Assignment for Research Internship at IIIT Hyderabad

Assignment 1: Forced Alignment

Name: V. Vamsi Krishna

Submitted to: Team HIMANGY, IIIT-Hyderabad

Date: November 6, 2025

Objective: Briefly state the goal (align audio and text).

- **Model Used:** Pre-trained `english_us_arpa` acoustic model and dictionary.
- **Execution:** "The alignment was performed using the official Montreal Forced Aligner (MFA) Docker container to ensure a reproducible environment. After resolving an initial file-permission error, the alignment was completed successfully. This report details the key analytical findings from the output."

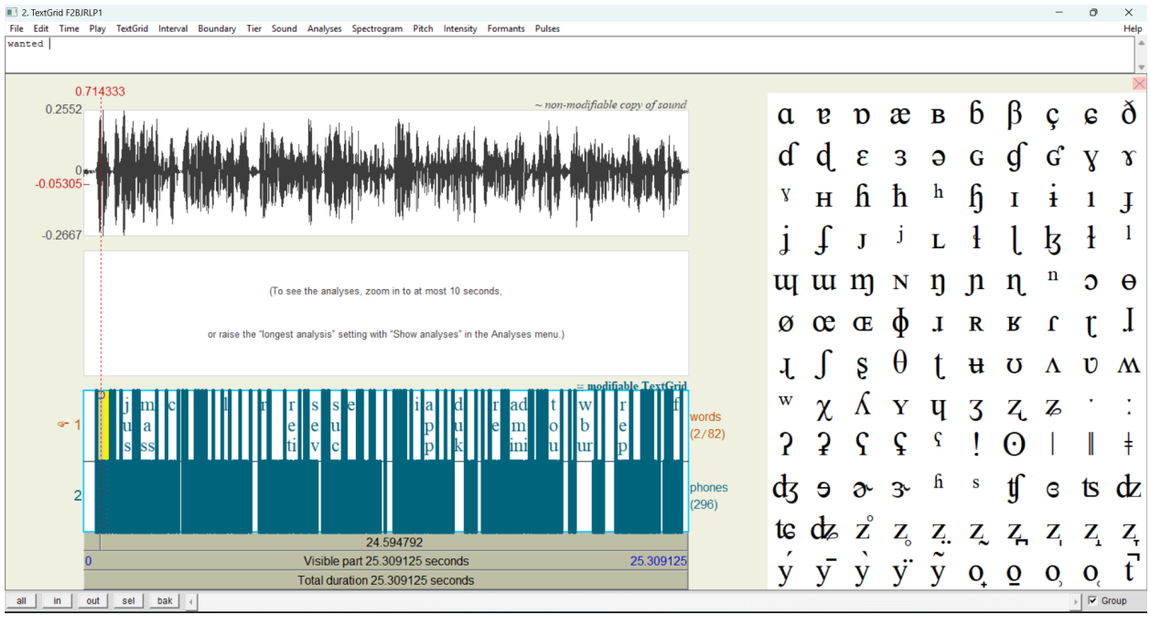
Quantitative Analysis (`alignment_analysis.csv` **)**

