

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

Catalina Torres

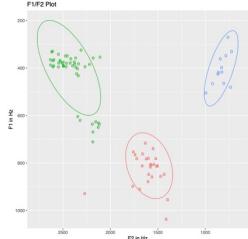
catalina.torres@ivs.uzh.ch

About the instructor

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

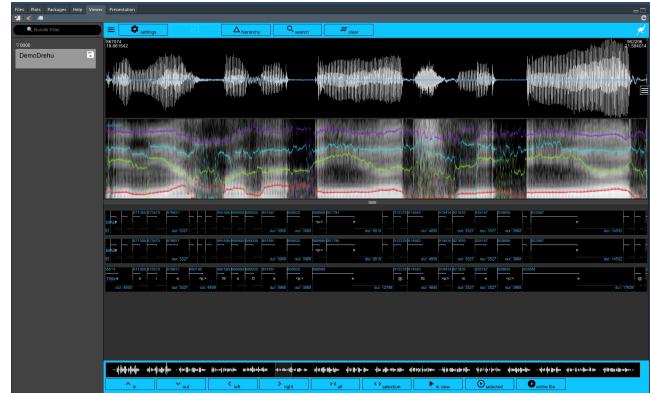


What is your experience
with data processing?



From data collection
to a data base

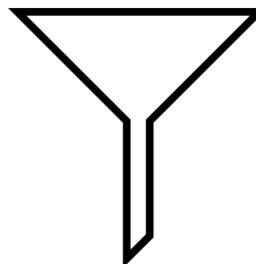
ML_Pt_Artur_Fernando-117_45.wav	2 Jul 2019 at 2:42 am	232 KB	Waveform audio
ML_Pt_Artur_Fernando-134_82.wav	2 Jul 2019 at 2:42 am	147 KB	Waveform audio
ML_Pt_Artur_Fernando-140_23.wav	2 Jul 2019 at 2:42 am	192 KB	Waveform audio
ML_Pt_Artur_Fernando-153_54.wav	2 Jul 2019 at 2:42 am	153 KB	Waveform audio
ML_Pt_Artur_Fernando-161_39.wav	2 Jul 2019 at 2:42 am	374 KB	Waveform audio
ML_Pt_Artur_Fernando-167_08.wav	2 Jul 2019 at 2:42 am	156 KB	Waveform audio
ML_Pt_Artur_Fernando-172_74.wav	2 Jul 2019 at 2:42 am	146 KB	Waveform audio
ML_Pt_Artur_Fernando-173_55.wav	2 Jul 2019 at 2:42 am	151 KB	Waveform audio
ML_Pt_Artur_Fernando-179_35.wav	2 Jul 2019 at 2:42 am	154 KB	Waveform audio
ML_Pt_Artur_Fernando-182_98.wav	2 Jul 2019 at 2:42 am	140 KB	Waveform audio
ML_Pt_Artur_Jairim-18_100.wav	2 Jul 2019 at 2:43 am	480 KB	Waveform audio
ML_Pt_Artur_Jairim-43_53.wav	2 Jul 2019 at 2:43 am	584 KB	Waveform audio
ML_Pt_Artur_Jairim-69_34.wav	2 Jul 2019 at 2:43 am	493 KB	Waveform audio
ML_Pt_Artur_Jairim-76_74.wav	2 Jul 2019 at 2:43 am	420 KB	Waveform audio
ML_Pt_Artur_Jairim-84_29.wav	2 Jul 2019 at 2:43 am	709 KB	Waveform audio
ML_Pt_Artur_Jairim-95_59.wav	2 Jul 2019 at 2:43 am	363 KB	Waveform audio
ML_Pt_Artur_Jairim-118_60.wav	2 Jul 2019 at 2:43 am	229 KB	Waveform audio
ML_Pt_Artur_Jairim-126_79.wav	2 Jul 2019 at 2:43 am	171 KB	Waveform audio
ML_Pt_Artur_Jairim-142_73.wav	2 Jul 2019 at 2:43 am	303 KB	Waveform audio
ML_Pt_Artur_Jairim-144_64.wav	2 Jul 2019 at 2:43 am	285 KB	Waveform audio
ML_Pt_Artur_Jairim-175_27.wav	2 Jul 2019 at 2:43 am	259 KB	Waveform audio
ML_Pt_Artur_Jairim-183_57.wav	2 Jul 2019 at 2:43 am	542 KB	Waveform audio
ML_Pt_Artur_Jairim-190_25.wav	2 Jul 2019 at 2:43 am	368 KB	Waveform audio
ML_Pt_Artur_Jairim-193_20.wav	2 Jul 2019 at 2:43 am	394 KB	Waveform audio
ML_Pt_Artur_Jairim-205_13.wav	2 Jul 2019 at 2:43 am	287 KB	Waveform audio
ML_Pt_Artur_Jairim-212_32.wav	2 Jul 2019 at 2:43 am	413 KB	Waveform audio
ML_Pt_Artur_Jairim-218_49.wav	2 Jul 2019 at 2:43 am	173 KB	Waveform audio
ML_Pt_Artur_Jairim-226_60.wav	2 Jul 2019 at 2:43 am	287 KB	Waveform audio
ML_Pt_Artur_Maria-5_54.wav	2 Jul 2019 at 2:43 am	224 KB	Waveform audio
ML_Pt_Laura_Maria-9_53.wav	2 Jul 2019 at 2:43 am	196 KB	Waveform audio
ML_Pt_Laura_Maria-12_77.wav	2 Jul 2019 at 2:43 am	159 KB	Waveform audio
ML_Pt_Laura_Maria-14_23.wav	2 Jul 2019 at 2:43 am	187 KB	Waveform audio



