

Forced Alignment Using Montreal Forced Aligner (MFA)

Project Date: Thursday, November 06, 2025

1. Environment & MFA Installation

- Installed **Anaconda/Miniconda** to manage packages and create isolated environments.
- Created and activated an MFA environment using:

```
conda config --add channels conda-forge
conda create -n aligner -c conda-forge montreal-forced-aligner
conda activate aligner
```

- Verified installation using `mfa --help` to see available MFA commands.

2. Downloaded Models and Dictionaries

- Downloaded the pretrained **English acoustic model** and **ARPA pronunciation dictionary** with:

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

- This ensured the alignment could use standard US English phonetic mapping.

3. Data Preparation

- Organized data with `.wav` audio and matching `.txt` transcript files placed in `C:\MFA\corpus`.
- Confirmed all filenames matched exactly and that every audio file had its transcript.
- Created an output directory `C:\MFA\output` for alignment results.

4. Forced Alignment Pipeline

- Initial attempts to run alignment led to SQLite database errors:

```
mfa align C:\MFA\corpus english_us_arpa english_us_arpa C:\MFA\output
```

- Received error: `extra argument: "word_interval_temp"`. This is a known Windows SQLite command-line incompatibility with MFA.
- Solution: Used PostgreSQL backend, which MFA supports for Windows reliability:

```
mfa align "C:\MFA\corpus" english_us_arpa english_us_arpa "C:\MFA\output" --  
use_postgres --auto_server
```

- The alignment completed and generated `.TextGrid` output files in the output directory.

5. Troubleshooting Summary

- **Initial Error:** `extra argument: "word_interval_temp"` due to SQLite bug on Windows.
- **Fix:** Switched MFA to the PostgreSQL backend using `--use_postgres --auto_server`, which resolved the error and allowed the alignment to complete.
- Occasional warnings about number of jobs due to single speaker setup (not critical).

6. Output Analysis

- Opened `.TextGrid` files in **Praat** by loading both the `.wav` audio and its corresponding TextGrid.
- Inspected word and phone alignment tiers:
 - Most boundaries matched the spoken audio closely.
 - Minor timing offsets observed, mainly in rapid speech.
- Alignment is suitable for research or language technology use. All files were successfully exported.

7. Key Observations & Report Documentation

- **Models Used:** `english_us_arpa` (acoustic & dictionary)
- **Workflow:** Environment setup, model download, data organization, alignment run (PostgreSQL backend), output inspection.
- **Known Issues:** Windows requires PostgreSQL backend to avoid SQLite errors.
- **Quality:** Alignment boundaries generally accurate.

8. Optional/Extra Credit

- *Not attempted* in this run: Training a custom dictionary or using alternate acoustic models. (For extra credit, use `mfa train_g2p` for custom G2P or try automation scripts.)

Conclusion:

- The complete forced alignment pipeline was successfully executed, with all major setup, troubleshooting, and analysis steps documented for reproducible research or academic use.