

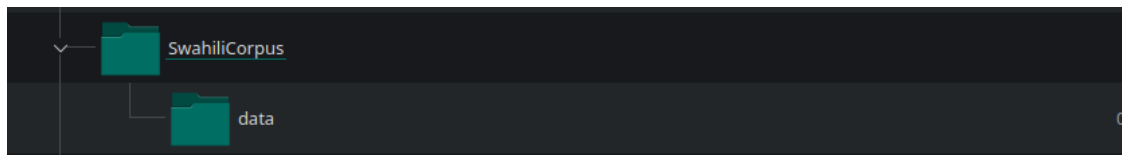
Use Case 4: Alignment of Swahili broadcast speech

In this tutorial we will align Swahili broadcast speech data. The data used here can be found in the UseCase4-Swahili_broadcast_data.zip. This is a subset of the ALFFA speech database for Swahili (Gelas et al. 2012). The entire dataset can be found here: <http://openslr.org/25/>

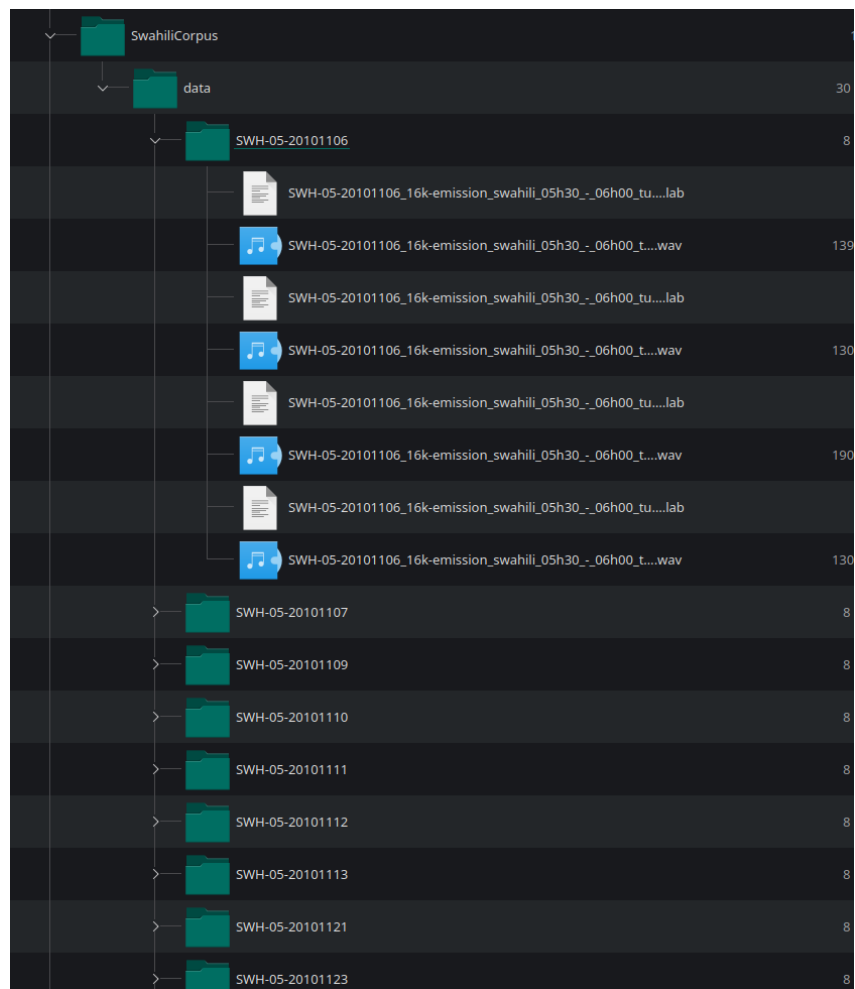
The data in the zip folder is already prepared and contains .wav and .lab files.

1) Create a corpus folder

As a first step, create a folder that will contain the data of the corpus. In this case, I used the name “SwahiliCorpus”.



Extract the data into the folder. The folder structure should look like this:



2) Activate your MFA environment

Open your terminal/console/command line and activate the environment, in which you installed the MFA.

3) Download the acoustic model and dictionary for Swahili

The pretrained acoustic model and dictionary for Swahili can be found here: <https://mfa-models.readthedocs.io/en/latest/>. Download them by using the following commands:

```
[ ]: mfa model download acoustic swahili_mfa
     mfa model download dictionary swahili_mfa
```

4) Align the Swahili data

Align the data by running the following command:

```
[ ]: mfa align ~/Documents/SwahiliCorpus/data/ swahili_mfa swahili_mfa ~/Documents/
     ↪SwahiliCorpus/alignments/
```

