Assignment 1: Forced Alignment using Montreal Forced Aligner (MFA)

Environment: Windows 11, WSL Ubuntu 22.04, conda env(mfa_env)

Objective

To align speech and text data from the ISLE corpus and F2BJRLP audio clips using the Montreal Forced Aligner (MFA), generating corresponding TextGrid files for precise phonetic and word-level time alignment.

Forced Alignment:

Forced alignment is the process of automatically matching spoken audio with its written transcript so that each word and sound (phoneme) is aligned with its exact timing in the recording. The Montreal Forced Aligner (MFA) performs this task by analyzing both the audio and text to produce TextGrid files, which mark the start and end times of each word and phoneme.

For example,

Audio:

A speaker says — *"Good morning"*

Transcript:

GOOD MORNING

A forced aligner (like MFA) analyzes the audio and produces a TextGrid file showing the word and phoneme boundaries (start and end times).

for example:

The audio spanning 0.00 to 1.20 seconds articulates the phrase "GOOD MORNING", where each phoneme corresponds to a specific time frame:

G (0.00–0.15), UH (0.15–0.35), D (0.35–0.55), M (0.55–0.65), AO (0.65–0.85), R (0.85–1.00), N (1.00–1.10), IH NG (1.10–1.20).

Dataset:

mfa_project/

│

├── corpus/

│ ├── F2BJRLP1.wav

│ ├── F2BJRLP1.txt

│ ├── F2BJRLP2.wav

│ ├── F2BJRLP2.txt

│ ├── F2BJRLP3.wav

│ ├── F2BJRLP3.txt

│ ├── ISLE_SESS0131_BLOCKD02_01_sprt1.wav

│ ├── ISLE_SESS0131_BLOCKD02_01_sprt1.txt

│ ├── ISLE_SESS0131_BLOCKD02_02_sprt1.wav

│ ├── ISLE_SESS0131_BLOCKD02_02_sprt1.txt

│ ├── ISLE_SESS0131_BLOCKD02_03_sprt1.wav

│ └── ISLE_SESS0131_BLOCKD02_03_sprt1.txt

├── output/

│ ├── F2BJRLP1.TextGrid

│ ├── F2BJRLP2.TextGrid

│ ├── F2BJRLP3.TextGrid

│ ├── ISLE_SESS0131_BLOCKD02_01_sprt1.TextGrid

│ ├── ISLE_SESS0131_BLOCKD02_02_sprt1.TextGrid

│ ├── ISLE_SESS0131_BLOCKD02_03_sprt1.TextGrid

├── custom.dict

├── praat_view/

└── wordlist.txt

Method: Tools & Commands:

Environment

create & activate (if not already)

- **conda create -n mfa_env python=3.10 -y**
- **conda activate mfa_env**

install MFA & Praat

- **pip install montreal-forced-aligner**
- **sudo apt-get update && sudo apt-get install -y praat dos2unix**

Project layout

- **mkdir -p ~/mfa_project/{corpus,output}**

copy wav + txt into corpus

- **ls ~/mfa_project/corpus**

Fix line-endings (Windows → Unix)

- **dos2unix ~/mfa_project/corpus/*.txt**

Get models

dictionary or acoustic models can be fetched by MFA

- **mfa model download dictionary english_us_arpa**
- **mfa model download acoustic english_us_arpa**

Validate corpus

- **mfa validate ~/mfa_project/corpus english_us_arpa**

or using your lexicon

- **mfa validate ~/mfa_project/corpus ~/mfa_project/custom.dict**

Align

with stock dictionary

- **mfa align ~/mfa_project/corpus english_us_arpa english_us_arpa ~/mfa_project/output --overwrite --num_jobs 4**

or with your custom dictionary

- **mfa align ~/mfa_project/corpus ~/mfa_project/custom.dict english_us_arpa ~/mfa_project/output --overwrite --num_jobs 4**

