An introduction to forced alignment using the Munich Automatic Segmentation System (MAUS)

Catalina Torres

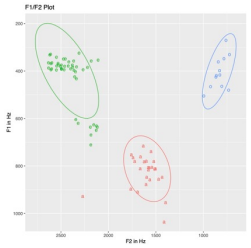
she/her

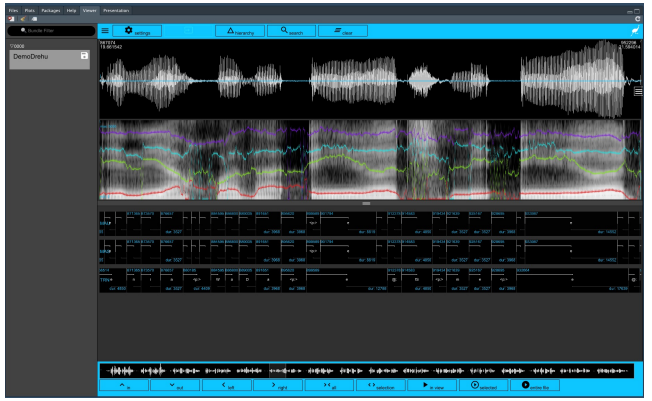
catalina.torres@ivs.uzh.ch

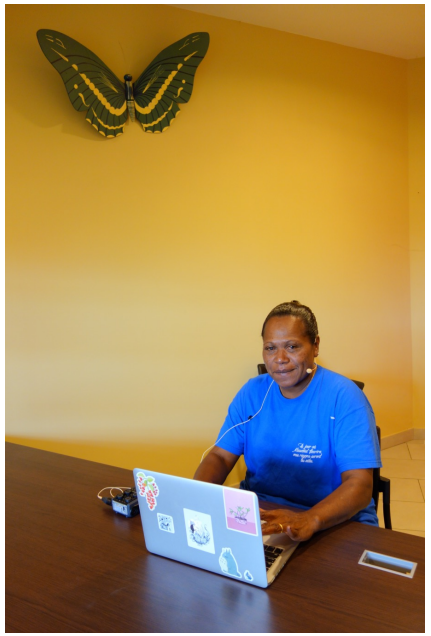
About the instructor

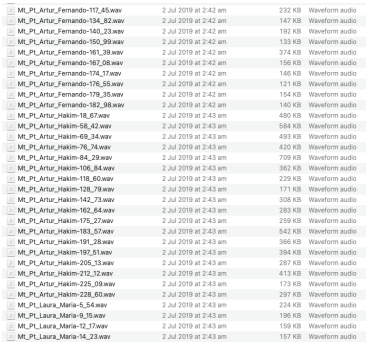
- Field Phonetician
- Research acoustic phonetics (prosody, segments, vowels)
- Field work in New Caledonia (Drehu, French)
- PhD from University of Melbourne, Australia
- I speak 5 languages (English, Spanish, German, French and Portuguese)

What is your experience with data processing?









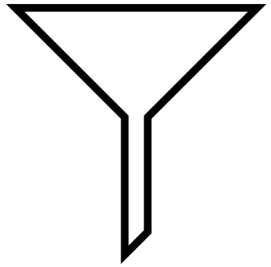
From data collection to a data base

What can acoustic analysis do for us?

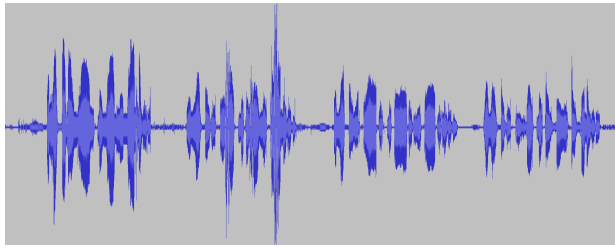
- You just started investigating a previously undocumented language and would like to define a vowel inventory
- You are investigating a language and suspect there is a vowel merger (Central Pame, Otomanguean)
- You are interested in socio-linguistic aspects and suspect there are differences

The speech data processing bottleneck

Eni a qaja ööhöö me ööhöö nge ööhöö hmaca
Eni a qaja eötr me eötr nge eötr hmaca
Eni a qaja axöö me axöö nge axöö hmaca
Eni a qaja hoho me hoho nge hoho hmaca
Eni a qaja eöötr me eöötr nge eöötr hmaca
Eni a qaja öhö me öhö nge öhö hmaca
Eni a qaja öö me öö nge öö hmaca
Eni a qaja keejë me keejë nge keejë hmaca
Eni a qaja axö me axö nge axö hmaca
Eni a qaja kejë me kejë nge kejë hmaca
Eni a qaja Kooko me Kooko nge Kooko hmaca
Eni a qaja möö me möö nge möö hmaca
Eni a qaja ö me ö nge ö hmaca
Eni a qaja Kokoo me Kokoo nge Kokoo hmaca
Eni a qaja gööime me gööime nge gööime hmaca



[a]



1 minute recording = 800 mins (transcribing, aligning) = 13+h

Steps in phonetic transcription

Create reliable segmentations between the acoustic signal and phonemic segments:

1. Orthographically transcribe speech
2. Time-align phonemes with respect to the waveform and spectrogram

- Manual alignment is incredibly time-consuming
- Reports of it being approximately 800 times longer than the duration of the speech itself

(Schiel et al. 2012)

Forced-alignment

What is forced alignment?

- Computational method:
  - Allows for automated phonemic transcriptions
  - Time-alignment at the segment level
  - Derived from orthographic transcriptions, time-aligned at the utterance level.

Different tools

- Munich Automatic Segmentation System (MAUS)
- Montreal forced aligner (MFA)
- Forced Alignment & Vowel Extraction (FAVE)
- Language, Brain and Behaviour Corpus Analysis Tool (LaBB-CAT)

(Gonzalez et al, 2020; Barth et al, 2020; Gnevsheva et al, 2020; Biczysko, 2022)

Advantages of MAUS

- Web based service
- No need to install any software
- No need to register
- Documentation and tutorials available online

Getting started with MAUS

What do we need?

