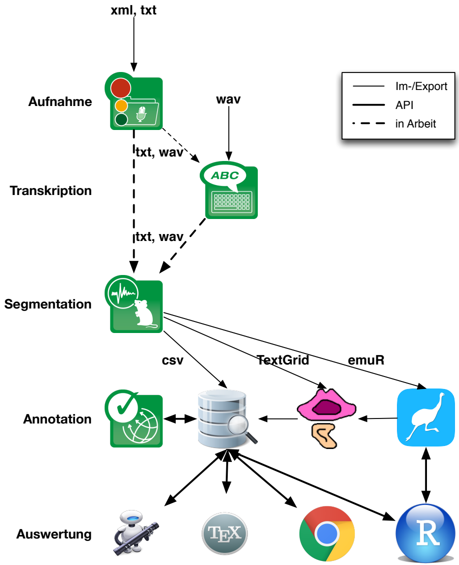
CLARIN-D New Web Services at BAS

Christoph Draxler
Florian Schiel, Thomas Kisler, Julian Pömp

25.04.2018

Workflow, tools and data



Current work at BAS

- speech recordings via browser or app
- external ASR
- editors in the workflow
- chunker
- pipeline services

Annotation: experiences

Task	Data	Cost	Time factor
chunking[1]	edit list	€	2
raw transcriptiton	orthography, markup	€	10-25
canonical transcription	SAMPA	€€	60
auditive transcription	IPA	€€€	300
manual segmentation	IPA, timestamps	€€€	1200

[Kva93], [WMA^{+}11]

[1] visual and/or auditive setting of boundaries in the signal

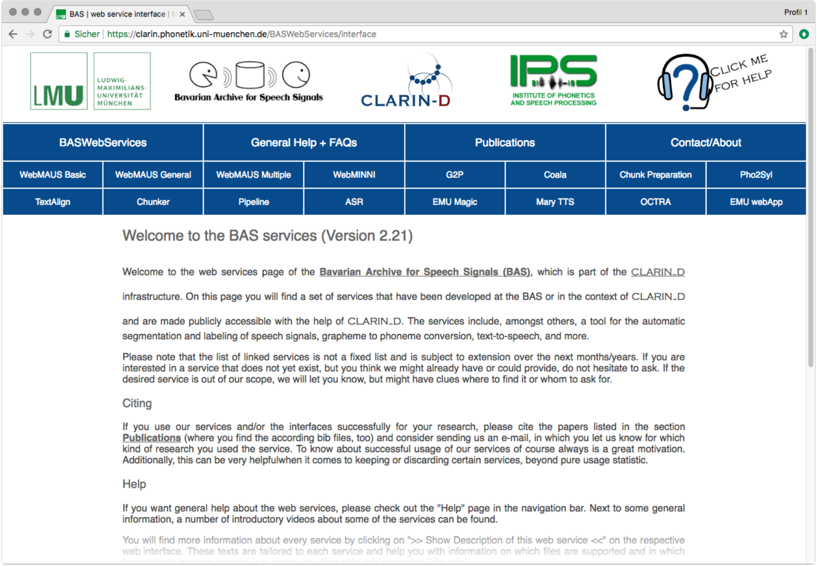
New frontiers

For *well-resourced* languages

- improve ASR of *difficult* audio signals
- optimise transcription task

For *less-resourced* languages

- provide manually prepared materials for training automatic services
- i. e. spontaneous speech of many speakers, manual phonetic segmentation, pronunciation dictionaries



Welcome to the BAS services (Version 2.21)

Welcome to the web services page of the **Bavarian Archive for Speech Signals (BAS)**, which is part of the CLARIN-D infrastructure. On this page you will find a set of services that have been developed at the BAS or in the context of CLARIN-D and are made publicly accessible with the help of CLARIN-D. The services include, amongst others, a tool for the automatic segmentation and labeling of speech signals, grapheme to phoneme conversion, text-to-speech, and more.

Please note that the list of linked services is not a fixed list and is subject to extension over the next months/years. If you are interested in a service that does not yet exist, but you think we might already have or could provide, do not hesitate to ask. If the desired service is out of our scope, we will let you know, but might have clues where to find it or whom to ask for.

