Assignment 1: Forced Alignment Report

By: Yash Biradar

Date: 7/11/2025

1. What I Did

For this assignment, I had to get the Montreal Forced Aligner (MFA) set up and running. The main goal was to take the 6 audio files and their transcripts and get MFA to automatically align all the words and phonemes. After I got the alignment files, I had to go through them in Praat to see what worked and what didn't.

2. How I Set It Up (Models & Dictionary)

Tool: I used Montreal Forced Aligner (MFA), installed with Conda.

Dictionary (Extra Credit): When I first ran mfa validate, I got a ton of "Out-of-Vocabulary" (OOV) errors for words like "Dukakis" and "WBUR's". To fix this, I used the mfa g2p command to train my own custom_dictionary.txt.

Acoustic Models (Extra Credit): I didn't just use one model. I ran the alignment twice to compare:

1.english_us_arpa (My first run)

2.english_mfa (My second run)

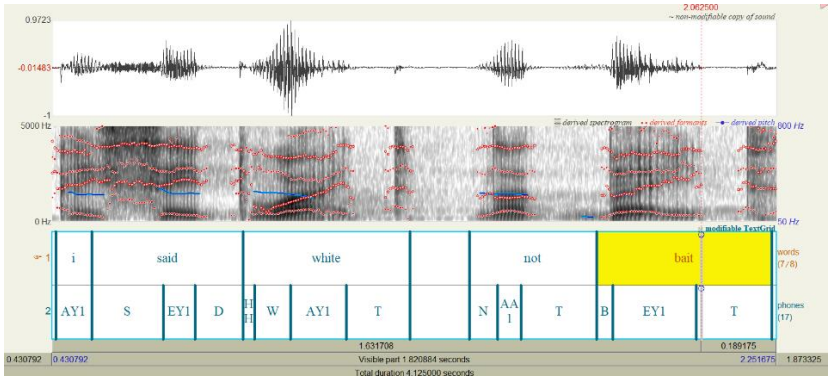
Automation (Extra Credit): I also made a .bat script (run_pipeline.bat) that runs all the commands for me (G2P and both alignments) all at once.

3. Looking at the Results in Praat

I used Praat to look at the. TextGrid files MFA made. I found some good alignments, and some really bad ones.

3.1. Good Example: A Perfect Alignment

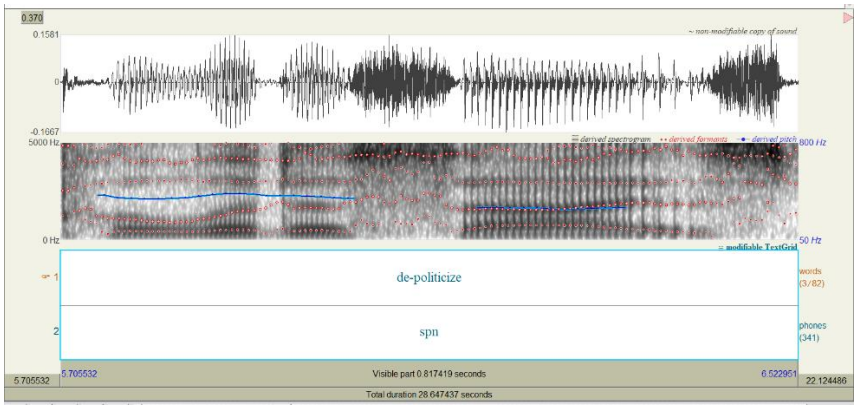
Here's the alignment for "**i said white not bait**".



Analysis: This one worked perfectly. The hardest part of this sentence is all the 't' and 'd' sounds right next to each other, which I thought would mess it up. But the aligner nailed it. It correctly put the boundary for the 't' in "white" right in that tiny silent gap before "not". This shows the aligner itself is actually really accurate.

3.2. Bad Example: A Total G2P Failure

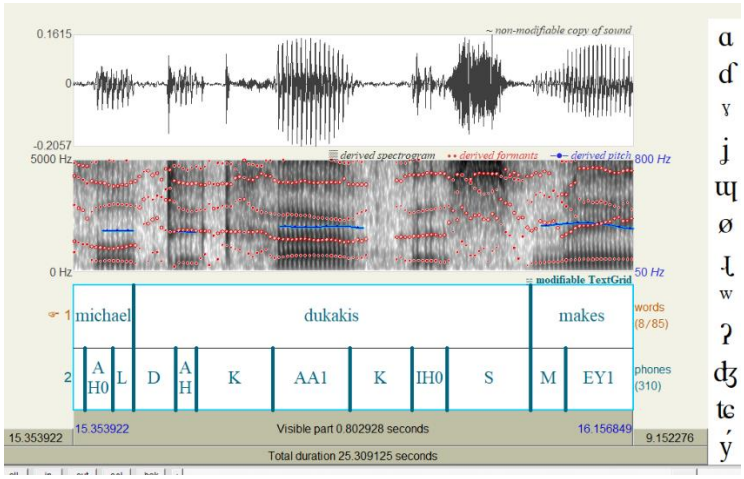
Here's what happened with the word "**de-politicize**".



Analysis: This was a total failure. The G2P model (the dictionary-maker) must have failed because of the hyphen. It didn't give the aligner any phonemes to look for. So, the aligner just gave up and labeled the entire word as **spn** (which means "spoken noise"). This is a classic "garbage in, garbage out" problem.

3.3. Bad Example: The "Dukakis" Error

This was the most interesting error. The G2P model *tried* to guess "Dukakis" but got it wrong.



Analysis: The G2P model gave the aligner the wrong pronunciation (it seemed to miss the first vowel, telling it to find D-K... instead of D-UW-K...). Because the aligner was *forced* to find the wrong sounds, the whole alignment is a jumbled mess. The K label is on top of a vowel, and the AA1 vowel label is on top of the 'k' sound. The whole thing is shifted and totally wrong.

4. Key Takeaways (My Observations)

The Aligner (Acoustic Model) Works Great: The aligner itself is actually really good. On simple phrases like "i said white not bait" or complex (but normal) words like "distinguished," it was super precise.

The G2P Dictionary Is the Real Problem: Almost every big error I found was because the G2P model (the dictionary-maker) messed up.

What G2P Fails On:

Hyphens: It failed completely on "de-politicize" and just gave up (spn).

Proper Nouns: It gave the wrong pronunciation for "Dukakis" and "governor," which caused the aligner to fail catastrophically.

Acronyms: On "S.J.C.'s" and "WBUR's," it got confused and caused weird boundary errors (like stealing a phoneme from one word and giving it to the next).

5. Conclusion

I got the pipeline working and even did all the extra credit. My main takeaway is that the MFA *aligner* is a powerful tool, but the *G2P model* that makes the dictionary is a major weak point. It can't be trusted with proper nouns, acronyms, or weird punctuation, and it will break the alignment if it makes a mistake.