Forced Alignment Workshop

Patrycja Strycharczuk

28/02 – 01/03/2022

About this workshop

This workshop is an introduction to forced alignment, using web-based interfaces. Forced alignment is a technique for detecting segment boundaries in audio files, using orthographic transcription provided by the user. You will learn how to analyse segment duration and vowel formants in audio recordings, using two online interfaces: DARLA and MAUS. Specifically, the topics we'll cover include:

- preparing transcription and audio files
- running the forced alignment
- formant analysis
- comparison between DARLA and MAUS functionality
- handling larger corpora and multiple files

This workshop is intended for beginners who have no prior experience with forced alignment, and it will focus on web-based FA interfaces. If you have used forced alignment before and are looking for advice on more advanced functions (e.g. training a forced aligner on a new language, adding dictionaries), this workshop is not for you.

Before attending, you should install Praat, which can be downloaded here. Some previous familiarity with Praat is an advantage, but not necessary for participation.

How forced alignment works

- orthographic transcription is converted to phonemic one, based on a dictionary
- the audio signal is matched to the phones, based on statistical information about the phone realisation in a pre-trained model

DARLA

Online interface for Montreal Forced Aligner (MFA)

URL: http://darla.dartmouth.edu/semi

Input: WAV file + corresponding TextGrid (recommended) or TXT file with transcription Output: fully aligned TextGrid and a CSV file with segment boundaries and formant values at midpoint

Languages: English (US, but works reasonably well on other varieties of English)

Transcription system: ARPABET

Things to look out for:

- no punctuation
- in the TextGrid, the transcription should be placed in a tier called 'sentence'

MAUS

URL: https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic

Input:

- for WebMAUS Basic: WAV file + TXT file with transcription (multiple pairs of files can be uploaded at once);
- for WebMAUS General: WAV file + TextGrid (this requires an extra step of G2P preparation)
  - G2P requires you to specify the language and the sampling rate. You can find out the sampling rate of your recording by selecting the sound file in Praat Objects and then ‘Query’ > ‘Query time sampling’ > ‘Get sampling frequency’
  - select tg as Input format
  - make sure that you specify the correct name for the Input TextGrid tier (it has to match the name of your annotation tier in your TextGrid file)
  - select x-sampa as the Output symbol inventory
  - you can leave the remaining settings unchanged

Output: fully aligned TextGrid, and optionally: EMU database, CSV file with formants, formant plots

Languages: multiple

Transcription system: X-SAMPA

Transcription guide for vowels

keyword	arpabet	xsampa
FLEECE	IY	i
KIT	IH	I
FACE	EY	eI
DRESS	EH	E
TRAP	AE	{
LOT	AA	A
THOUGHT	AO	O
STRUT	AH	V
GOAT	OW	oU
FOOT	UH	U
GOOSE	UW	u
PRICE	AY	aI
MOUTH	AW	aU
CHOICE	OY	OI
NURSE	ER	3‘

Good practice when working with forced alignment

1. You will reduce errors if you pre-segment the file yourself, inputting orthographic transcription. Using a text file with transcription as input might work, but you are likely to run into problems with longer sound files, especially with spontaneous speech.

2. Always *check the output* and either make corrections, or clean your data set to get rid of erroneous values.

3. Consider the differences between the variety you’re working with and a reference variety. Forced alignment has been shown to work relatively well on regional varieties, but the reference transcription may not match the variety you’re working with.

4. Both DARLA and MAUS have very informative Help and FAQ sections. Use them for troubleshooting if something is not working with your alignment.

5. Make sure you cite the forced alignment properly. See:

- http://darla.dartmouth.edu/cite for DARLA
- https://clarin.phonetik.uni-muenchen.de/BASWebServices/publications for MAUS

Some relevant sources

- Gonzalez, S., Grama, J., & Travis, C. E. (2020). Comparing the performance of forced aligners used in sociophonetic research. *Linguistics Vanguard*, 6(1).

- Mahr, T. J., Berisha, V., Kawabata, K., Liss, J., & Hustad, K. C. (2021). Performance of forced-alignment algorithms on children's speech. *Journal of Speech, Language, and Hearing Research*, 64(6S), 2213-2222.

- MacKenzie, L., & Turton, D. (2020). Assessing the accuracy of existing forced alignment software on varieties of British English. *Linguistics Vanguard*, 6(s1).

- Meer, P. (2020). Automatic alignment for New Englishes: Applying state-of-the-art aligners to Trinidadian English. The *Journal of the Acoustical Society of America*, 147(4), 2283-2294.

Exercises

Exercise 1

1. Open the file 'gne_spontaneous.wav' in Praat
   - In Praat press 'Open' > 'Read form file…' and select 'gne_spontaneous.wav'

2. Create a corresponding TextGrid with a single interval tier called 'sentence'
   - Select 'Sound gne_spontaneous' in Praat Objects and press 'Annotate' > 'To TextGrid…', and then type 'sentence' in the field 'All tier names:'

3. Provide orthographic transcription for the file in the 'sentence' tier. Do not use any punctuation.
   - Select the Sound and TextGrid together and press 'View & Edit'. Add boundaries wherever you can see a major pause (silence). Press on the spectrogram / waveform at time points corresponding to pauses and then click on the small circle that appears on top of the annotation tier. Once you have the boundaries, play the fragments of the sound file for each interval and type your transcription in the annotation tier).

4. Save your TextGrid as 'gne_spontaneous.TextGrid'

Exercise 2

1. Load the forced aligned TextGrids for the 'gne_spontaneous' file into Praat. You will find the TextGrids in the DARLA and MAUS subfolders in the file bundle for day 1 of the workshop.

2. Merge the two TextGrids in Praat
   - Select both TextGrids and press 'Merge'

3. Edit the merged TextGrid along with the Sound file in Praat.

4. Remove all tiers except 'MAU' and 'sentence - phones'

5. Compare the segmentation between the two systems. How do they compare to your own judgement?

Exercise 3

Record a sample of your own speech, using Praat, transcribe it and force align it, using DARLA or MAUS. If you speak one of the languages included in MAUS, other than English, try experimenting with a speech sample in that language. You will need ca. 1-2 minutes of speech for this exercise. Inspect the output. What is your evaluation of the segmentation accuracy?

Exercise 4

1. Load the sound file 'gne_read.wav' in to Praat, and the DARLA - aligned TextGrid (you will find it in the gne_read_DARLA).
2. Inspect the segmentation for the FOOT (UH1) and STRUT (AH1) vowels in the TextGrid. Use Praat search functions to identify the relevant vowel tokens. Make corrections if required.