Citing

If you use our services and/or the interfaces successfully for your research, please cite the papers listed in the section **Publications** (where you find the according bib files, too) and consider sending us an e-mail, in which you let us know for which kind of research you used the service. To know about successful usage of our services of course always is a great motivation. Additionally, this can be very helpfulwhen it comes to keeping or discarding certain services, beyond pure usage statistic.

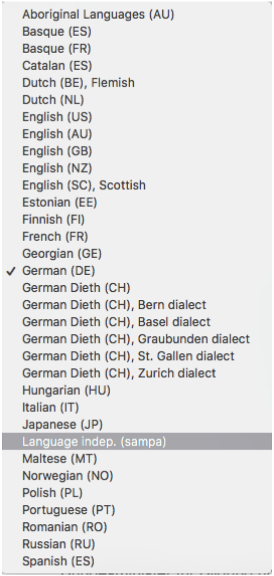
Help

If you want general help about the web services, please check out the "Help" page in the navigation bar. Next to some general information, a number of introductory videos about some of the services can be found.

You will find more information about every service by clicking on ">> Show Description of this web service <<" on the respective web interface. These texts are tailored to each service and help you with information on which files are supported and in which format those have to present (e.g. mono wav files only, tab separated lists, etc.)

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

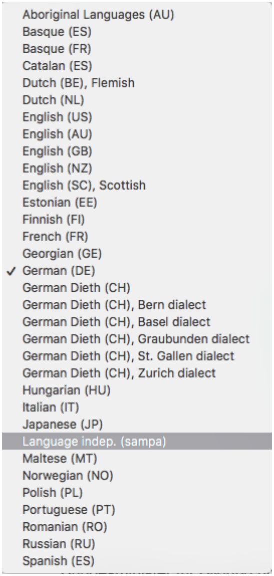
MAUS: Languages



Thank you!

- several variants of Swiss German
- collaboration with Uni ZH

MAUS: Languages



Thank you!

- several variants of Swiss German
- collaboration with Uni ZH

Your language not here?

- try the language independent settings
- send us a corpus of your language!

Automatic Speech Recognition ASR

Hey Siri! Google won't listen and Alexa is busy buying stuff I don't need!

Hey Siri! Google won't listen and Alexa is busy buying stuff I don't need!

ASR seems to work. Why not use it?

ASR works well, given

- well-resourced languages
- near-field microphone signals or microphone arrays
- processing power
- specific contexts
- standard transcripts

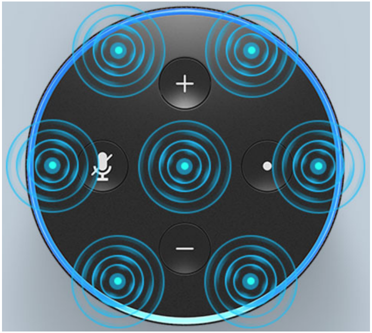
ASR seems to work. Why not use it?

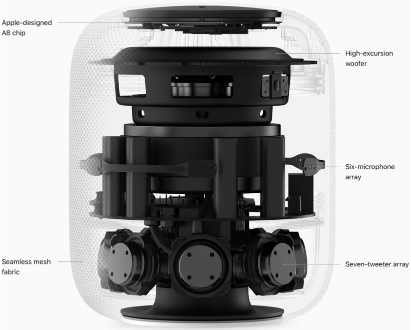
ASR works well, given

- well-resourced languages
- near-field microphone signals or microphone arrays
- processing power
- specific contexts
- standard transcripts

This is *not* what we have – and maybe not even want.

Inside Amazon Echo and Apple HomePod





https://www.amazon.de/dp/B06ZXQV6P8

https://www.apple.com/uk/homepod/

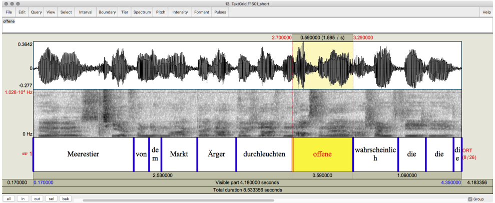
ASR as a web service

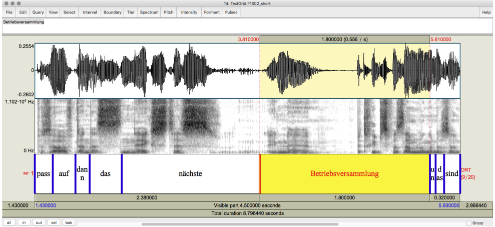
Use ASR to generate a raw orthographic transcript

- ASR interfaces available from third party providers
- some restrictions apply (max. duration, quota . . .)
- commercial providers store the audio signal (!)
- quality of the result varies greatly

Then, correct the ASR output manually[KRS17].

