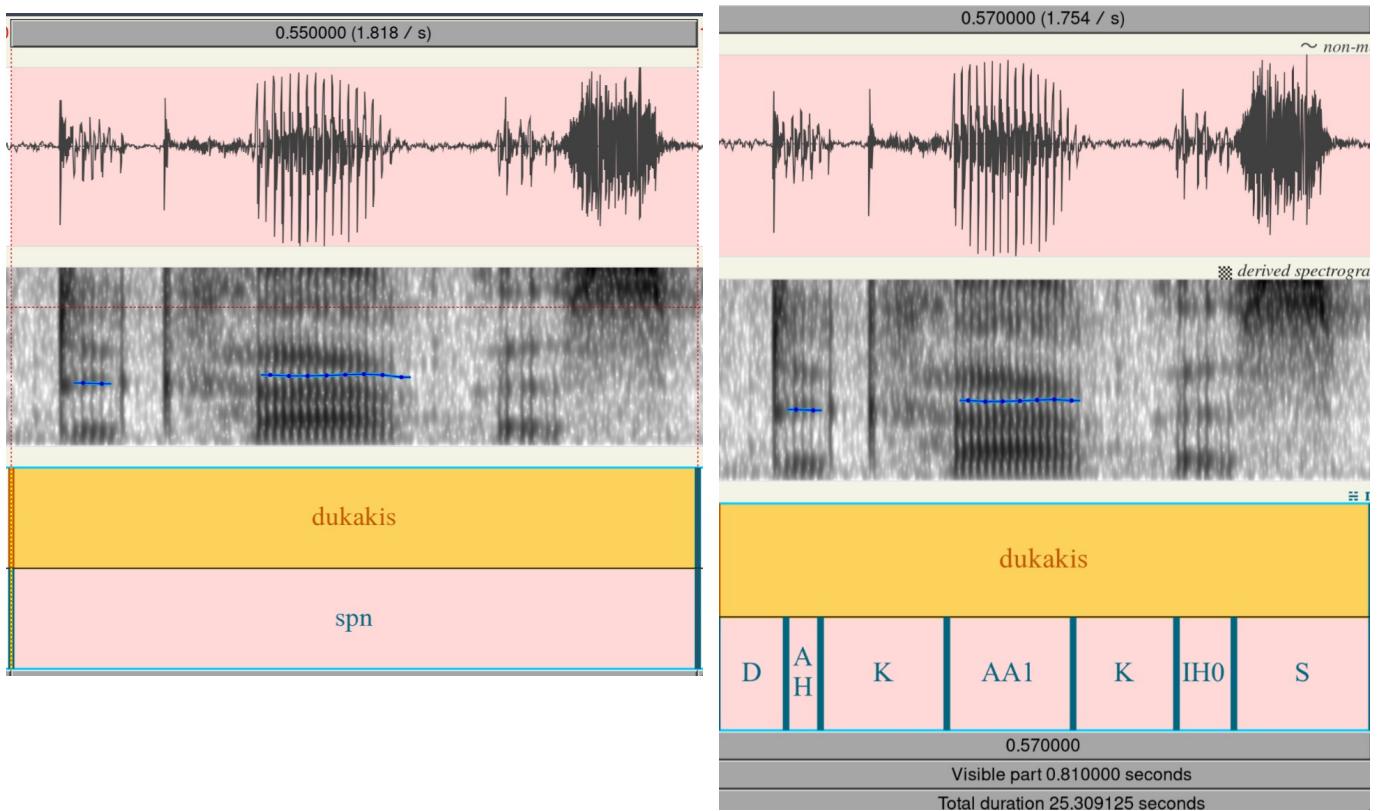


## Extra Credit Solution:

### 1. Custom Dictionary:

First we start by training a custom dictionary using g2p model:



the five words which weren't aligned before now get aligned with th help of the new dictionary.

Ex: screenshots for “Dukakis” before and after are given above.

Similarly all 5 words that weren't aligned before are found in the new dictionary.

So training a custom dictionary was a success.

### 2. Trying other acoustic models:

So there was only one other acoustic english model that I could try which was english\_mfa

acoustic model: english\_mfa

corresponding g2p for custom dict: english\_us\_mfa

I couldn't eyeball any differences on praat but the alignment\_analysis csv file basically tells me that the log\_likelihood scores for english\_mfa model is slightly higher in some cases but lower in other cases, so inconclusive comparison really.

But I have attached outputs of the textgrids of english\_mfa model aswell on github.