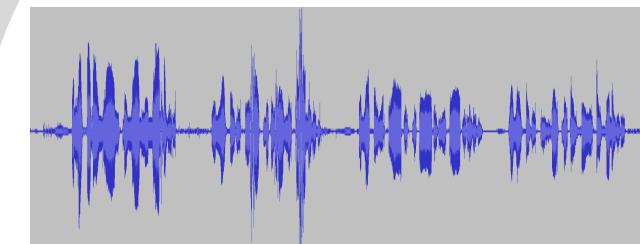
What can acoustic analysis do for us?

- 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

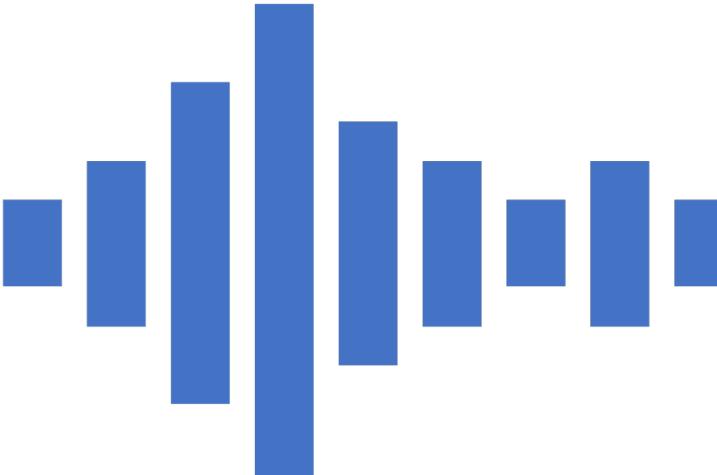
The speech data processing bottleneck



[a]



Eni a qaja ööhöö me ööhöö nge ööhöö hmaca
Eni a qaja eöötr me eöötr nge eöötr 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 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



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)

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

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 wl, 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 lmap 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

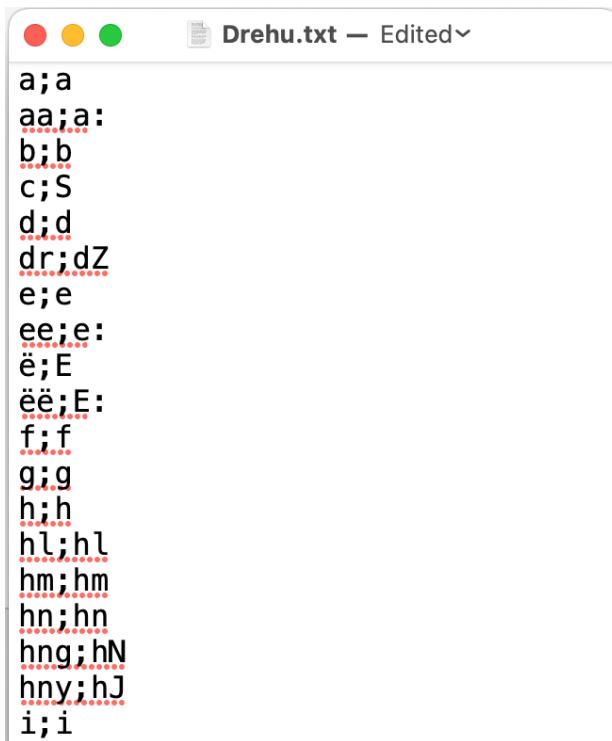


Machine readable alphabet

How to create an Imap mapping file

- [Instructions](#)
- [SAMPA symbols](#)
- [Inventory in MAUS](#)