ASR demo: 2 signal conditions





ASR vs. manual transcription

Haven on Demand	**Google**	**EML**	**manual**
und	und	und	und
Saft	pass	das	pass
	auf	auch	auf
an	dann	dann	dann
dessen	das	das	als
nächstes	nächste	nächste	nächstes
wegen		das	
		innen	irgendeine
Betriebsversammlungen	Betriebsversammlung	Betriebsversammlungen	Betriebsversammlung
	und		oder
	das	aus	so
seien	sind	und	und
die	die	die	die
Chefin	Chefin	Chefin	Chefin
von	von	von	von
diese	dieser	dieser	dieser
	diese	zur	diese
Filial	Filialkette	Filialkette	Filialkette
Kette			
30	**22**	**27**	

far-field microphone, studio environment, Levenshtein distance on characters

ASR supported by BAS Web Services

- **Google**: commercial, many languages, max. 10s
- **HP Haven on Demand**: commercial, limited set of languages
- **IBM Watson**: commercial, limited set of languages, monthly quota
- **European Media Lab**: non-commercial, limited set of languages
- **Radboud University**: academic, limited set of languages

CLARIN login required!

Octra – Transcription editor(s) for raw transcripts

Orthographic transcription – why?

ASR simply is not good enough for

- noisy signals
- under-resourced languages
- particular speaking styles
- transcriptions with markup
- …

Humans are *incredibly flexible*: it often takes only a few minutes to adapt to a speaker or a noisy condition

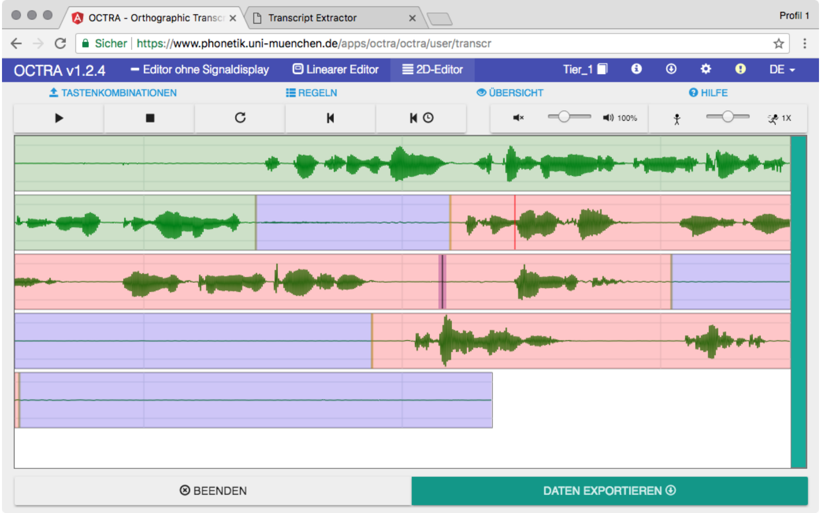
Octra motivation

Octra was developed from scratch, with efficiency as the main design goal

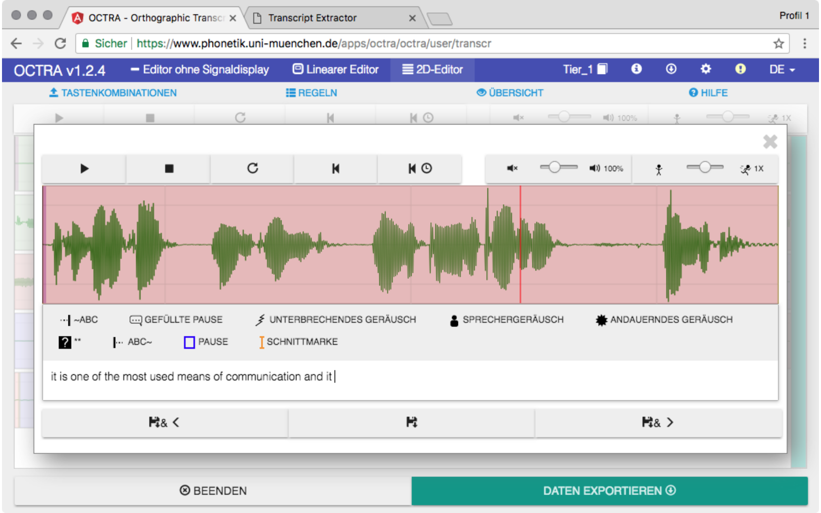
- web application – no installation
- local, online and URL mode of operation
- three different editors
- various import and export formats
- . . .