- Recording (mono + .wav)
- Praat or ELAN transcription (utterance level transcription)
- Internet access
- Browser Chrome
- Imap mapping file

What is an Imap mapping file?

File containing orthography-to-phoneme information

Needs to be a .txt file with UTF-8 encoding

Phonemic information encoded with SAMPA

What is SAMPA?

Speech Assessment Methods Phonetic Alphabet

Machine readable alphabet

How to create an Imap mapping file

- Instructions
- SAMPA symbols
- Inventory in MAUS

Orthography

SAMPA

Drehu.txt — Edited

```
a;a
aa;a:
b;b
c;S
d;d
dr;dZ
e;e
ee;e:
ë;E
ëë;E:
f;f
g;g
h;h
hl;hl
hm;hm
hn;hn
hng;hN
hny;hJ
i;i
```

SAMPA inventory in MAUS

AA	AA	ɑː	lengthened back open unrounded vowel in quantity II		vaaru	ekk	vowel	lowback	AA.ekk 0.04
AA:	AA:	ɑɑː	lengthened back open unrounded vowel in quantity III			ekk	vowel	lowback	AA:.ekk 0.04
ae	ae	æ	diphthong	eng-SC	diphthong	lowfront>midfront	Ae.eng-AU	0.04	
Ae	Ae	œ	diphthong	aus	diphthong	lowback>midfront	Ae.eng-AU	0.04	
Ae:	Ae:	œː	diphthong	ekk	diphthong	lowback>midfront	Ae.eng-AU	0.04	
a:i	a:i	aːi	diphthong	nld	diphthong	lowfront>fronthigh	a:i.nld 0.04		
a~i~	a~i~	ãĩ	diphthong nasalized	(Taiwanese Min Nan)	nan	diphthong	lowfront>highfront	a~i~.nan-TW	0.04
ai	ai	ai	diphthong	ita	diphthong	lowfront>fronthigh	ai.nan-TW	0.04	
ai:	ai:	aiː	diphthong	isl	diphthong	lowfront>fronthigh	aI.deu 0.04		
aI	aI	aɪ	diphthong Bein	deu	diphthong	lowfront>fronthigh	aI.deu 0.04		
Ai	Ai	ɑi	diphthong	ekk	diphthong	lowback>highfront	Ai.nor 0.04		
Ai:	Ai:	ɑiː	diphthong	ekk	diphthong	lowback>highfront	ai.ita 0.04		
AI	AI	ɑɪ	diphthong	ltz	diphthong	lowback>highfront	aI.deu 0.04		
Ao:	Ao:	ɑoː	diphthong	ekk	diphthong	lowback>midback aU.deu	0.04		
au	au	au	diphthong	isl	diphthong	lowfront>backhigh	au.nan-TW	0.04	
au:	au:	auː	diphthong	isl	diphthong	lowfront>backhigh	aU.deu 0.04		
aU	aU	aʊ	diphthong Haus	deu	diphthong	lowfront>backhigh	aU.deu 0.04		

Questions

Group discussion

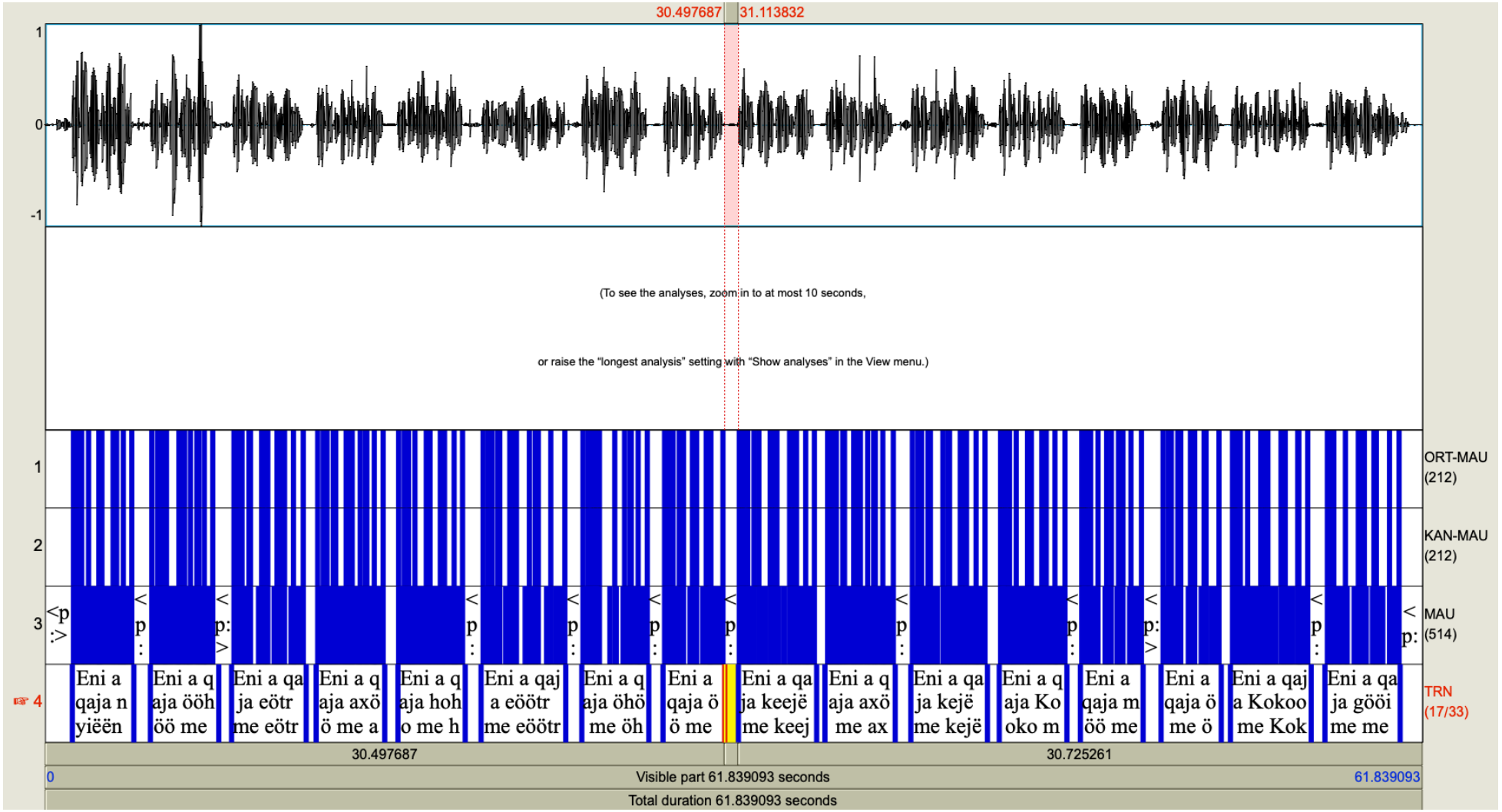
How do you decide which symbols to use for an Imap mapping file?

