

Forced Alignment

The [Bavarian Archive for Speech Signals \(BAS\)](#) has kindly published a set of [speech processing web services](#) including one for forced alignment called [WebMAUS](#). You can use this service yourself directly, using your web browser, but LaBB-CAT also has a module for using it automatically, called the *BAS Services Manager*.

Caution

Using WebMAUS for forced alignment requires LaBB-CAT to send your recordings and transcripts over the internet to a third party.

Although the [BAS Web Services Terms of Service](#) make clear that uploaded data is deleted after 24 hours, using the service is only suitable in situations in which you have consent from participants to do so.

(We *do* have such permission for the recordings used in this exercise.)

Install the layer manager

LaBB-CAT has a specialised ‘layer manager’ module for integrating with the BAS Web Services.

- (1) In LaBB-CAT, select the *layer managers* option on the menu, which gives you a list of the layer managers already installed.
- (2) At the bottom of the page, click the *List of layer managers that are not yet installed* link.
- (3) Look for *BAS Web Services Manager* in the list, and press its *Install* button.
- (4) Follow the “terms of usage” link and read the terms.
- (5) Close the terms page, returning to LaBB-CAT.
- (6) Tick the “Accept Terms of Usage” option.
- (7) Press *Install*.

You will see a page of information about the Layer Manager.

Set up a layer for triggering forced alignment

Forced alignment is a process that works on each ‘utterance’ the recording - i.e. each line in the transcript. As it processes all the words on a line at once, we need to create a ‘phrase’ layer.

- (8) Select the *phrase layers* option on the menu
- (9) At the top of the page, there’s a blank form for creating a new layer; fill in the following details:

- **Layer ID:** MAUS
- **Type:** *Text*
- **Alignment:** *Intervals*
- **Manager:** *BAS Web Services Manager*
- **Generate:** *Always*
- **Description:** WebMAUS forced alignment

(10) Press *New*.

You will see a form that allows you to configure the layer; hover the mouse pointer over each setting to see what its purpose is.

(11) The main choice is the *Phoneme encoding*; choose *DISC* for this setting.

(12) Press *Set Parameters*

We are going to force-align only a single participant's for now, so don't click *Regenerate* on this page.

Forced Alignment

- (13) Select the *participants* option on the menu at the top.
- (14) Tick the checkbox next to the first participant's name.
- (15) Press the *All Utterances* button above.
- (16) Press *List*.
LaBB-CAT will identify all the selected participant's utterances, and then list the first twenty.
- (17) Press the *MAUS* button at the bottom of the page.

LaBB-CAT will now upload each utterance – both the audio and the transcript – to the BAS Web Services server. LaBB-CAT must be able to access the internet for this to work, so if there is no internet connectivity, forced alignment will fail at this point.

Once WebMAUS has completed the forced alignment, LaBB-CAT downloads the resulting phone alignment times, and saves them in the *segment* layer.

Inspection/Correction

Once forced alignment is complete, you can inspect/correct alignments using LaBB-CAT's integration with Praat.

Tip

You can follow the following steps without having to wait for the forced alignment to complete.

Processing will continue even if you leave the current page.

If you later want to check back on the progress of the forced alignment task select the *activity* option on the menu. If the task is still running, a progress bar will be shown there.

- (18) Select the *transcripts* option on the menu.
- (19) Open the transcript you aligned.
- (20) On the *Layers* tab at the top, tick both the *MAUS* layer and the *segment* layer.
You will see which lines have been force-aligned, as they have a timestamp, and have the *segment* layer filled in.

The interactive transcript page doesn't show you the alignments of the words or phones, but you can see those using Praat. You can open individual utterances in Praat directly from the transcript page, but first, the LaBB-CAT/Praat integration has to be set up; this only has to be done once:

- i. On the top-right of the page, above the playback controls, there's a Praat icon - click it.
- ii. Follow the instructions that appear (these vary depending on what web browser you use).

Tip

You may need to grant a browser extension permission to install, and it's possible you will need a connection to the internet in order to download this extension.

You also may be asked where Praat is installed; Navigate to the location where Praat is installed, and double-click the "Praat.exe" file (on some systems the file may simply be called "Praat"). The Praat program may open, and then immediately close, as LaBB-CAT tests it can communicate with Praat.

Now Praat integration has been set up, and you should be able to access Praat options in the transcript page from now on...

- (21) Click on a line that has been aligned, and select the *Open TextGrid* option on the menu. Praat should open, and show you a spectrogram of the line's audio, with a TextGrid below that includes the words and the segments.
- (22) If you click on a word, and hit the tab key, the word's interval is played. Try out various words, and see what you think about how accurate the forced aligner has been.

Try this out with different lines in the transcript. You will see that in some cases the alignment is pretty good, and in other cases, it's not so good. In the not-so-good cases, see if you can figure out why the forced aligner got it wrong.

You may have noticed that, each time you open an utterance in Praat, a button appears in the transcript to the left of the line, labelled *Import Changes*. This button allows you to save any adjustments you might want to make to the alignments back into the LaBB-CAT database.

- (23) If you feel confident using Praat, open an utterance TextGrid, adjust the alignments of the words and phones so that they're more accurate, and then press the *Import Changes* button in the transcript.