Octra is developed by Julian Pömp and Christoph Draxler [PD17]

2D-Editor


OCTRA - Orthographic Transcr
Transcript Extractor
Profil 1
Sicher
https://www.phonetik.uni-muenchen.de/apps/octra/octra/user/transcr
OCTRA v1.2.4
Editor ohne Signaldisplay
Linearer Editor
2D-Editor
Tier_1
DE
TASTENKOMBINATIONEN
REGELN
ÜBERSICHT
HILFE
100%
1X
BEENDEN
DATEN EXPORTIEREN

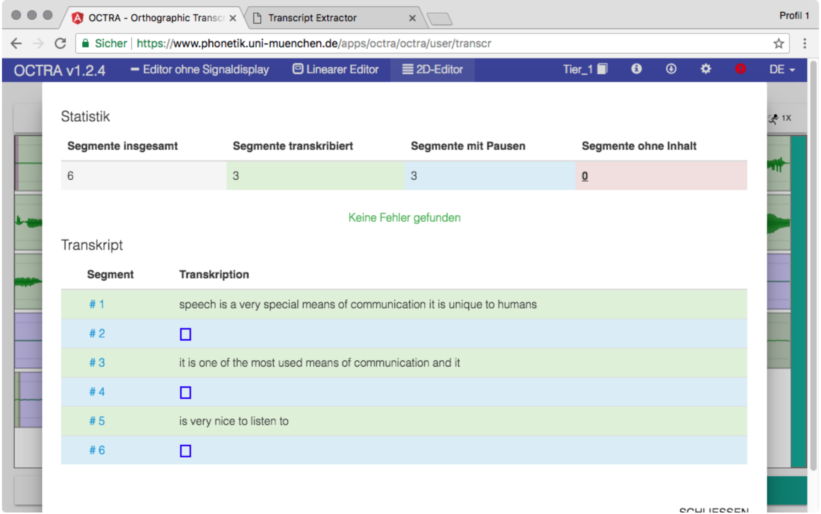
Detail editor



https://www.phonetik.uni-muenchen.de/apps/octra/octra/user/transcr

it is one of the most used means of communication and it

Check transcripts in overview



OCTRA - Orthographic Transcr
Transcript Extractor
Profil 1
Sicher | https://www.phonetik.uni-muenchen.de/apps/octra/octra/user/transcr
OCTRA v1.2.4
Editor ohne Signaldisplay
Linearer Editor
2D-Editor
Tier_1
DE
1X

Statistik

Segmente insgesamt	Segmente transkribiert	Segmente mit Pausen	Segmente ohne Inhalt
6	3	3	0

