MFA: Train-and-Align

You can use 'MFA' to train new speaker-specific acoustic models on your speech data, and then to force align the data on those models. You may decide to do this if:

- You can't share your data with third parties and so can't use WebMAUS.
- MFA has no dictionaries and pre-trained acoustic models for the language of your data and so you can't use MFA and Pretrained Acoustic Models.
- You have 3-5 hours' speech.

The general process is illustrated in Figure 1.

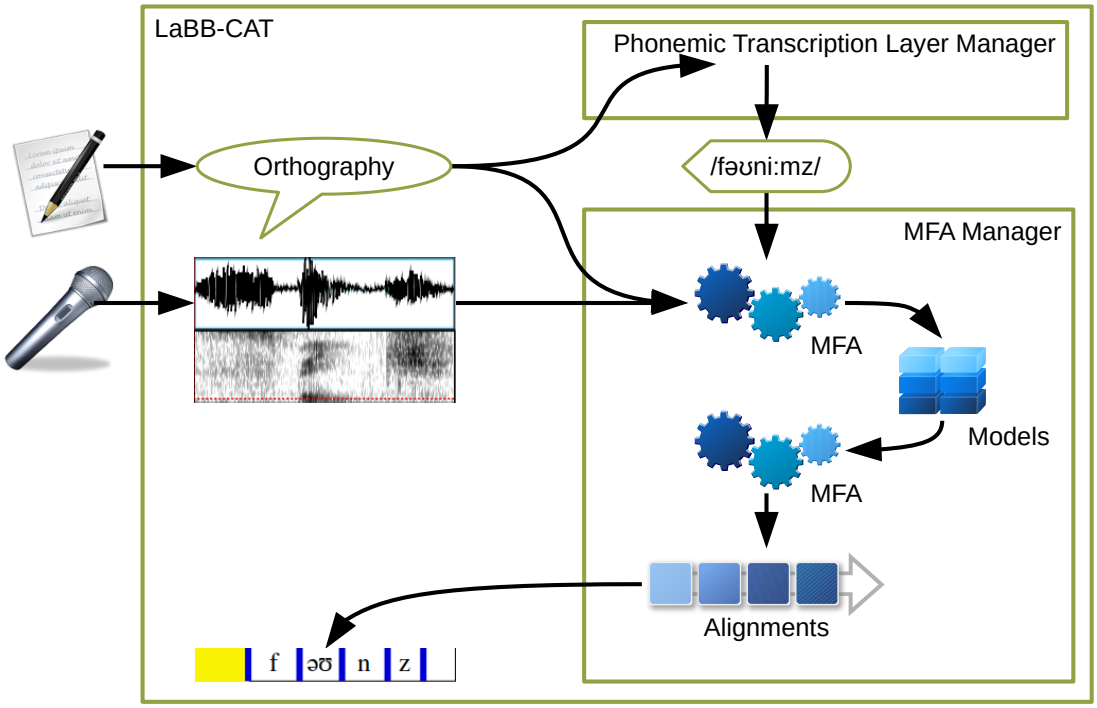

Figure 1: Pronunciations are generated from transcripts, and then combined with the recordings to train acoustic models, which are then used to compute phone-level alignements, which are saved to LaBB-CAT's database

Prerequisites

In order to be able to force-align transcripts to the word and/or segment level, you first need the following:

1. Transcripts that are aligned at the utterance level (i.e. there's a known time-point every 20 or so words), whch have been uploaded into LaBB-CAT
2. A WAV file for each transcript, on the LaBB-CAT server
3. A phonemic transcription word layer, that has at least one pronunciation for every word. If there are some lines/utterances that contain words with missing pronunciations, those lines will be ignored by the HTK Layer Manager.

Depending on your speech data, there are several ways to obtain phonemic transcriptions for words:

- Lexical tagging
  - CELEX - for British English, German, Dutch, using one of the CELEX layer managers.
  - CMU Pronouncing Dictionary - for US English, using th CMU Pronouncing Dictionary layer manager.
  - Unisyn - for various English varieties, using the Unisyn layer manager.
  - Define your own lexicon, and use the Flat File Dictionary layer manager to integrate it into LaBB-CAT.
- Inferring pronunciation from orthography
  - Spanish, using the Spanish Phonological Transcriber layer manager
  - Bas Web Service: G2P - for various languages.
  - Define your own simple mapping rules from orthography to phonology, using the Character Mapper layer manager.

If the speech corpus includes data in more than one language, it is possible to ensure that the utterances are phonemically tagged in a way that's sensitive to the language of the specific utterance, using the *language* layers and attributes, and auxiliary layer managers.