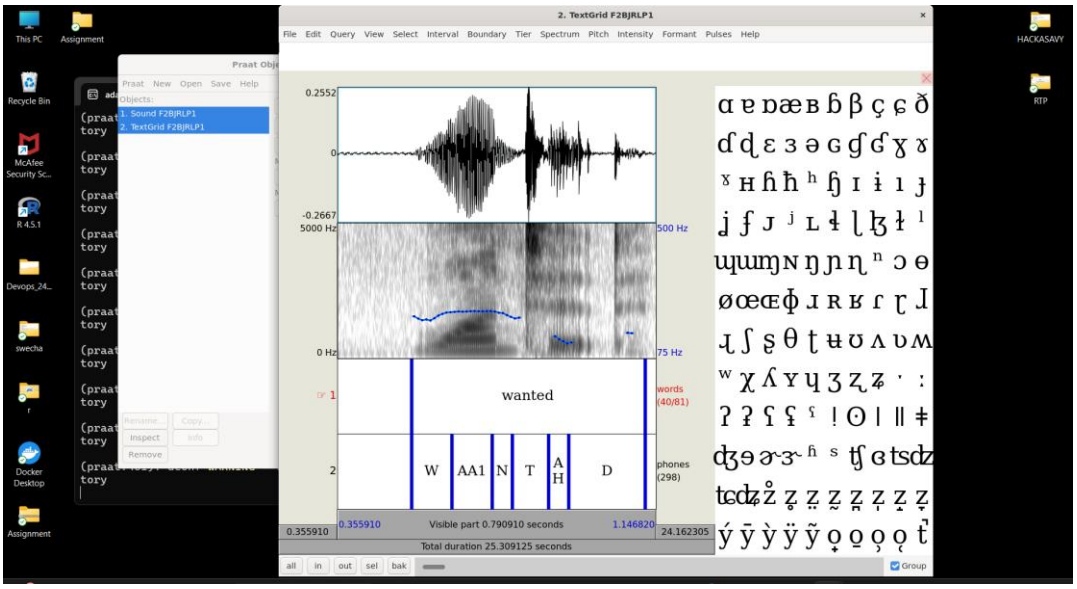
Inspect outputs

- **ls ~/mfa_project/output**
- **praat &**

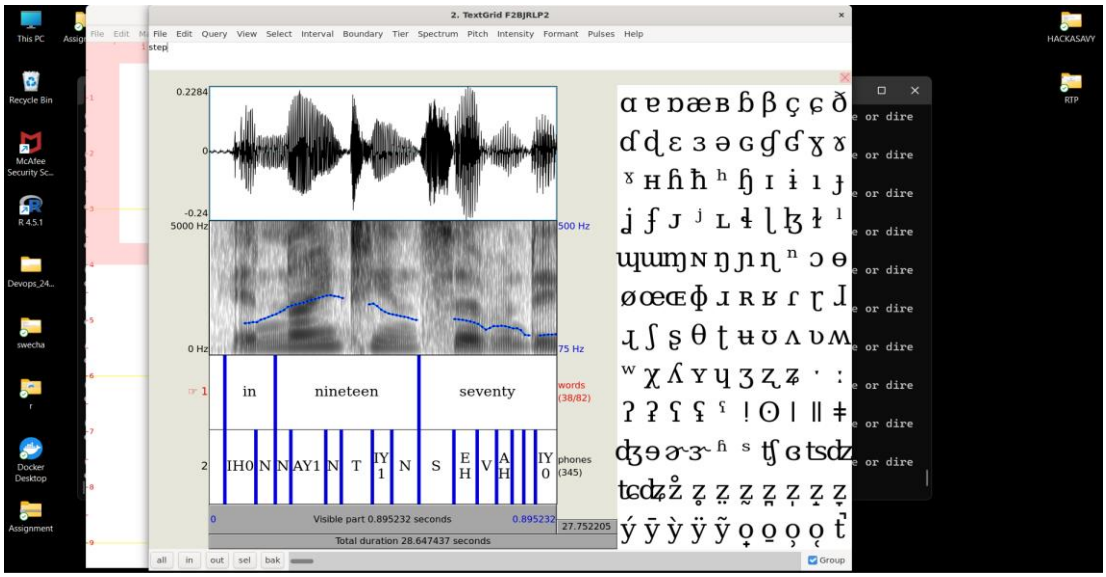
Results

Visualization pf Praat Objects

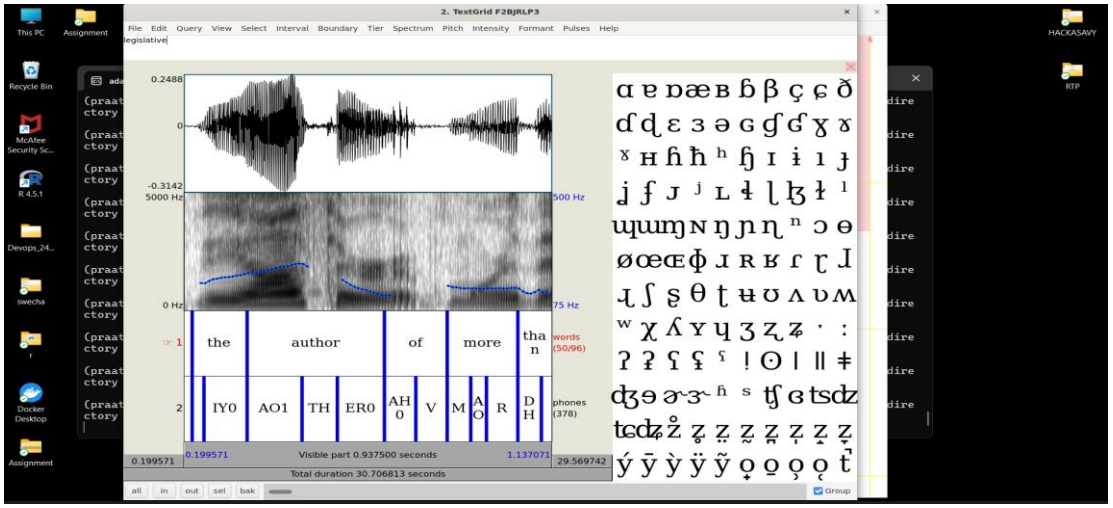
Output of F2BJRLP1.TextGrid



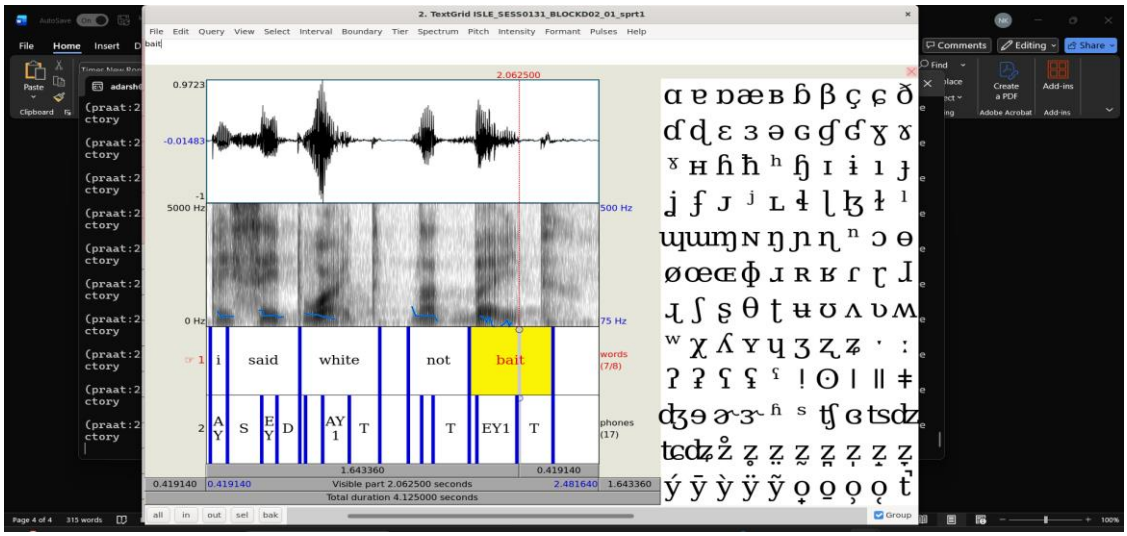