What are the challenges when dealing with a previously undocumented language?

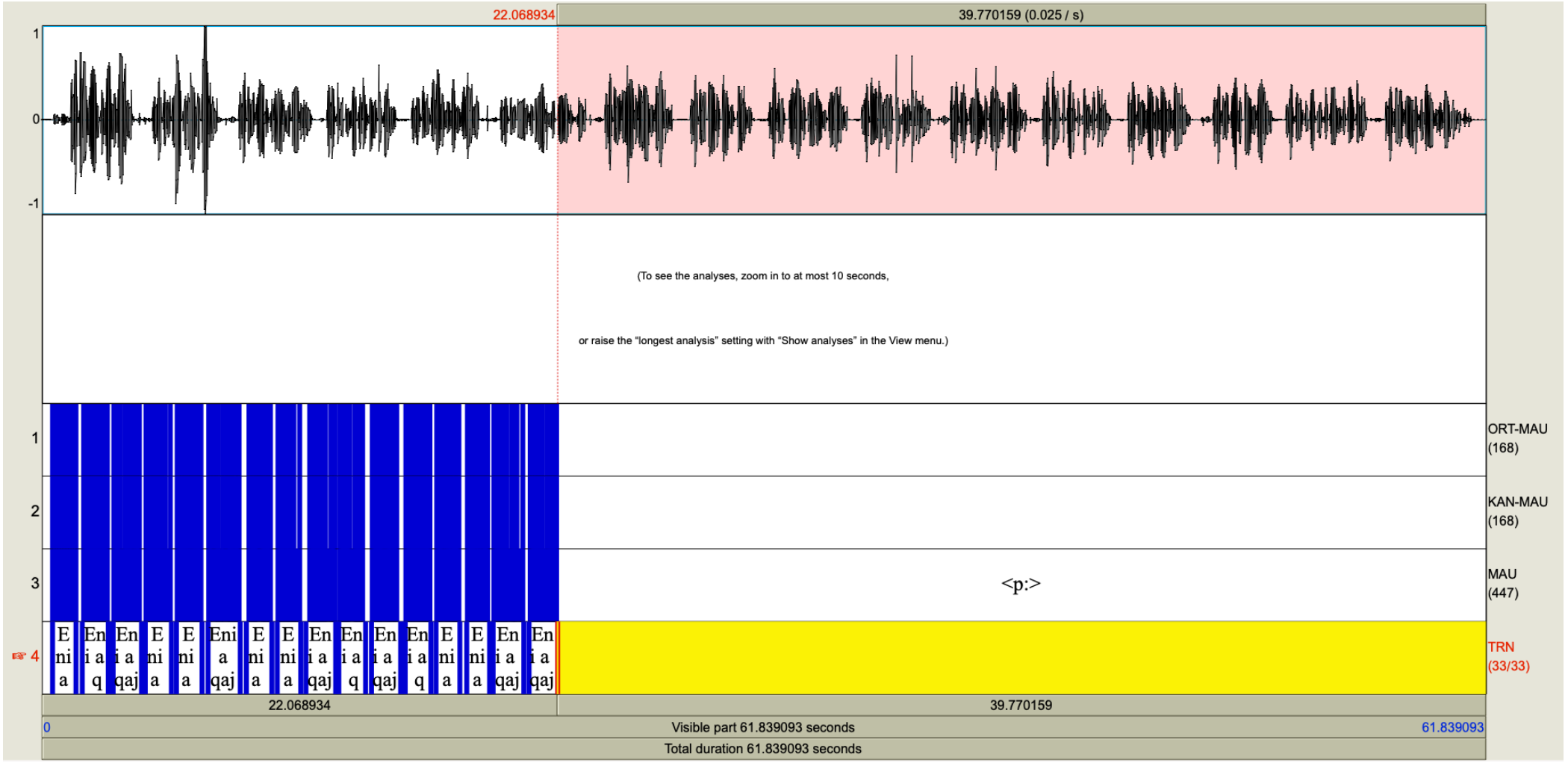
What should we consider before force aligning recordings from an under-documented language?