Whichever method you choose, you need a phonemes 'word layer' on which each word token is tagged with its pronunciation, before you can proceed with the forced-alignment steps below.

Procedure for MFA Forced Alignment

The broad steps for getting forced-alignments from MFA are:

1. Install MFA on the same computer that LaBB-CAT is installed on

2. Install the MFA Layer Manager, which integrates LaBB-CAT with MFA
3. Create and configure a new MFA layer in LaBB-CAT
4. Pick a speakers/participants in your database and identify their utterances
5. Fill in the missing pronunciations for those utterances
6. Run forced alignment
7. Repeat steps 4-6 for all the participants in your database

MFA Installation

MFA is not included as part of LaBB-CAT, and so it must be installed on the server you have installed LaBB-CAT on before you can integrate LaBB-CAT with it.

If MFA has not been installed already, please follow the following steps, depending on the operatings system of your LaBB-CAT server:

Linux

To install the Montreal Forced Aligner on Linux systems for all users, so that your web server can access it if required:

1. Download Miniconda:
   ```
   wget https://repo.anaconda.com/miniconda/Miniconda3-py38\_4.10.3-Linux-x86\_64.sh
   ```
2. Start the installer:
   ```
   sudo bash Miniconda3-py38\_4.10.3-Linux-x86\_64.sh
   ```
3. When asked the location to install Miniconda, use:
   ```
   /opt/conda
   ```
4. When asked whether the installer should initialize Miniconda, this is unnecessary so you can respond `no`
5. Change ownership of the conda files):
   ```
   sudo chown -R $USERNAME:$USERNAME /opt/conda
   ```
6. Make conda accessible to all users (so you web server can access MFA):
   ```
   chmod -R go-w /opt/conda
   chmod -R go+rX /opt/conda
   ```
7. Install the Montreal Forced Aligner
   ```
   /opt/conda/bin/conda create -n aligner -c conda-forge
   montreal-forced-aligner=2.2.17
   ```

Windows

To install the Montreal Forced Aligner on Windows systems for all users, so that your web server can access it if required:

1. Download the Miniconda installer: https://repo.anaconda.com/miniconda/Miniconda3-latest-Windows-x86_64.exe
2. Start the installer by double-clicking it.
3. When asked, select the "Install for all users" option. This will install conda somewhere like *C:\ProgramData\Miniconda3*
4. When asked, tick the *add to PATH* option.
5. Install the Montreal Forced Aligner by specifying a path to the environment
   ```
   conda create -c conda-forge -p C:\ProgramData\Miniconda3\envs\aligner
   montreal-forced-aligner=2.2.17
   ```

Windows Troubleshooting

The 3rd party MFA software requires:

- the possibility of running command-line programs during installation and forced alignement
- the possibility that these programs can download data from the internet

On Windows, this can sometimes be complicated by the fact that Apache Tomcat and LaBB-CAT are installed as a 'Windows Service'. Windows Services usually run using the permissions of a special anonymous login account called 'Local System', which in some environments has restricted permissions to access different resources.
If you install the MFA Manager LaBB-CAT integration module, but you find it returns errors when trying to interact with MFA, the problem may be that the Windows Service:

- does not have permission to access the folder where MFA is installed, or
- is not allowed to execute other programs, or
- cannot access the internet.

Sometimes problems can be resolved by:

- running the Apache Tomcat Windows Service as a different user other than 'Local System'. (or if it was running as some other used, try setting it back to 'Local System'), or
- adjusting the permissions of the Windows Service users, or
- adjusting the permissions of the folders where MFA is installed

- configuring the service to use the local Internet Proxy settings to enable connecting to the internet.

PSexec is a tool that can be used to diagnose and solve problems on Windows.

PSexec

1. Download PStool.zip from Microsoft:
   http://technet.microsoft.com/en-us/sysinternals/bb897553.aspx
2. Unzip it
3. Put *PSexec.exe* into `C:\Windows\System32`
4. Open `cmd` using "Run as Administrator"
5. Run the command:
   `Psexec.exe -i -s cmd.exe`
   This opens a new command prompt for local system account
6. In the new command prompt window, check you have the correct account type with the command:
   `whomai`

Then you can use the command prompt to run MFA commands to diagnose errors - e.g.:

- `conda activate montreal-forced-aligner` - activates the MFA environment
- `mfa version` - ensures MFA is installed and accessible, and confirms the version
- `mfa model download dictionary` - ensures MFA can connect to the internet to get models etc.; this command should return a long list of language dictionaries, and not report errors.