### **3. Automating the pipeline:**

Script used :

```
#!/usr/bin/env bash
#
# Automates the full MFA alignment pipeline for the assignment.
#
# Exit immediately if any command fails
set -e

echo "===== STARTING ALIGNMENT PIPELINE ====="

# --- 1. Paths and Model Names ---
INPUT_DIR="/home/somesh/Desktop/input"
MODEL_DIR="/home/somesh/Documents/MFA/pretrained_models"
OUTPUT_DIR="/home/somesh/Desktop/output"

ACOUSTIC_MODEL="english_us_arpa"
G2P_MODEL="english_us_arpa"
CUSTOM_DICT="custom_dictionary.dict"

# --- 2. Activate Conda Environment ---
echo "Activating Conda environment 'mfa'..."
source /opt/miniconda3/etc/profile.d/conda.sh
conda activate mfa

# --- 3. Download Required Models ---
echo "Downloading required models (if not present)..."
mfa model download acoustic ${ACOUSTIC_MODEL}
mfa model download g2p ${G2P_MODEL}

# --- 4. Generate Custom Dictionary ---
echo "Generating custom dictionary: ${CUSTOM_DICT}..."
mfa g2p \
    ${INPUT_DIR} \
    ${MODEL_DIR}/g2p/${G2P_MODEL}.zip \
    ${MODEL_DIR}/dictionary/${CUSTOM_DICT} \
    --clean

# --- 5. Run Alignment ---
echo "Running alignment with ${ACOUSTIC_MODEL} model..."
mfa align \
    ${INPUT_DIR} \
    ${MODEL_DIR}/dictionary/${CUSTOM_DICT} \
    ${MODEL_DIR}/acoustic/${ACOUSTIC_MODEL}.zip \
    ${OUTPUT_DIR} \
```

```
--clean
```

```
# --- 6. Finalization ---
echo ""
echo "===== AUTOMATION COMPLETE ====="
echo "Alignment TextGrids are in: ${OUTPUT_DIR}"
```

### My output screenshot from running the script:

```
<frozen importlib._bootstrap>:488: Warning: OpenSSL 3's legacy provider failed to load. Legacy algorithms will not be available. If you need those algorithms, check your OpenSSL configuration.
Downloading required models (if not present) ...
Local version of model already exists (/home/somesh/Documents/MFA/pretrained_models/acoustic/english_us_arpa.zip). Use the --ignore_cache flag to force redownloading.
Local version of model already exists (/home/somesh/Documents/MFA/pretrained_models/g2p/english_us_arpa.zip). Use the --ignore_cache flag to force redownloading.
Generating custom dictionary: custom_dictionary.dict ...
INFO      Setting up corpus information ...
INFO      Loading corpus from source files ...
100% ━━━━━━━━━━━━━━━━━━━━━━━━ 6/1 [ 0:00:02 < 0:00:00 , ? it/s
INFO      Found 1 speaker across 6 files, average number of utterances per speaker: 6.0
INFO      Initializing multiprocessing jobs ...
WARNING   Number of jobs was specified as 3, but due to only having 1 speakers, MFA will only use 1 jobs. Use the --single Speaker flag if you would like to split utterances across jobs regardless of their speaker.
INFO      Normalizing text ...
100% ━━━━━━━━━━━━━━━━ 6/6 [ 0:00:01 < 0:00:00 , ? it/s
INFO      Generating pronunciations ...
100% ━━━━━━━━━━━━━━ 149/149 [ 0:00:46 < 0:00:00 , 4 it/s
INFO      Done! Everything took 50.498 seconds
Running alignment with english_us_arpa model ...
INFO      Setting up corpus information ...
INFO      Loading corpus from source files ...
6% ━━━━ 6/100 [ 0:00:02 < -:--:-- , ? it/s
INFO      Found 1 speaker across 6 files, average number of utterances per speaker: 6.0
INFO      Initializing multiprocessing jobs ...
WARNING   Number of jobs was specified as 3, but due to only having 1 speakers, MFA will only use 1 jobs. Use the --single Speaker flag if you would like to split utterances across jobs regardless of their speaker.
INFO      Normalizing text ...
100% ━━━━━━━━━━━━━━ 6/6 [ 0:00:01 < 0:00:00 , ? it/s
INFO      Generating MFCCs ...
100% ━━━━━━━━━━━━ 6/6 [ 0:00:01 < 0:00:00 , ? it/s
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 [ 0:00:01 < 0:00:00 , ? it/s
INFO      Analyzing alignment quality ...
100% ━━━━━━ 6/6 [ 0:00:01 < 0:00:00 , ? it/s
INFO      Exporting alignment TextGrids to /home/somesh/Desktop/output ...
100% ━━━━━━ 6/6 [ 0:00:00 < 0:00:00 , ? it/s
INFO      Finished exporting TextGrids to /home/somesh/Desktop/output!
INFO      Done! Everything took 12.256 seconds

===== AUTOMATION COMPLETE =====
Alignment TextGrids are in: /home/somesh/Desktop/output
```