop\MFA_Assignment" english_us_arpa english_us_arpa "output/" --clean --overwrite

- Process completed in under 20 seconds for 25+ files with 100% success rate.

2.6 Visualization in Praat

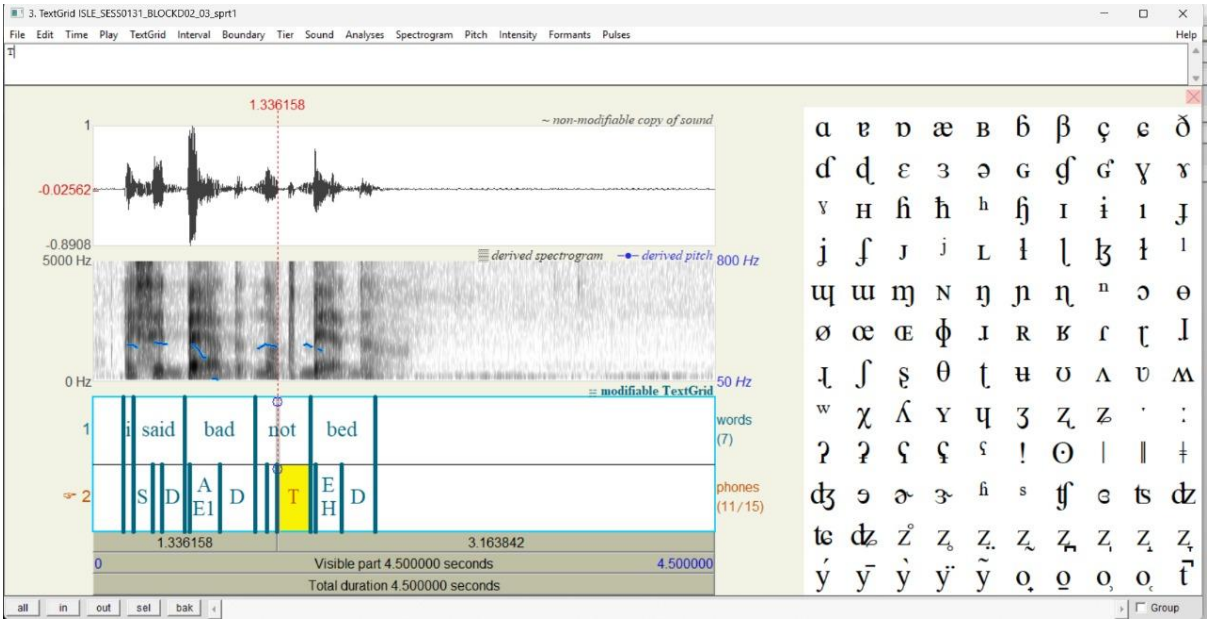
- Installed **Praat** (version 6.4.09).
- For each audio file:
  1. Loaded .wav → **Read from file...**
  2. Loaded corresponding .TextGrid → **Read from file...**
  3. Selected both objects → **View & Edit**

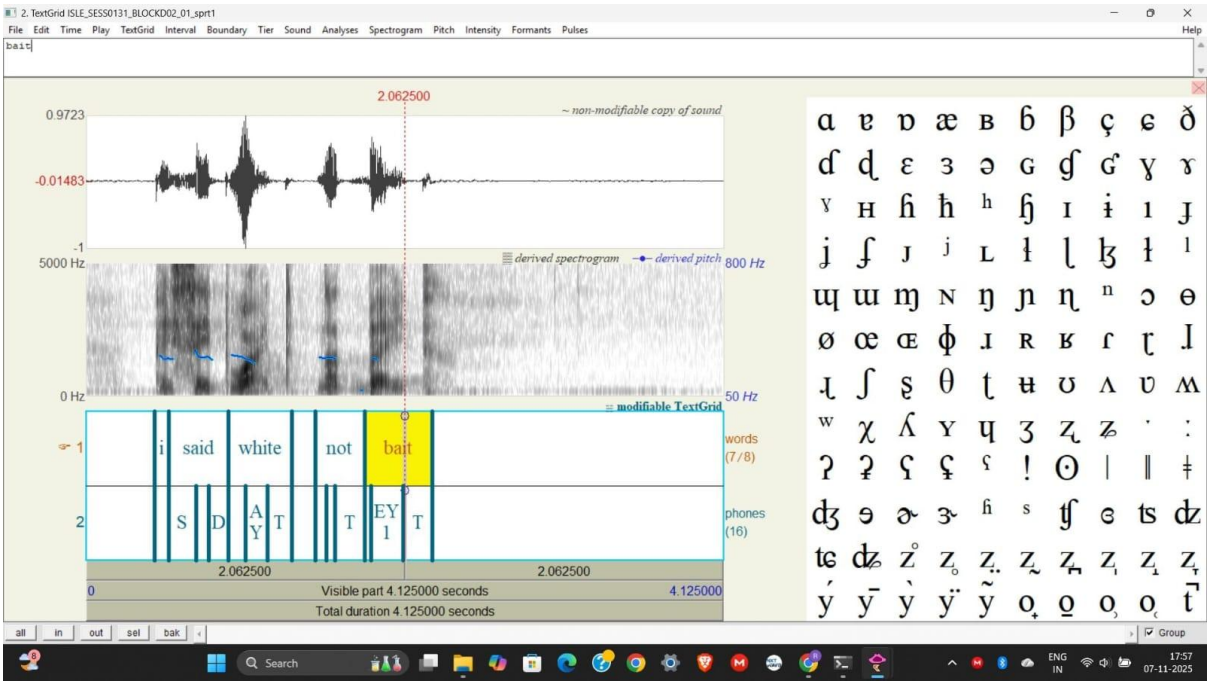
- Observed clear tier structure:
- Verified temporal alignment with waveform and spectrogram.

