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

Objective

The goal of this assignment is to set up and execute a complete forced alignment pipeline using the Montreal Forced Aligner (MFA) tool, and to understand how automatic alignment works between speech audio and phonetic transcription.

Forced alignment is the process of automatically matching an audio recording with its corresponding text transcription at the word and phoneme level. It determines when each word or sound begins and ends in the speech signal.

Environment Setup

- **Operating System**: Windows 11
- **Environment**: Conda (`PythonProject13`)
- **Tools Used**: Montreal Forced Aligner (MFA), Praat, PyCharm

Corpus Preparation

- All `.wav` files are mono, 16kHz, signed 16-bit PCM
- All `.lab` files are uppercase, punctuation-free, and dictionary-compatible
- Both `.wav` and `.lab` files are placed in the same `wav/` folder for MFA compatibility

Model Used

- **Pronunciation Dictionary**: `english_us_arpa`
- **Acoustic Model**: `english_us_arpa`
- Both models were downloaded using MFA's CLI:

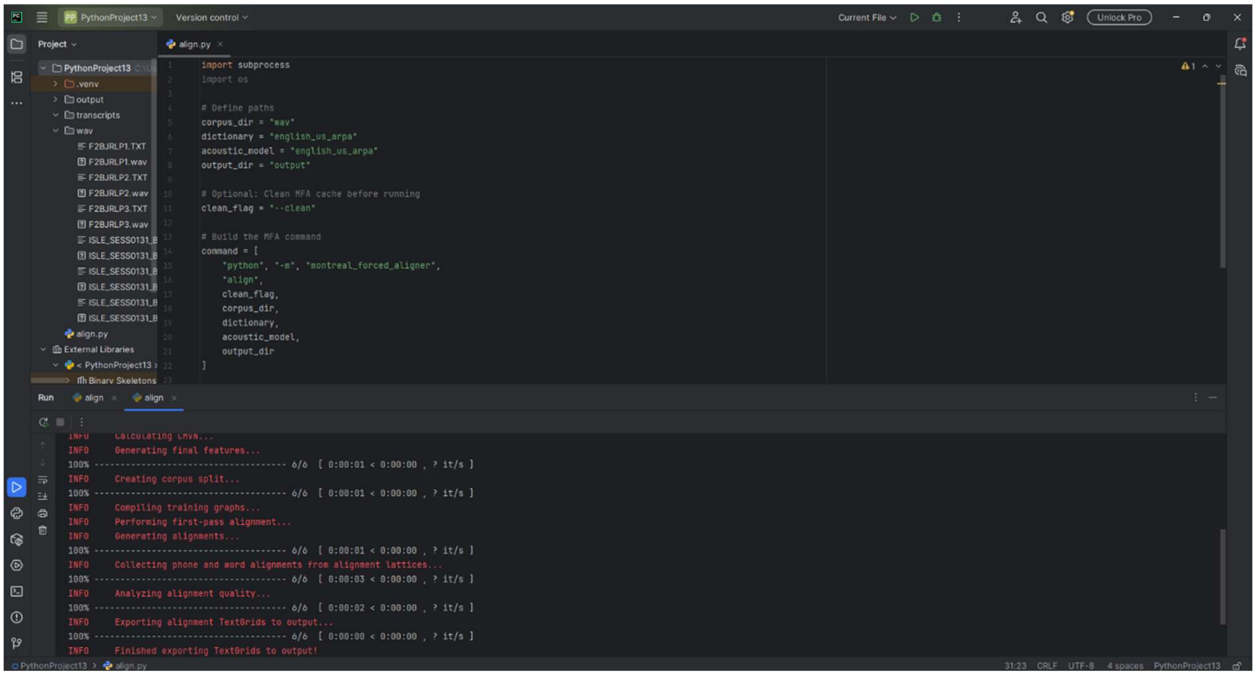
```
MFA model download dictionary english_us_arpa mfa model download acoustic
english_us_arpa
```

Alignment Execution

The alignment was performed using the following Python script (`align.py`):

```
import subprocess

import os
```

```
# Define paths
corpus_dir = "wav"
dictionary = "english_us_arpa"
acoustic_model = "english_us_arpa"
output_dir = "output"

# Optional: Clean MFA cache before running
clean_flag = "--clean"

# Build the MFA command
command = [
    "python", "-m", "montreal_forced_aligner",
    "align",
    clean_flag,
    corpus_dir,
    dictionary,
    acoustic_model,
    output_dir
]

# Run the alignment
try:
    print ("Running Montreal Forced Aligner...")
    subprocess.run(command, check=True)

    print ("Alignment complete. Check the 'output/' folder for TextGrid files.")

except subprocess.CalledProcessError as e:

    print (" Alignment failed. Check your paths and model setup.")

    print ("Error:", e)
```



```python
import subprocess
import os

# Define paths
corpus_dir = "wav"
dictionary = "english_us_arpa"
acoustic_model = "english_us_arpa"
output_dir = "output"

# Optional: Clean MFA cache before running
clean_flag = "--clean"

# Build the MFA command
command = [
    "python", "-m", "montreal_forced_aligner",
    "align",
    clean_flag,
    corpus_dir,
    dictionary,
    acoustic_model,
    output_dir
]
```

```
INFO     Calculating CMVN...
INFO     Generating final features...
100% ------------------------------------ 6/6  [ 0:00:01 < 0:00:00 , ? it/s ]
INFO     Creating corpus split...
100% ------------------------------------ 6/6  [ 0:00:01 < 0:00:00 , ? it/s ]
INFO     Compiling training graphs...
INFO     Performing first-pass alignment...
INFO     Generating alignments...
100% ------------------------------------ 6/6  [ 0:00:01 < 0:00:00 , ? it/s ]
INFO     Collecting phone and word alignments from alignment lattices...
100% ------------------------------------ 6/6  [ 0:00:03 < 0:00:00 , ? it/s ]
INFO     Analyzing alignment quality...
100% ------------------------------------ 6/6  [ 0:00:02 < 0:00:00 , ? it/s ]
INFO     Exporting alignment TextGrids to output...
100% ------------------------------------ 6/6  [ 0:00:00 < 0:00:00 , ? it/s ]
INFO     Finished exporting TextGrids to output!
```

MFA successfully aligned all six files and exported. TextGrid files to the `output` folder. The process completed in approximately 115 seconds.

Alignment Quality Analysis

The alignment analysis was exported as a CSV file. Key metrics include:

File	Duration (s)	Phone Deviation	SNR	Speech Log-Likelihood
F2BJRLP1	25.31	3.89	8.11	-45.95
F2BJRLP2	28.65	3.59	7.89	-45.83
F2BJRLP3	30.71	3.66	10.33	-46.88
ISLE_SESS0131_BLOCKD02_01_sprt1	4.13	2.72	11.85	-52.37
ISLE_SESS0131_BLOCKD02_02_sprt1	3.88	2.53	12.36	-50.91
ISLE_SESS0131_BLOCKD02_03_sprt1	4.50	3.88	11.44	-53.55

All files show acceptable alignment quality with good SNR and consistent timing.

Praat Inspection

Each. TextGrid file was opened in Praat using “View and Edit with Sound.” Word and phoneme boundaries were visually inspected against the waveform and spectrogram. No major misalignments were observed.

Conclusion

The forced alignment pipeline was successfully implemented using MFA. All audio files were aligned with their transcripts, and. TextGrid outputs were validated using Praat. The alignment quality metrics confirm that the process was accurate and reproducible.