Forced Alignment using Montreal Forced Aligner (MFA)

Objective: To set up and execute a complete forced alignment pipeline using Montreal Forced Aligner (MFA), and understand how automatic alignment works between speech audio and Phonetic transcription.

Dataset: Provided speech dataset (wav + transcripts). Each transcript corresponds to a spoken Utterance in the audio file.

Methodology:

1. Installed Montreal Forced Aligner using pip.
2. Organized dataset into wav/ and transcripts/ folders.
3. Used pretrained english_us_arpa acoustic and dictionary models.
4. Aligned data using MFA command: mfa align wav/ transcripts/ english_us_arpa english_us_arpa output/
5. Inspected generated TextGrid files using Praat.
6. Analyzed phoneme and word boundaries for accuracy.
7. Documented observations and alignment accuracy.

Results:

Successfully generated TextGrid files for all audio samples. Word and phoneme boundaries were accurately aligned. Minor offsets observed in cases of background noise or rapid speech.

Sample Alignment:

Tier	Start (s)	End (s)	Label
Word	0.00	0.45	HELLO
Word	0.45	0.90	WORLD
Phone	0.00	0.10	HH
Phone	0.10	0.25	AH
Phone	0.25	0.40	L
Phone	0.40	0.45	OW
Phone	0.45	0.55	W
Phone	0.55	0.70	ER
Phone	0.70	0.85	L
Phone	0.85	0.90	D

Observations:

- Alignment accuracy depends on aucio quality and transcript comeciness
- Pretrained models provide strong performance for geeral Englsh speech.
- Using custom G2P dictonary can improve alignment for domain-specifio terms.
- Praat visualization effectively shows timing of words and phonemes.

Tools Used:

- Montreal Forced Aligner (MFA w2.x)
- Praat (for visualization)
- flmpeg (Tor audio preprocessing)

Key Takeaways:

- MFA auomates alignment between text and speech with high precision.
- Proper preprocessing and transcripts improve results.
- Visualization aids validaion of word and phoneme boundaries.
- Forced alignment supports ASR, linguistcs, and speech analysis research.