Output of F2BJRLP2.TextGrid



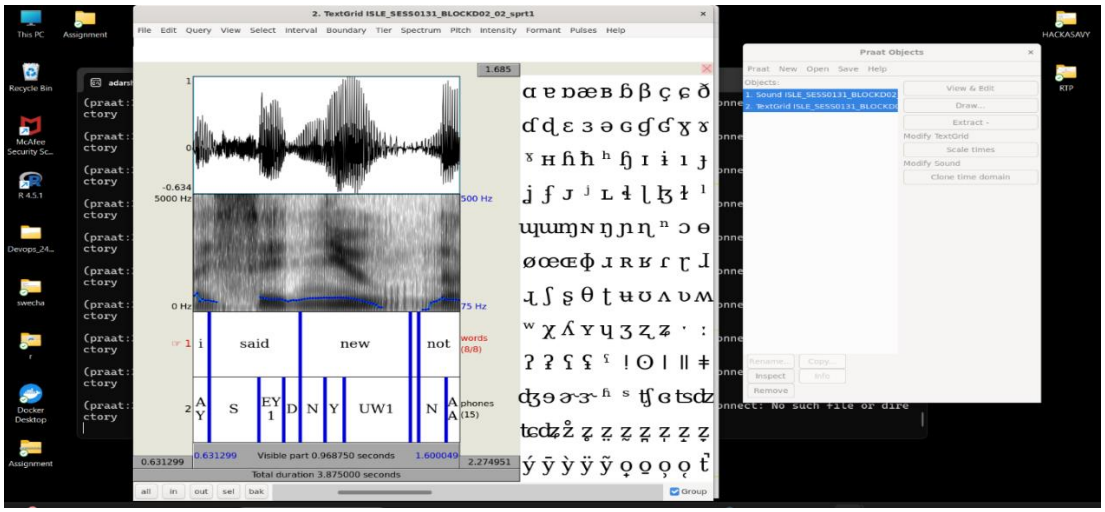
Output of F2BJRLP3.TextGrid



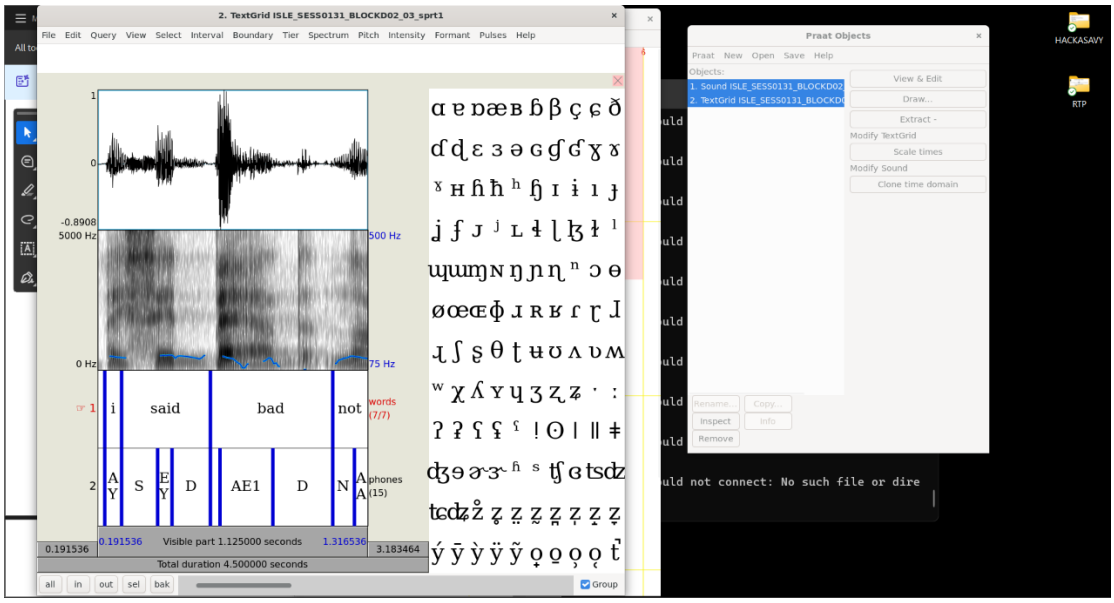
Output of ISLE_SESS0131_BLOCKD02_01_sprt1.TextGrid



Output of ISLE_SESS0131_BLOCKD02_02_sprt1.TextGrid



Output of ISLE_SESS0131_BLOCKD02_03_sprt1.TextGrid



Github: https://github.com/KompallyNikshitha/montreal_forced_aligner

Google Drive:
https://drive.google.com/file/d/1vnN_GNKTVnUdkeGEebL6mmiJ_O7yow4o/view?usp=sharing

Contact Us:

KOMPALLY NIKSHITHA

nikshithakompally08@gmail.com