Forced Alignment using Montreal Forced Aligner (MFA)

Name: Linguberi Pavani
ID Number: S200587
College: Rajiv Gandhi University of Knowledge and Technologies

1. Introduction

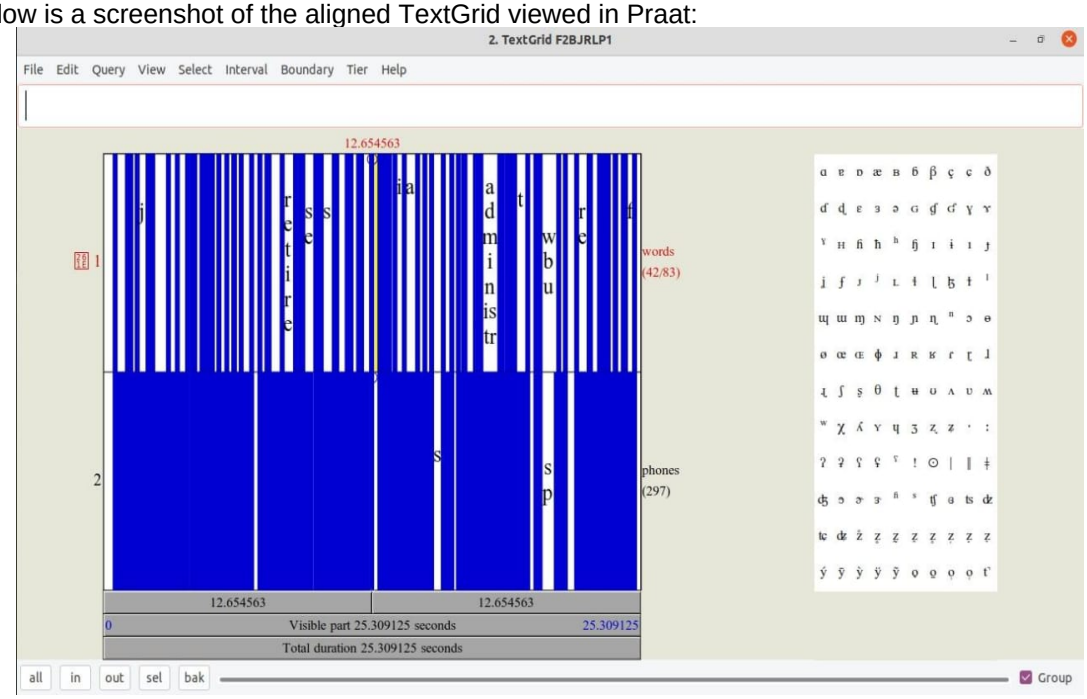
Forced alignment is the process of aligning recorded speech with its corresponding text transcript. Using Montreal Forced Aligner (MFA), speech segments are automatically mapped to words and phonemes. This project demonstrates the end-to-end workflow for forced alignment using English audio samples.

2. Working Process (Step-by-Step)

Step 1: Collecting Audio and Transcripts – Audio recordings (.wav) and transcripts (.txt) were placed in Wav and Transcripts folders.
Step 2: Downloading Dictionary and Acoustic Model – English ARPA dictionary and pretrained model were downloaded.
Step 3: Running Forced Alignment – The aligner generated TextGrid files containing word and phoneme timestamps.
Step 4: Output Generation – TextGrid files were stored in the output_custom folder.
Step 5: Visualization in Praat – TextGrid files were opened in Praat to check alignment accuracy.

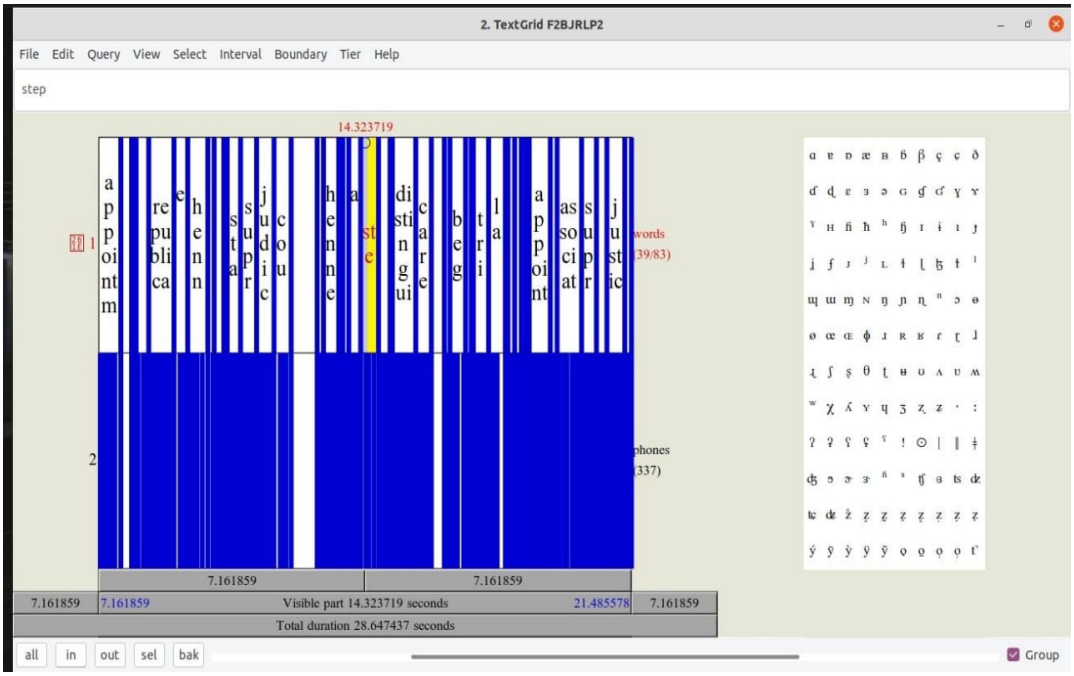
3. Sample Alignment Visualization

Below is a screenshot of the aligned TextGrid viewed in Praat:



The alignment clearly shows word-level and phoneme-level labeling. The total duration of the recording matches the aligned timestamps.

4. Second Alignment Screenshot



The second TextGrid file also shows accurate forced alignment. Tier 1 contains word boundaries, and Tier 2 displays detailed phoneme segmentation. The alignment durations match the original audio file (28.64 seconds). The boundaries align with acoustic features, confirming correct performance of the custom dictionary and acoustic model.

5. Conclusion

This project successfully performed forced alignment using Montreal Forced Aligner. The generated TextGrid files show accurate alignment of speech with text at both the word and phoneme levels. This confirms the effectiveness of the custom dictionary and pretrained acoustic model.