File inspection

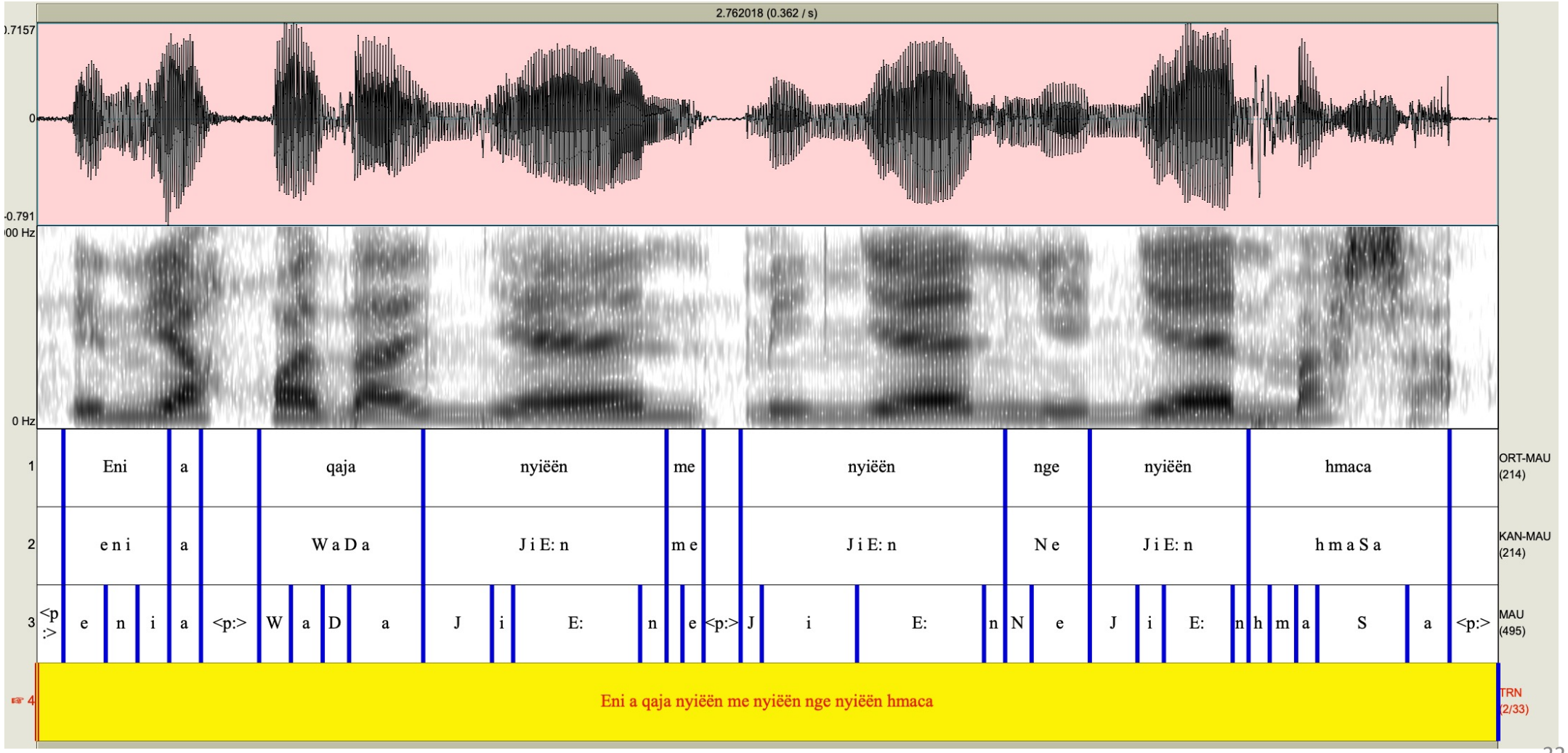
Successful forced-alignment



Infelicitous forced-alignment



Alignment details

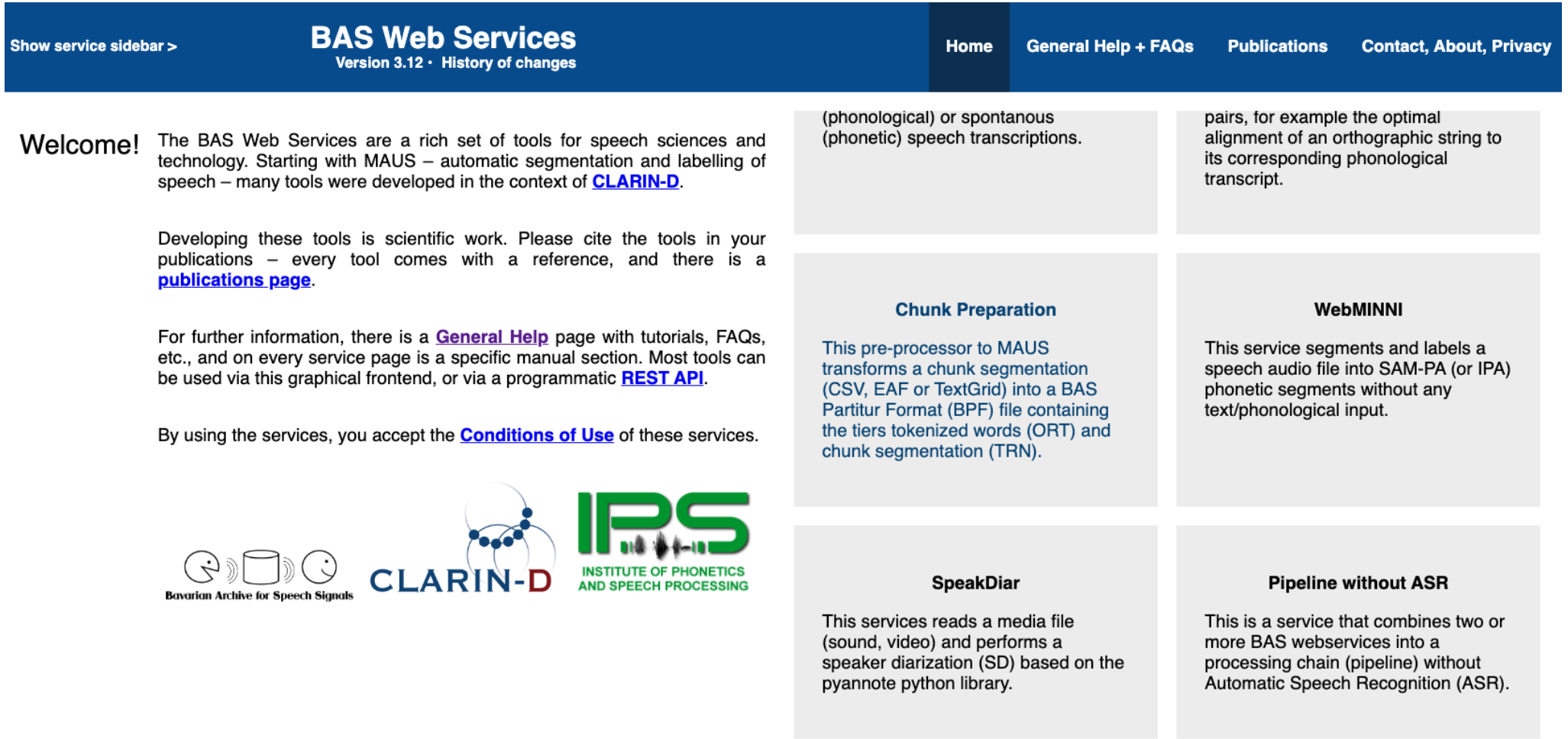


How to create a well aligned TextGrid

1. Chunk Preparation (Prepare Chunks)
2. G2P (Carry out text to transcript conversion)
3. WebMAUS general (Force align)

Starting point

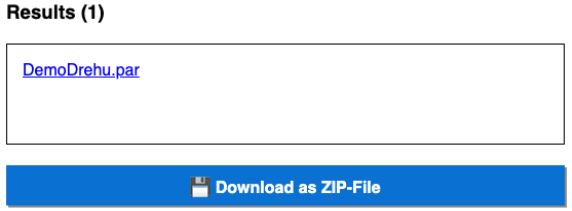