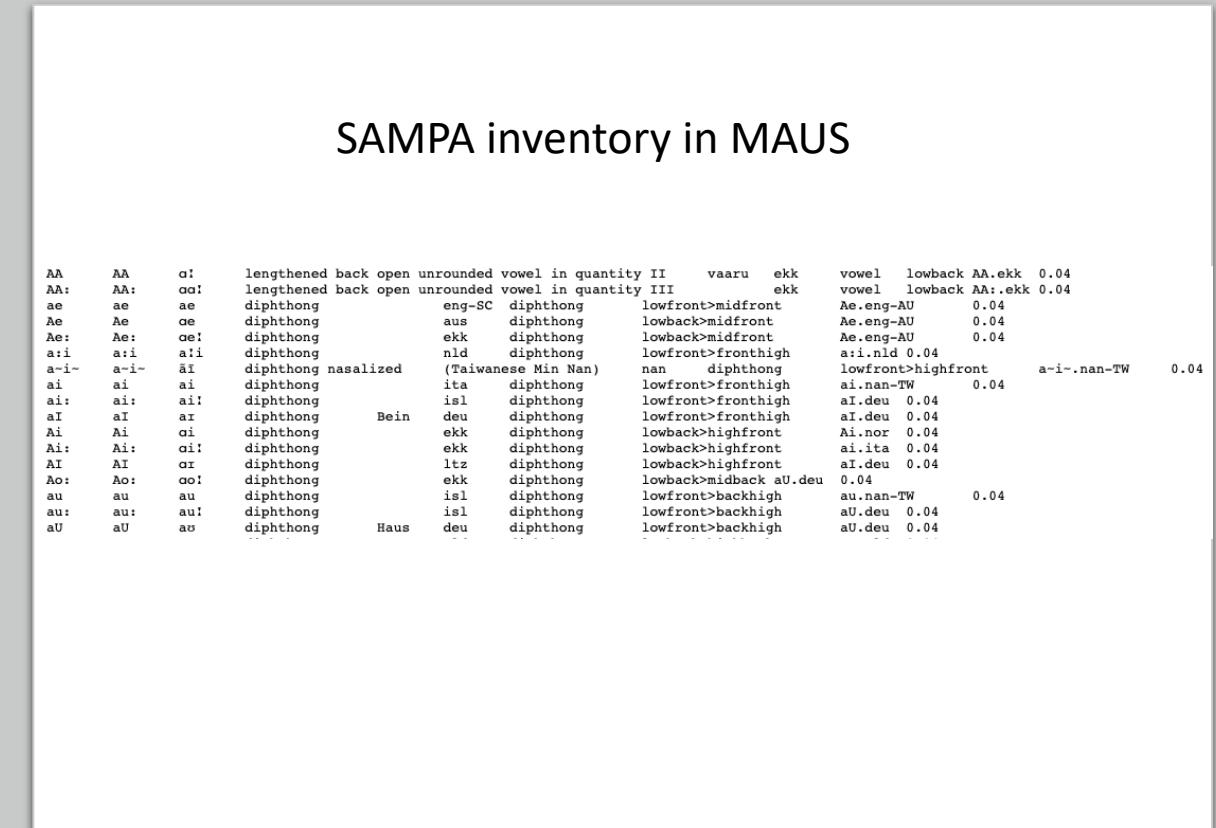
Orthography



Drehu.txt — Edited

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

SAMPA



SAMPA inventory in MAUS

AA	AA:	o:	lengthened back open unrounded vowel in quantity II	vaaru	ekk	vowel	lowback AA.ekk	0.04	
ae	ae	ae	diphthong	eng-SC	diphthong	lowfront>midfront	Ae.eng-AU	0.04	
Ae	Ae	oe	diphthong	aus	diphthong	lowback>midfront	Ae.eng-AU	0.04	
Ae:	Ae:	oe:	diphthong	ekk	diphthong	lowback>midfront	Ae.eng-AU	0.04	
a:i	a:i	a:i	diphthong	nld	diphthong	lowfront>fronthigh	a:i.nld	0.04	
a:-i-	a:-i-	áí	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	
a:i:	a:i:	ai:	diphthong	isl	diphthong	lowfront>fronthigh	aI.deu	0.04	