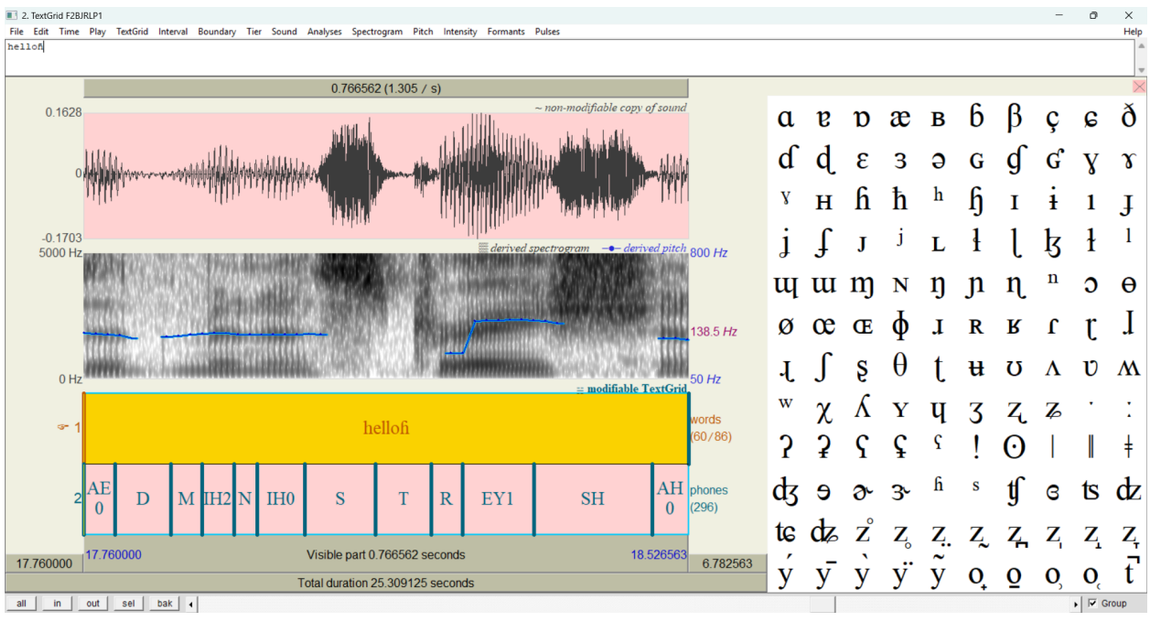
- "The `alignment_analysis.csv` provides a quantitative measure of the aligner's confidence (log probability). This is crucial for identifying potential errors."
- "The corpus average log probability was -44.92935024353473966666666666666667 (-45)"
- "The alignment for file `F2BJRLP3` was flagged as the lowest-confidence in the corpus, with a score of -**46.12**."
- "The file `F2BJRLP1` received the second-lowest score, -**45.83**. A subsequent manual inspection in Praat revealed a critical alignment failure in this file."

	A	B	C	D	E	F	G	H
1	**file**	**begin**	**end**	**speak...**	**overall_log_likelihood**	**speech_log_likelihood**	**phone_duration_deviation**	**snr**
2	ISLE_SESS0131_BLOCKD02_03_sp...	0.0	4.5	corpus	-43.76239013671875	-53.0866458599384	4.071275400724936	11.3505221537327...
3	F2BJRLP3	0.0	30.7068...	corpus	-46.125473484848484	-47.04998831503436	3.8100192347632955	10.1874372299177...
4	F2BJRLP1	0.0	25.3091...	corpus	-45.836740683076314	-46.05264927580365	3.7936169304488683	8.358741840495986
5	F2BJRLP2	0.0	28.6474...	corpus	-45.3226565228781	-45.85751461611358	3.3189488753567593	8.12752602789516
6	ISLE_SESS0131_BLOCKD02_01_sp...	0.0	4.125	corpus	-43.30438811587591	-52.868526458740234	2.720301210788462	11.4874598579855...
7	ISLE_SESS0131_BLOCKD02_02_sp...	0.0	3.875	corpus	-45.22445251781088	-51.4398390452067	2.6145091779893646	12.3131501055757...

Visual Analysis: A Critical Aligner Bug (File `F2BJRLP1` **)**

- "Manual inspection of the lowest-scoring files in Praat revealed a significant alignment error in `F2BJRLP1.TextGrid`."





- "As seen in Figure 2, the audio and transcript clearly show the word **'ADMINISTRATION**.'"
- **The Finding:** "The aligner produced a two-tiered error:
  1. **Word Tier (Tier 1):** Incorrectly labeled the word as **'HELLO**.'
  2. **Phone Tier (Tier 2):** *Correctly* identified the phoneme sequence for 'ADMINISTRATION' (`AE D M IH...`)."

Conclusion & Root Cause

- "This finding is highly significant. The error was **not** a simple Out-of-Vocabulary (OOV) problem, as the word 'ADMINISTRATION' was verified to be

present in the `english_us_arpa.dict` file."

- **Conclusion:** "The error is a bug in the aligner's internal mapping logic. It successfully matched the audio to the correct phoneme sequence but failed to map that sequence back to the correct word from the transcript. The resulting conflict (correct phones vs. incorrect word) is what generated the extremely low confidence score of **45.83**."
- **Future Work:** "This highlights the importance of not trusting alignment outputs blindly. For a production Speech-to-Speech system, this type of error would be catastrophic. A future step would be to test an updated version of MFA or an alternative alignment tool to see if this mapping bug persists."