http://clarin.phonetik.uni-muenchen.de/BASWebServices



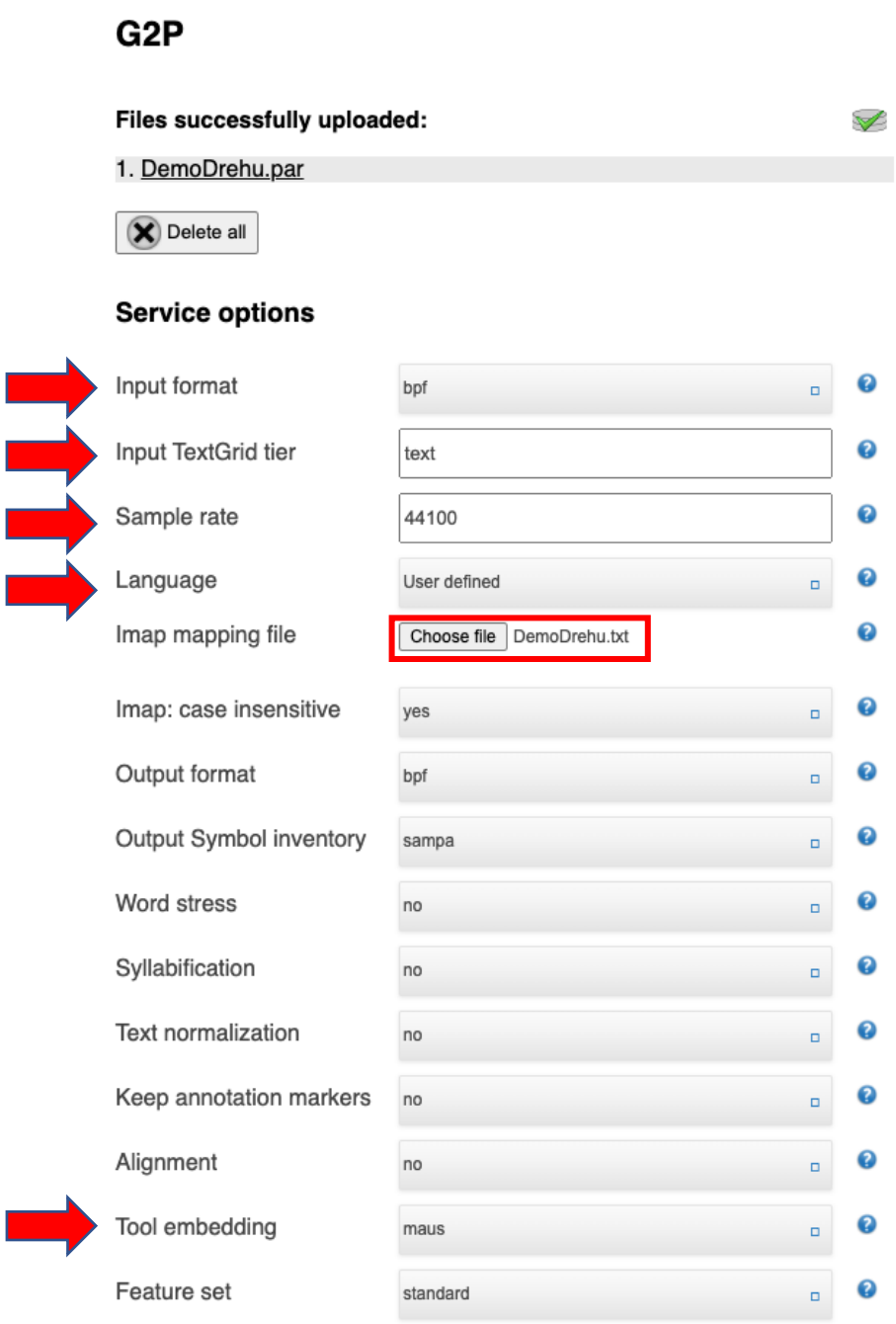
1. Chunk preparation

Steps:

1. Drag & drop .TextGrid
2. Upload
3. Adjust options:
   - Language = Language independent
   - Input tier name = text
   - Sampling rate = 44100
   - Keep annotation = yes
   - all other options default
4. Run service
5. Download zip file
6. Unzip file (needed for next step)