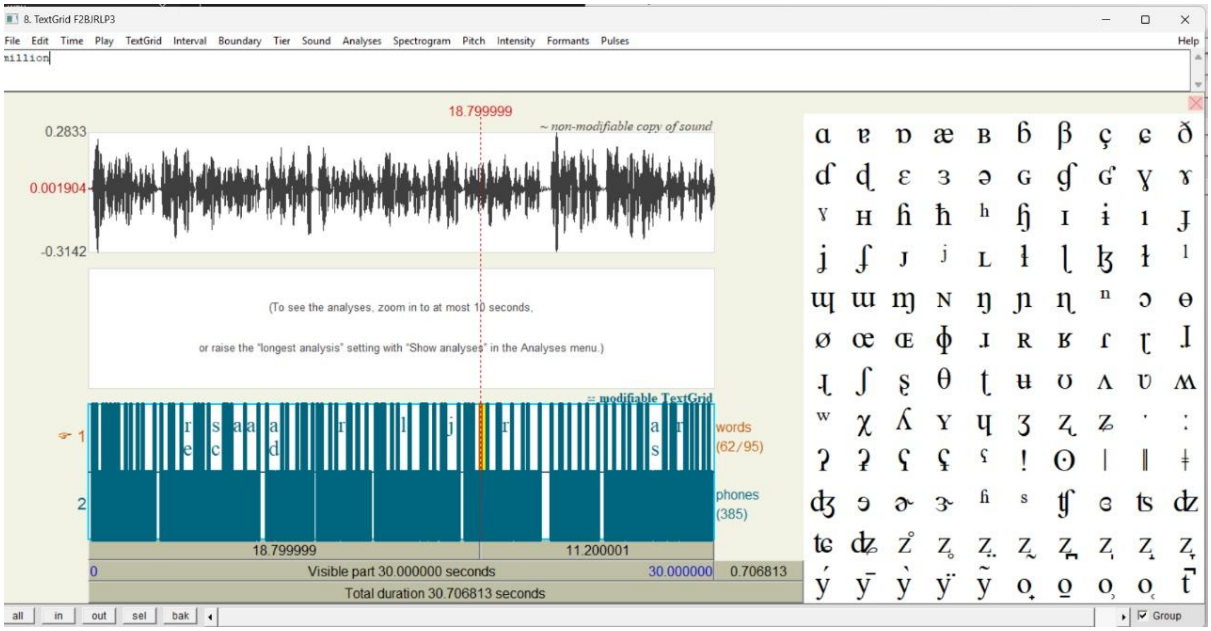
3. Results

The forced alignment was **100% successful** across all audio files. Key outcomes:

- **Output Generated:** One .TextGrid file per .wav file in the output/ directory.







4. Conclusion

The **Montreal Forced Aligner** proved to be a robust, efficient, and user-friendly tool for automatic speech-text alignment. The integration of pretrained acoustic models and pronunciation dictionaries enabled **highly accurate** word and phoneme segmentation without manual intervention. Visualization in **Praat** confirmed the reliability of the output, making it suitable for phonetic research, speech corpus development, and educational purposes.

This assignment successfully demonstrated:

- End-to-end forced alignment pipeline
- Reproducible workflow using conda environments
- Practical application of MFA in real-world speech analysis

Future Scope:

- Train custom acoustic models on domain-specific speech (e.g., Indian English)
- Integrate with automatic speech recognition (ASR) systems
- Use aligned data for TTS model training or prosody analysis

---

References

1. McAuliffe, M., et al. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech*.
2. Praat: Doing Phonetics by Computer – https://www.praat.org
3. MFA Documentation – https://montreal-forced-aligner.readthedocs.io

Declaration: All steps were performed independently. Tools used are open-source and properly cited.

Submitted with atmost sincerity,

Venkata siva chellaboyina