aI	aI	ar	diphthong	Bein	deu	diphthong	lowfront>fronthigh	aI.deu	0.04
Ai	Ai	oi	diphthong	ekk	diphthong	lowback>highfront	Ai.nor	0.04	
Ai:	Ai:	oi:	diphthong	ekk	diphthong	lowback>highfront	ai.ita	0.04	
AI	AI	or	diphthong	ltz	diphthong	lowback>highfront	aI.deu	0.04	
Ao:	Ao:	oo:	diphthong	ekk	diphthong	lowback>midback aU.deu	0.04		
au	au	au	diphthong	isl	diphthong	lowfront>backhigh	au.nan-TW	0.04	
au:	au:	au:	diphthong	isl	diphthong	lowfront>backhigh	aU.deu	0.04	
aU	aU	ao	diphthong	Haus	deu	diphthong	lowfront>backhigh	aU.deu	0.04

Exercise

Group discussion

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

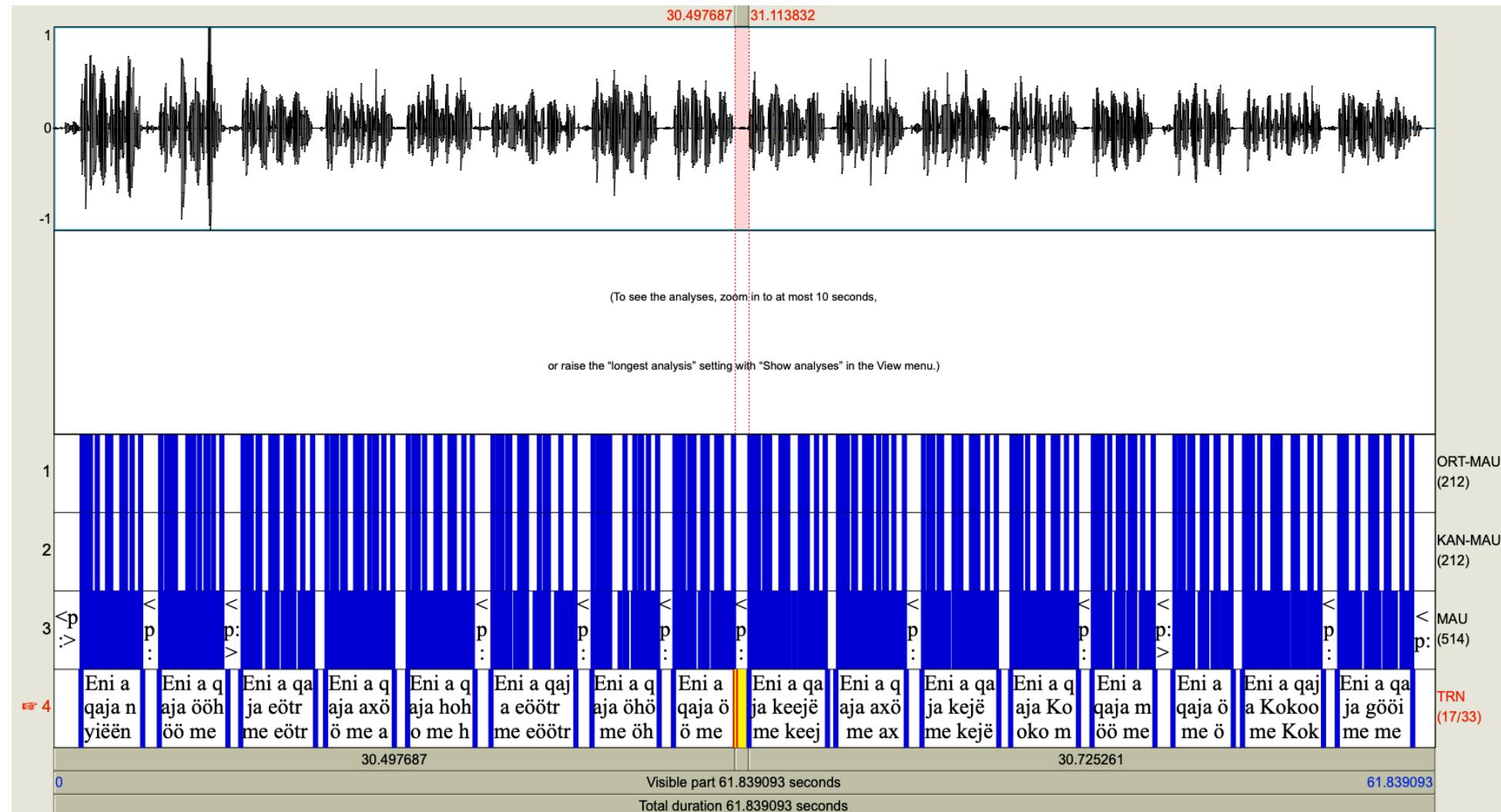
What are the challenges when dealing with a previously undocumented language?