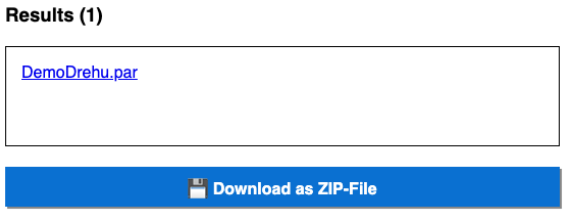


2. G2P

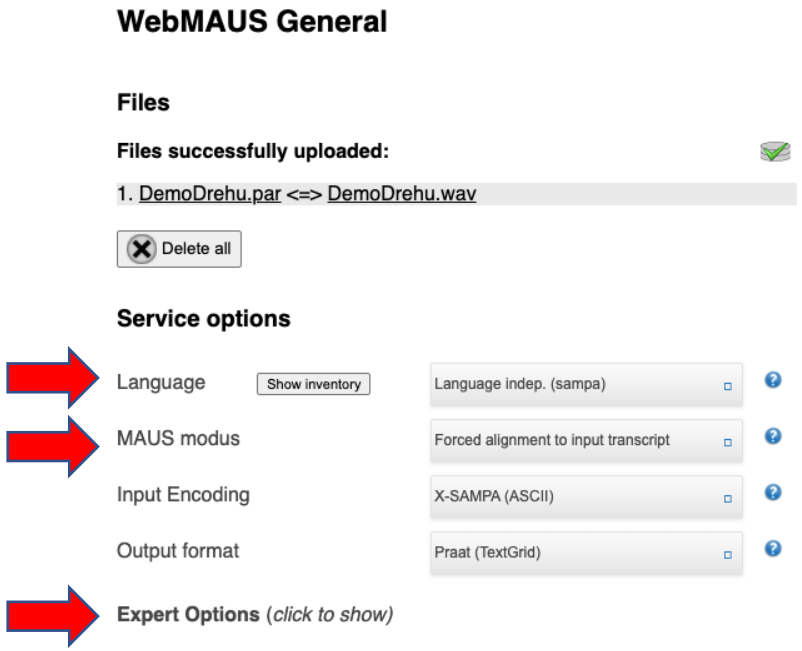


Steps:

1. Drag & drop .par
2. Upload
3. Adjust options:
   - Input format = bpf
   - Input TextGrid tier = text
   - Sample rate = 44100
   - Language = User defined
   - Imap mapping file = Drehu.txt
   - Tool embedding = maus
   - all other options default
4. Run service
5. Download zip file
6. Unzip file (needed for the next step)