Proxy Settings

To update proxy server settings:

1. type `inetcpl.cpl`
2. goto *Connections* tab
3. click on the *LAN Settings* button
4. Fill in the *Proxy* section with the correct details

i Docker Container

If your LaBB-CAT server is installed in a Docker Container, it can download and install Miniconda and MFA itself, as part of the process of installing the MFA Manager LaBB-CAT module.

> There is no need for a separate installation of the MFA software.

Layer Manager Installation

Once MFA has been installed, you have to install the MFA Manager, which is the LaBB-CAT module that provides MFA with all the data it needs, and then saves to alignments MFA produces back to your database.

1. Select the *layer managers* menu option.
2. Follow the *List of layer managers that are not yet installed* link.
3. Find *MFA Manager* in the list, and press its *Install* button and then press *Install* again. As long as MFA has been installed for all users, you should see a box that's already filled in with the location that MFA was installed to.
4. Click *Configure* to continue the layer manager installation.
   You will see a window open with some information about integrating with MFA, including the information you've already read above.

Dictionary and segment layer labels

The labels used for phonemes layer (or whichever layer tags each word token with its pronunciation) will use a specific encoding for the phonemes. Encodings include:

- CELEX DISC: Exactly one ASCII character per phoneme,
  e.g. there'll → `D8r@l`
- Unicode IPA: One or more Unicode character per phoneme, possibly including diacritics, delimited by spaces:
  e.g. there'll → `ð ə l`
- ARPAbet: Phonemes are one or two uppercase ASCII characters, possibly suffixed with a digit indicating stress, delimited by spaces:
  e.g. there'll → `DH EH1 R AX0 L`

If it uses CELEX DISC encoding, the phonemes layer should have its *Type* set to *Phonological* on the word layers page. Otherwise its *Type* should be set to *Text*.

In order to ensure that the labels that the *MFA Manager* will create on the segment layer use the same encoding, the segment layer must have the same *Type* as the *phonemes* layer. In order to ensure that:

1. Select the *segment layers* option on the menu.
2. The *segment* layer is the first on the list (and may be the only layer there).
3. Check the Type of the segment layer. If it's not the same as the *phonemes* layer, change the *Type* so that it matches, and press the *Save* button that appears.

Create the MFA layer

Once you've installed MFA and the MFA Layer Manager, you need to create a new layer for triggering and controlling forced alignment. This layer will itself contain a timestamp for each line/utterance it has force-aligned (and so it's a 'phrase' layer), but during that process, the word and phone alignments will also be set on other layers.

1. Select the *phrase layers* option on the menu
2. Fill in the form at the top of the page (which doubles as column headings) with the following details:
    - **Layer ID**: `mfa`
    - **Type**: *Text*
    - **Alignment**: *Intervals*
    - **Manager**: *MFA Manager*
    - **Generate**: *never* (this is because we will manually select utterance for forced alignment, to ensure there is enough data to train acoustic models)
    - **Description**: *MFA alignment time*
3. When you configure the layer, set the following options:
    - **Pronunciation Layer**: select the *phonemes* layer (or whichever word layer that tags each word with its pronunciation)
    - **Dictionary Name**: *[none]*
    - **Pretrained Acoustic Models**: *[none]* (this ensures that the train/align procedure is used)
    - The rest of the options can be left as their default values.
    - If you're curious about what the configuration options do, hover your mouse over each option to see a 'tool tip' that describes what the option is for.
4. Press *Set Parameters*

Batch Alignment

MFA is data-hungry when training acoustic models for forced alignment, and needs 3-5 hours' speech

To start a forced-alignment process for a batch of selected participants, you need to first select the participants who will be aligned. Then you need to list all their utterances, and MFA will first training acoustic models from scratch from them, and then using those acoustic models, align those same utterances.

1. In LaBB-CAT, select the *participants* link on the menu
2. Filter the list to display the desired participants, and tick the checkbox next to each participant you want to include

3. Press the *All Utterances* button above
4. Press *List* to list all of their utterances.
   A progress bar will appear while LaBB-CAT identifies all the selected participant's utterances. Once this is done, the first twenty utterances will be listed (like search results, the first twenty are listed for convenience, but you can process all matching utterances)
5. Click the *Mfa* button below the list.
   A progress bar will appear while MFA gathers up all the utterance data, trains acoustic models, force-aligns the utterances, and then saves the resulting alignments back to LaBB-CAT. This process may take some time.

Once the progress bar reaches 100% and the process is complete, the selected utterances will have word start/end times set, and aligned phones will have been added to the *segment* layer, using the same labels as appear in the *phonemes* layer