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

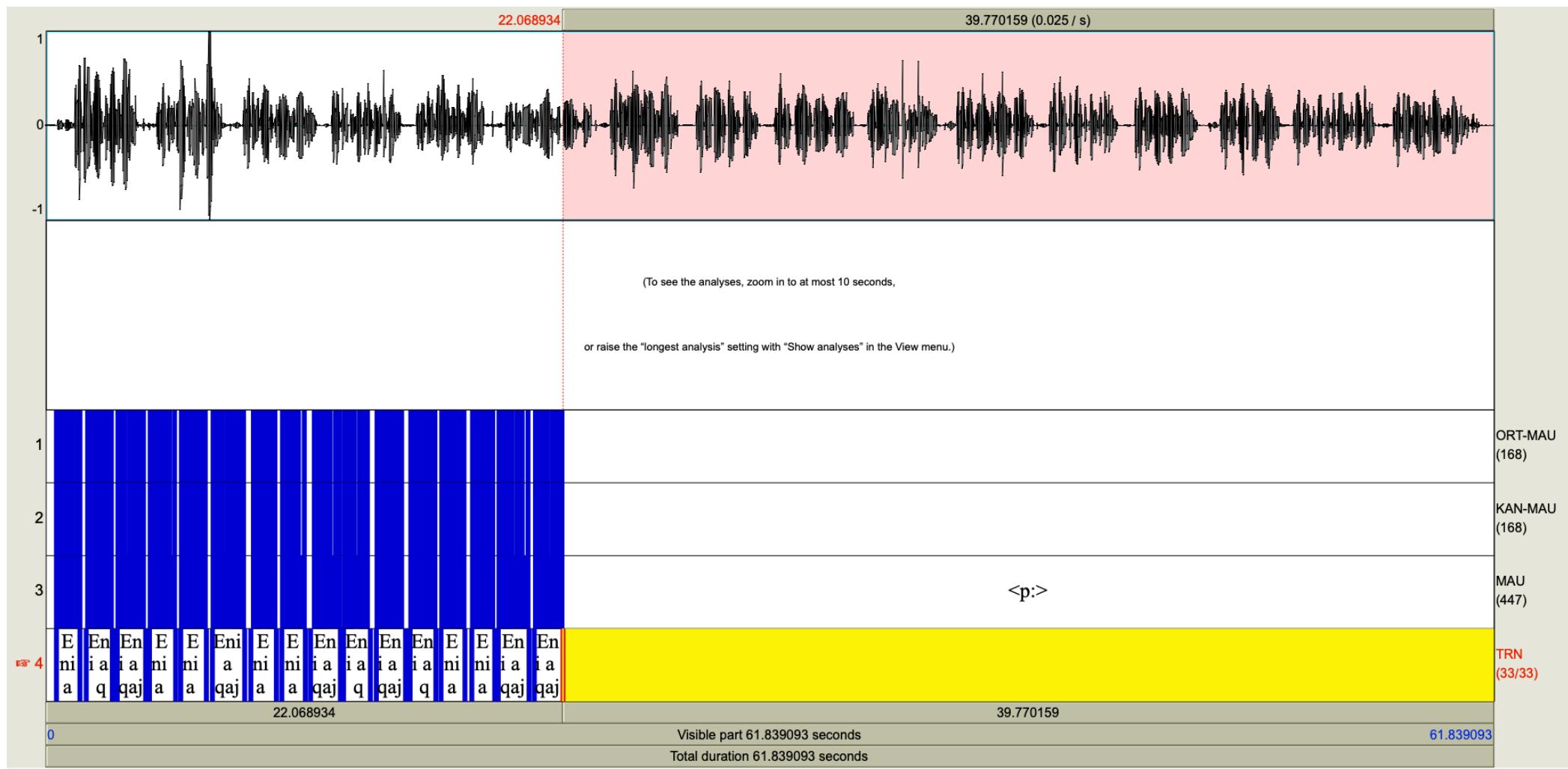


File inspection

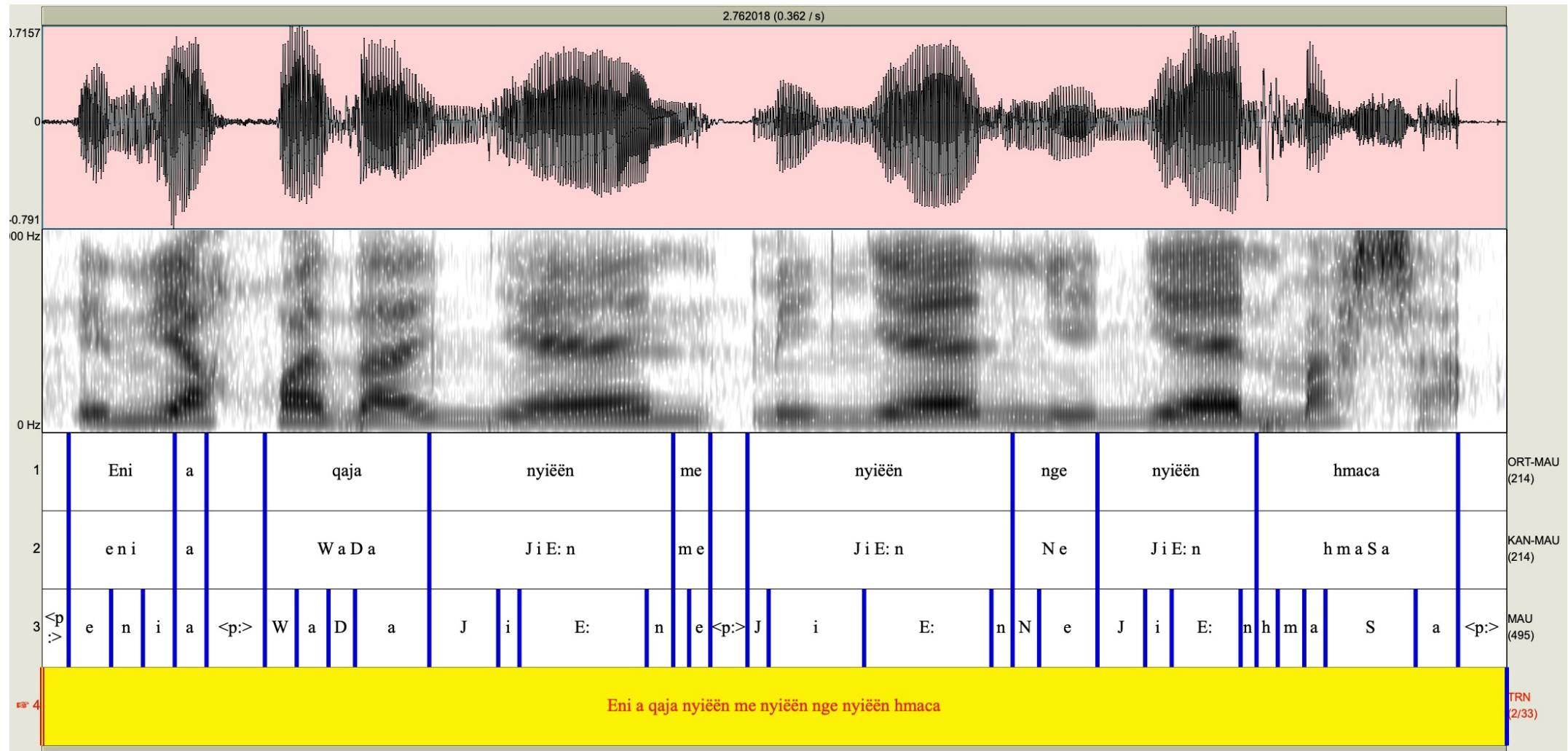
Successful forced-alignment



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>

Show service sidebar >

BAS Web Services

Version 3.12 · History of changes

Home General Help + FAQs Publications Contact, About, Privacy

Welcome!

The BAS Web Services are a rich set of tools for speech sciences and technology. Starting with MAUS – automatic segmentation and labelling of speech – many tools were developed in the context of [CLARIN-D](#).

Developing these tools is scientific work. Please cite the tools in your publications – every tool comes with a reference, and there is a [publications page](#).

For further information, there is a [General Help](#) page with tutorials, FAQs, etc., and on every service page is a specific manual section. Most tools can be used via this graphical frontend, or via a programmatic [REST API](#).

By using the services, you accept the [Conditions of Use](#) of these services.



(phonological) or spontaneous (phonetic) speech transcriptions.

pairs, for example the optimal alignment of an orthographic string to its corresponding phonological transcript.