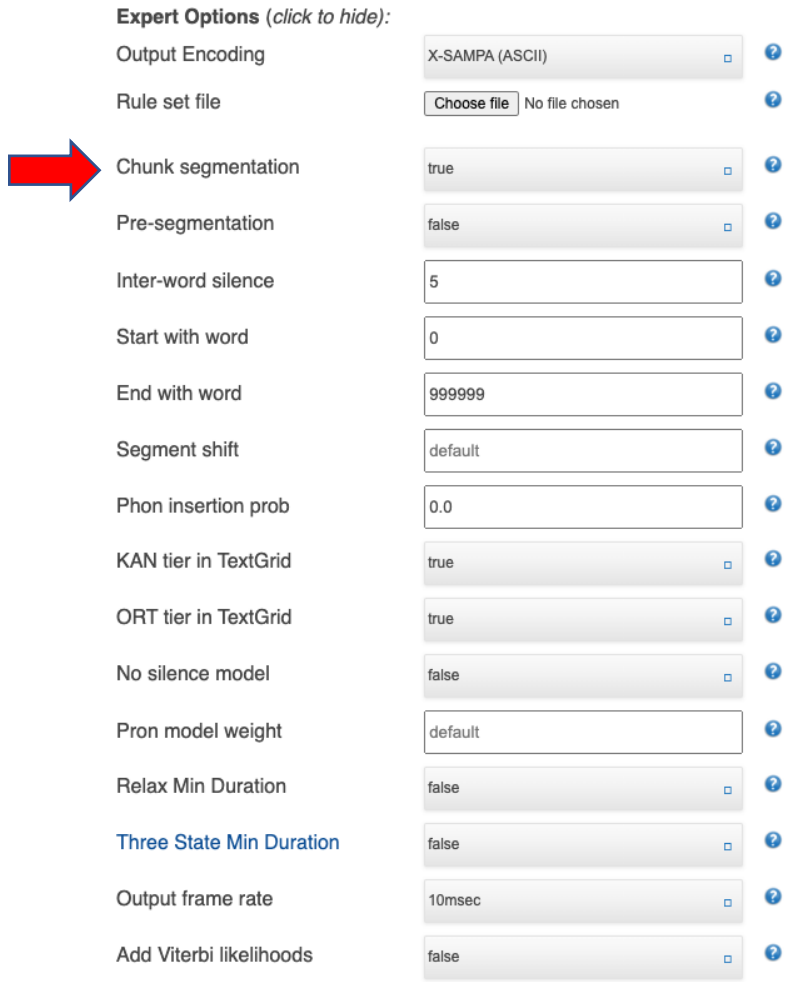


3. WebMAUS General



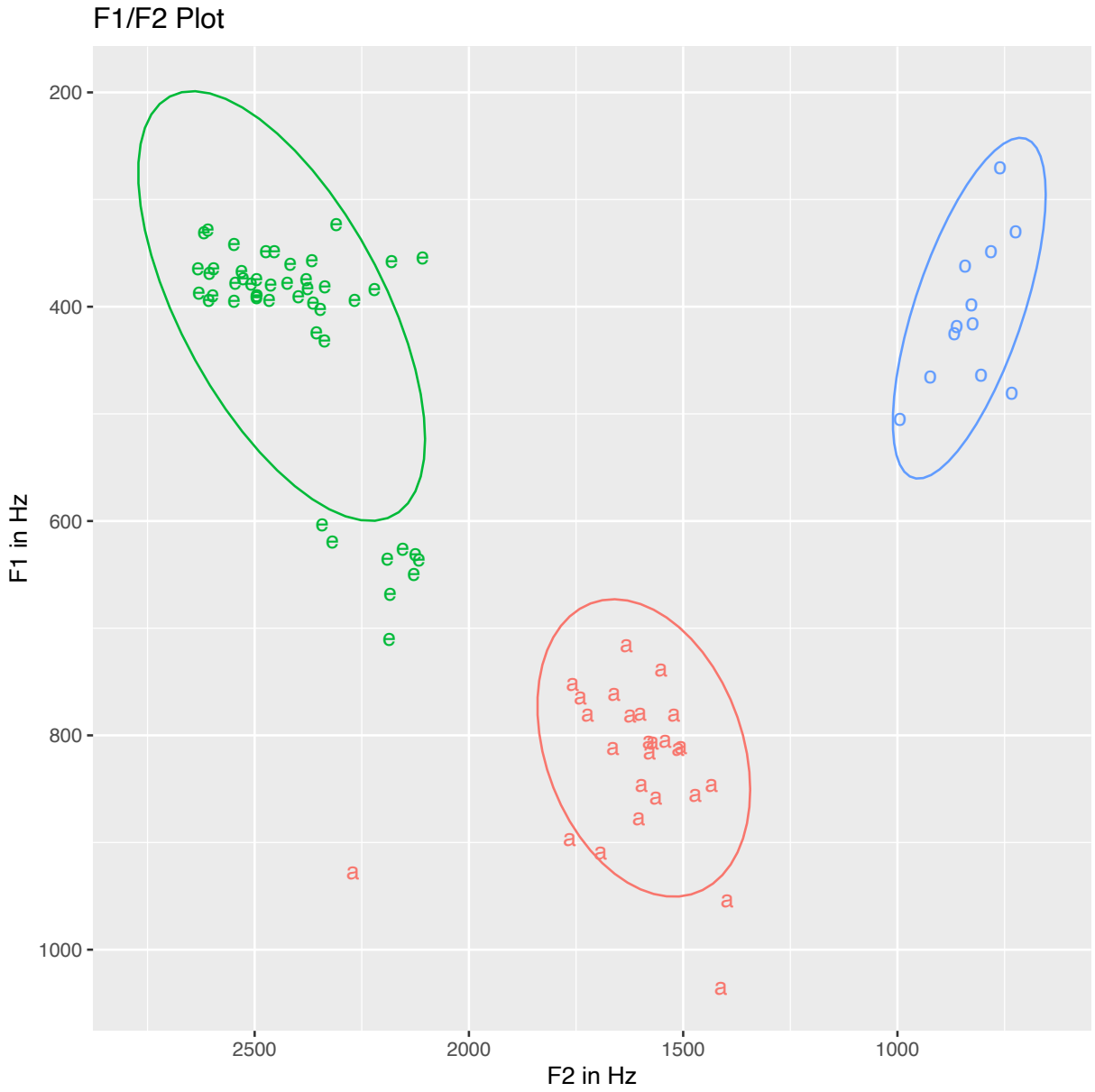
Steps:

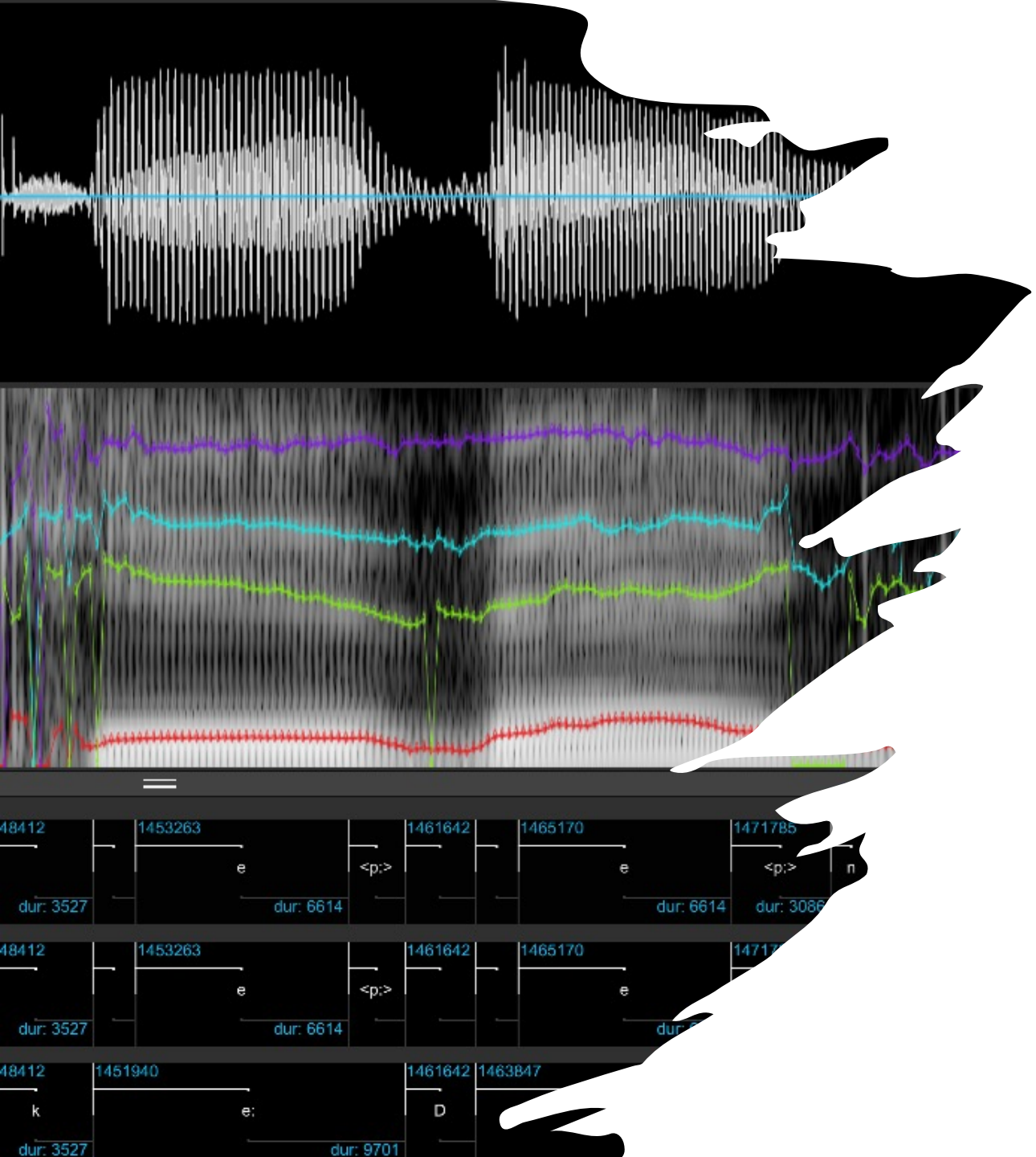
1. Drag & drop (second) .par + .wav
2. Upload
3. Adjust options:
   - Language = Language indep.
   - MAUS modus = Forced alignment..
   - ➢ Click on **Expert options:**
   - Chunk segmentation = true
   - All other options default
4. Run service (it will take some time)
5. Download zip file
6. Unzip file
7. Open new .TextGrid in Praat