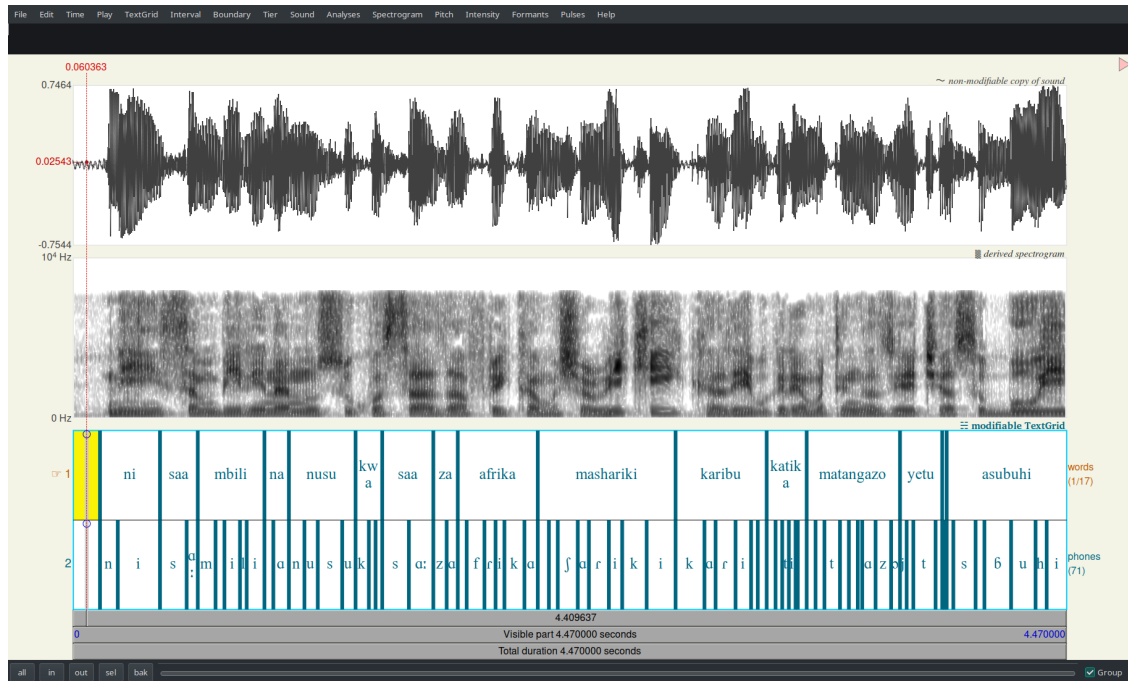
Here is an example for all command on Windows:



```

(base) C:\Users\p>conda activate myMFA
(myMFA) C:\Users\p>mfa model download acoustic swahili_mfa
(myMFA) C:\Users\p>mfa model download dictionary swahili_mfa
(myMFA) C:\Users\p>mfa align .\clean C:\Users\p\Documents\SwahiliCorpus\data\ swahili_mfa swahili_mfa C:\Users\p\Documents\SwahiliCorpus\alignments/
INFO - Setting up corpus information...
INFO - Loading corpus from source files...
INFO [9849, 21.00it/s]
INFO Found 30 speakers across 120 files, average number of utterances per speaker: 4.0
INFO - Initializing multiprocessing jobs...
INFO - Normalizing text...
100% | 120/120 [00:04<00:00, 28.46it/s]
INFO - Creating corpus split for feature generation...
100% | 240/240 [00:02<00:00, 82.22it/s]
INFO - Generating MFCCs...
122it [00:04, 26.57it/s]
INFO - Calculating GMM...
INFO - Generating final features...
100% | 120/120 [00:02<00:00, 41.43it/s]
INFO - Creating corpus split with features...
100% | 120/120 [00:03<00:00, 38.73it/s]
INFO - Compiling training graphs...
100% | 120/120 [00:03<00:00, 39.39it/s]
INFO - Performing first-pass alignment...
INFO - Generating alignments...
100% | 120/120 [00:04<00:00, 29.72it/s]
INFO - Calculating FMLLR for speaker adaptation...
100% | 30/30 [00:03<00:00, 8.88it/s]
INFO - Performing second-pass alignment...
INFO - Generating alignments...
99% | 119/120 [00:03<00:00, 31.07it/s]
99% | 0/120 [00:00?, 71it/s]
INFO - Collecting phone and word alignments from alignment lattices...
99% | 119/120 [00:03<00:00, 32.05it/s]
INFO - Exporting alignment TextGrids to C:\Users\p\Documents\SwahiliCorpus\alignments...
100% | 120/120 [00:04<00:00, 28.74it/s]
INFO - Finished exporting TextGrids to C:\Users\p\Documents\SwahiliCorpus\alignments!
INFO - Done! Everything took 58.831 seconds
(myMFA) C:\Users\p>
```

You should see now the TextGrids in the alignments directory. Here is an example of the alignments:



References

Gelas, H., Besacier, L., & Pellegrino, F. (2012). Developments of Swahili resources for an automatic speech recognition system. Workshop on Spoken Language Technologies for Under-resourced Languages.