 **Warning**

These changes are flagged as manual edits, so if forced-alignment is run again, they will not be over-written with new bad alignments. Therefore it's important that the changes you make are actually improvements, because automated forced alignment will never change them again.

 **Important**

There are some rules about what you can change when correcting word/segment alignments:

1. You're not allowed to add or delete words (if this is necessary, it should be done by correcting the transcript instead).
2. All the segments must be within the bounds of their own word.
3. The start of the first phone should line up with the start of the word, and the end of the last phone should line up with the end of the word.
4. You should not change the alignment of the utterance itself (which would only be possible if you select the *Open Text Grid incl. ± 1 utterance in Praat* option).

Segment Layer Searches

Now that WebMAUS has created individually aligned phones in the corpus, those speech sounds can be searched and their labels and alignment information exported.

Let's say you're particularly interested in the NEAR /ɪə/ and SQUARE /ɛə/ monophthongs. You can now identify and extract instances of those phonemes, using their CELEX DISC labels 7 and 8.

- (24) Select the *search* option on the menu.
- (25) Tick the *segment* layer.

The segments layer contains annotations at the sub-word level - i.e. there are potentially multiple annotations per word, each annotation representing a single phone of the word.

On the search page you will see that, as with other layers, there is a box on the segments layer for a regular expression.

As with other patterns in the search matrix, the pattern that you enter in the box is matched against individual annotations. So if you enter 7 (i.e. CELEX DISC for /ɪə/) in the box, it will match each ‘NEAR’ vowel segment in each word in the database.

! Important

If you enter a pattern that would match more than a single character on this layer (i.e. more than a single phoneme) then no search results will be returned, because each annotation on this layer is only a single character long (remember the DISC encoding uses one character per phoneme).

For example, if you enter n7 for your search, intending to match all instances of the word “near”, then no results will be returned, because no single segment will ever match that pattern.

You *can* match multiple segments, by adding another *segment* box using the **+** button in it, and entering for example n in the first box and 7 in the second box.

- (26) We want to search for all instances of the ‘NEAR’ and ‘SQUARE’ vowels, so enter [78] in the *segment* pattern box.
- (27) Press *Search*
After a short delay, you should see a list of words that have either /ɪə/ or /ɛə/.

You can export all these vowel tokens to a CSV file for analysis or further processing

- (28) Press the *CSV Export* button.
- (29) Save and open the resulting file.

You will see that the file includes the following columns:

- **Target segment** – the label of the phone that matched the search pattern,
- **Target segment start** – the start time of the phone and
- **Target segment end** – the end time of the phone, as determined by WebMAUS.

This data can be used directly to calculate vowel duration, and in combination with Praat to make acoustic measurements.

Acoustic Measurement

The CSV file includes the columns “Target segments start” and “Target segments end”; these columns have the start and end time of the matching segment tokens. Given this information, LaBB-CAT can extract acoustic measurements on the speech sounds using Praat.

- (30) In LaBB-CAT, select the *extract* menu option.
- (31) Click the *process with praat* option.

i Note

If you have no *process with praat* option, and you have LaBB-CAT installed locally on your own computer (as opposed to it running on a server you access over the internet):

1. Find out where Praat is installed on your own computer.
On a Windows computer this might be somewhere like C:\Program Files\Praat
On a Mac this will probably be /Applications
2. Select the *system attributes* option on the LaBB-CAT menu.
3. In the **Praat Path** box, enter the location you found above.
4. Press the *Save* button at the bottom right of the page.

Then go back the the *upload* page, and you should find that the *process with praat* option has appeared.

- (32) Press *Browse* or *Choose File* and select the CSV results that you saved above.
You will see a form to fill in, and the first couple of settings (*Transcript Name column* and *Participant column* should be already filled in).
- (33) For the *Start Time column*, ensure that the *Target segment start* option is selected.
- (34) For the *End Time column*, ensure the *Target segment end* option is selected.

These two settings define the start/end times of the phone. For some measurements you might extract from Praat, processing signal that includes surrounding context is usually a good idea. You'll see there's a setting for that (which you can leave at the default of 0.025s), and you will see options for various measurements.

The default options are for *F1* and *F2* only, but if you feel like getting other measurements, feel free to tick those options too. You can expand each section with the button to reveal more settings, which allow you to specify more detail about how Praat should do its computations. Again, feel free to look at those and try different settings.

6. Press *Process*.
You will see a progress bar while LaBB-CAT generates Praat scripts and runs them.
7. Once Praat has finished processing the intervals, you will get a CSV file (you might have to click the *CSV file with measurements* link) - save and open it.

You will see that it's a copy of the CSV file you uploaded, with some extra columns added on the right.

Depending on your settings, this will include at least one column per measurement you selected (the formant columns also include a column containing the time at which the measurements were taken), and a final column called *Error* which is hopefully blank, but

which might contain errors reported back by Praat (e.g. if it couldn't find the audio file or ran into any other problem during processing).

In this exercise, you have seen how WebMAUS can be used to compute word and phone alignments automatically from your data. Perfect automatic labels and alignments are not guaranteed, but LaBB-CAT has a mechanism for manually correcting poor alignments.

Once forced alignment has been done, individual speech sounds can be identified by searching, extracted, and measured using Praat.