How are your files looking?

Additional material



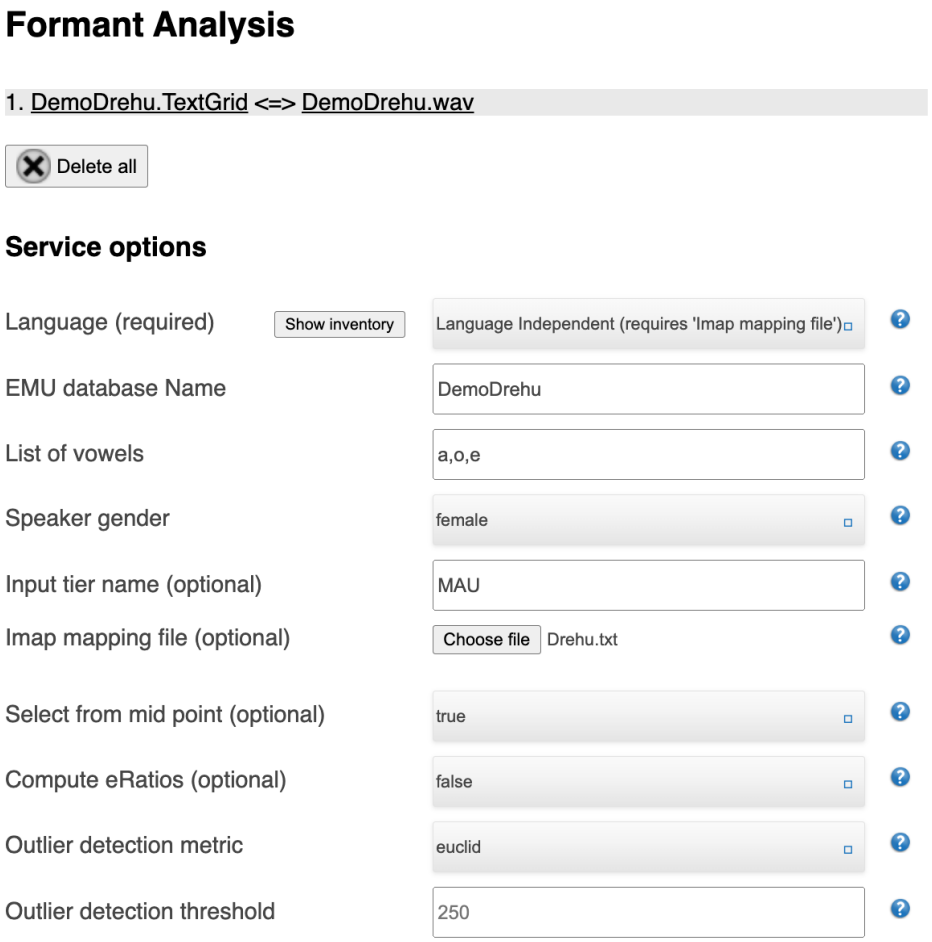


DemoDrehu_emuDB

- Go to > Additional_Material
- Go to > DemoDrehu_emuDB
- Go to > FormantAnalysis folder
- Open > plots : you will find an array of plots!
- This is helpful for data exploration!
- To inspect the emuDB and try out the R script you need to have R + R studio installed
- You will also need additional packages:

"wrassp", "emuR", "ggplot2", "dplyr", "tidyr", "argparse", "gridExtra", "docopt"

Formant Analysis



1. Steps:
2. Drag & drop new .TextGrid + .wav (Supplementary_Step)
3. Upload
4. Adjust options:
   - Language = Language independent
   - EMU data base name = DemoDrehu
   - List of vowels = a,o,e
   - Speaker gender = female
   - Input tier name = MAU
   - Imap mapping file = Drehu.txt
   - Select from midpoint = true
   - all other options default
5. Run service
6. Download zip file
7. Unzip file (needed for next step)

- Approx 9-10 minutes for the DemoDrehu

Thank you!

References

- Barth, D., Grama, J., Gonzalez, S., & Travis, C. (2020). Using forced alignment for sociophonetic research on a minority language. *University of Pennsylvania Working Papers in Linguistics*, *25*(2), 2.
- Biczysko, K. (2022). Automatic Annotation of Speech: Exploring Boundaries within Forced Alignment for Swedish and Norwegian. MA thesis Uppsala University.
- Gonzalez, S., Grama, J., & Travis, C. E. (2020). Comparing the performance of forced aligners used in sociophonetic research. *Linguistics Vanguard*, *6*(1).
- Gnevsheva, K., Gonzalez, S., & Fromont, R. (2020). Australian English bilingual corpus: automatic forced-alignment accuracy in Russian and English. *Australian Journal of Linguistics*, *40*(2), 182-193.
- Reichel, U.D., Kisler, T. (2014). Language-independent grapheme-phoneme conversion and word stress assignment as a web service. In: Hoffmann, R. (Ed.): Elektronische Sprachverarbeitung. Studientexte zur Sprachkommunikation 71, pp 42-49, TUDpress, Dresden.