Chunk Preparation

This pre-processor to MAUS transforms a chunk segmentation (CSV, EAF or TextGrid) into a BAS Partitur Format (BPF) file containing the tiers tokenized words (ORT) and chunk segmentation (TRN).

WebMINNI

This service segments and labels a speech audio file into SAM-PA (or IPA) phonetic segments without any text/phonological input.

SpeakDiar

This service reads a media file (sound, video) and performs a speaker diarization (SD) based on the pyannote python library.

Pipeline without ASR

This is a service that combines two or more BAS webservices into a processing chain (pipeline) without Automatic Speech Recognition (ASR).

Chunk Preparation

Files

Please drag & drop the annotation files from which the chunk segmentation is extracted, e.g. 'file1.TextGrid', 'file2.eaf' (allowed formats are: textgrid, eaf, csv).

Selected files (not yet uploaded):

- 1. DemoDrehu.TextGrid

Upload Delete all

Service options

Language: Language independent, use with X-SAMPA input ?

Input format: tg ?

Input tier name: text ?

Sampling rate: 44100 ?

Keep annotation: yes ?

Run

I have read and accepted the [terms of usage](#) for this service, including the policy of monitoring access to the services (paragraph 5). I hereby confirm that I am a member of an academic institution or that I have obtained a BAS user license for this service. In case of a publication of my results I will use a proper citation to this service.

Run Web Service

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)

Results (1)

DemoDrehu.par

Download as ZIP-File

2. G2P

Files successfully uploaded:

1. [DemoDrehu.par](#)

[Delete all](#)

Service options

 Input format  Input TextGrid tier  Sample rate  Language  Imap mapping file  Imap: case insensitive  Output format  Output Symbol inventory  Word stress  Syllabification  Text normalization  Keep annotation markers  Alignment  Tool embedding  Feature set	<input type="text" value="bpf"/> ? <input type="text" value="text"/> ? <input type="text" value="44100"/> ? <input type="text" value="User defined"/> ? <div style="border: 1px solid red; padding: 2px; display: inline-block;">Choose file DemoDrehu.txt</div> ? <input type="text" value="yes"/> ? <input type="text" value="bpf"/> ? <input type="text" value="sampa"/> ? <input type="text" value="no"/> ? <input type="text" value="maus"/> ? <input type="text" value="standard"/> ?
---	---

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)

Results (1)

DemoDrehu.par
Download as ZIP-File

3. WebMAUS General

WebMAUS General

Files

Files successfully uploaded:

1. DemoDrehu.par <=> DemoDrehu.wav

 Delete all

Service options

 Language

Show inventory

Language indep. (sampa)



 MAUS modus

Forced alignment to input transcript



Input Encoding

X-SAMPA (ASCII)



Output format

Praat (TextGrid)



 Expert Options (click to show)

Steps:

1. Drag & drop (second) .par + .wav
2. Upload
3. Adjust options:
 - Language = Language indep.
 - MAUS modus = Forced alignment..

➤ Click on **Expert options**:

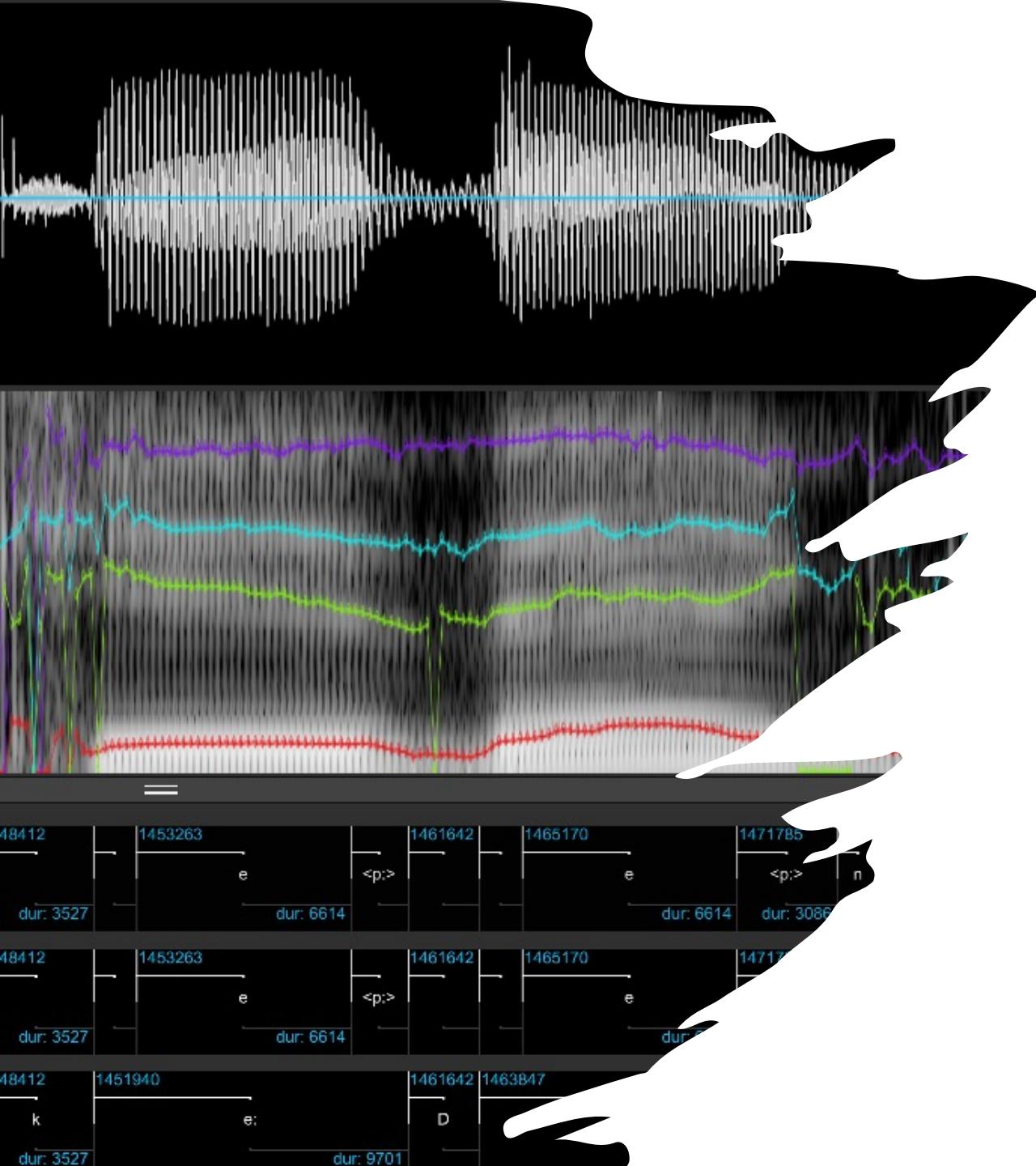
 - Chunk segmentation = true
 - All other options default
4. Run service (it will take some time)
5. Download zip file
6. Unzip file
7. Open new .TextGrid in Praat

Expert Options (click to hide):

Output Encoding	X-SAMPA (ASCII)	
Rule set file	Choose file No file chosen	
Chunk segmentation	true	
Pre-segmentation	false	
Inter-word silence	5	
Start with word	0	
End with word	999999	
Segment shift	default	
Phon insertion prob	0.0	
KAN tier in TextGrid	true	
ORT tier in TextGrid	true	
No silence model	false	
Pron model weight	default	
Relax Min Duration	false	
Three State Min Duration	false	
Output frame rate	10msec	
Add Viterbi likelihoods	false	

A close-up photograph of a stack of papers. The top sheet is white, followed by several green and brown ones. A silver paperclip is visible on the left side, holding some of the papers together.

How are your
files looking?



DemoDrehu_emuDB

- 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 it out the R script you need to have R + R studio installed
- You will also need additional packages:
"wrassp", "emuR", "ggplot2", "dplyr",
"tidy", "argparse", "gridExtra", "docopt"

Formant Analysis

Formant Analysis

1. DemoDrehu.TextGrid <=> DemoDrehu.wav

 Delete all

Service options

Language (required)	<input type="button" value="Show inventory"/>	Language Independent (requires 'Imap mapping file')	
EMU database Name	<input type="text" value="DemoDrehu"/> 		
List of vowels	<input type="text" value="a,o,e"/> 		
Speaker gender	<input type="text" value="female"/> 		
Input tier name (optional)	<input type="text" value="MAU"/> 		
Imap mapping file (optional)	<input type="button" value="Choose file"/>	Drehu.txt	
Select from mid point (optional)	<input type="text" value="true"/> 		
Compute eRatios (optional)	<input type="text" value="false"/> 		
Outlier detection metric	<input type="text" value="euclid"/> 		
Outlier detection threshold	<input type="text" value="250"/> 		

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).
- Gnevshева, 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.