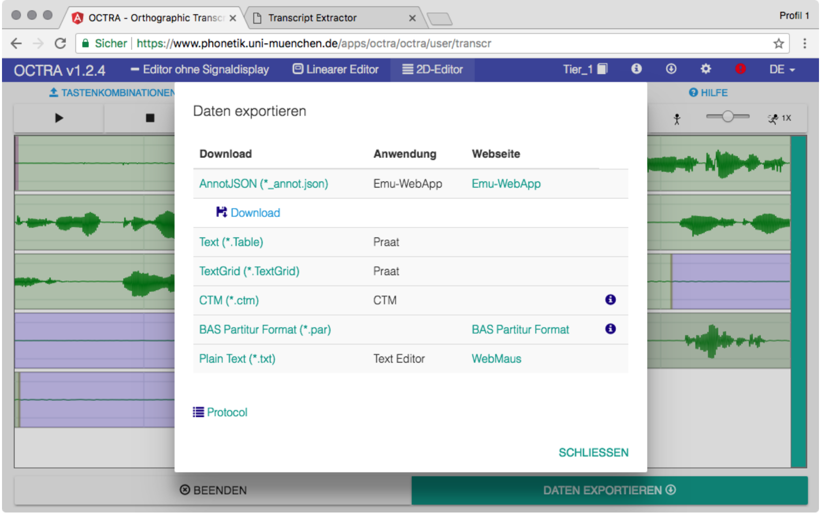
Keine Fehler gefunden

Transkript

Segment	Transkription
# 1	speech is a very special means of communication it is unique to humans
# 2	□
# 3	it is one of the most used means of communication and it
# 4	□
# 5	is very nice to listen to
# 6	□

SCHLIESSEN

Export transcripts



OCTRA - Orthographic Transcr
Transcript Extractor
Profil 1
Sicher | https://www.phonetik.uni-muenchen.de/apps/octra/octra/user/transcr
OCTRA v1.2.4
Editor ohne Signaldisplay
Linearer Editor
2D-Editor
Tier_1
DE
TASTENKOMBINATIONEN
HILFE
1X
BEENDEN
DATEN EXPORTIEREN

Daten exportieren

Download	Anwendung	Webseite	
AnnotJSON (*_annot.json)	Emu-WebApp	Emu-WebApp	
Download			
Text (*.Table)	Praat		
TextGrid (*.TextGrid)	Praat		
CTM (*.ctm)	CTM		ℹ
BAS Partitur Format (*.par)		BAS Partitur Format	ℹ
Plain Text (*.txt)	Text Editor	WebMaus	

Protocol

SCHLIESSEN

AnnotJSON-format

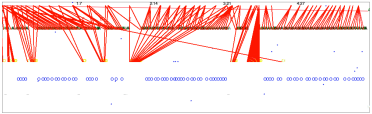
```
{"name": "DRCH0001Y1",
 "annotates": "DRCH0001Y1.wav",
 "levels": [
  {
   "name": "Tier_1",
   "type": "SEGMENT",
   "items": [
    {
     "id": 1,
     "labels": [
       {"name": "Tier_1",
        "value": "speech is a very special means of communication it is unique
     "sampleStart": 0,
     "sampleDur": 284721
   },
   ...
```

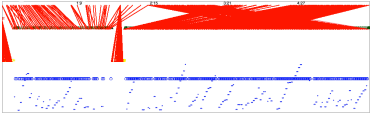
Octra – pilot study

Task: transcribe 3-5 minute long speech on "Communication"

- two transcribers, no prior experience with Octra
- manual correction of ASR output vs. full manual transcription
- basic transcription guidelines

Individual transcription styles and preferences!



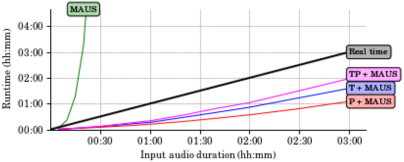


Chunker – processing long audio files

Chunker: Motivation

Chunker speeds up segmentation of long audio files

- WebMAUS requires $O(n^2)$ processing time
- practical limit approx. 20 min



Chunker was developed by Nina Poerner [PS16]

Chunker Procedure

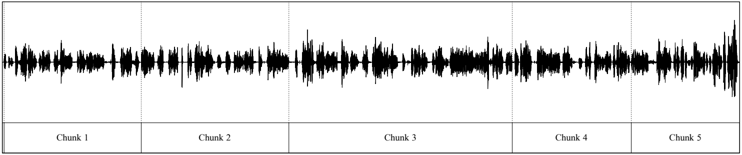
Chunker prerequisites: orthographic transcript and audio file

- generate raw transcript using ASR
- search for matching word sequences in ASR output and transcript
- extract words from manual transcript and cut audio file using ASR timestamps
- run WebMAUS for the paired text and audio fragments
- recombine everything

Chunker results

File `F1S02_SPM.wav` (length 2:01 minutes)

- 5 chunks
- length between 17 and 36 seconds
- nicely cut in longer pauses



Pipeline services – automating the workflow

Pipeline services: Motivation

User request: Simplify using web services!

- file upload and result downloads needed for every service
- which file formats work for which tool?
- too many options – with intransparent dependencies
- too much clicking. . .

There must be an easier way!

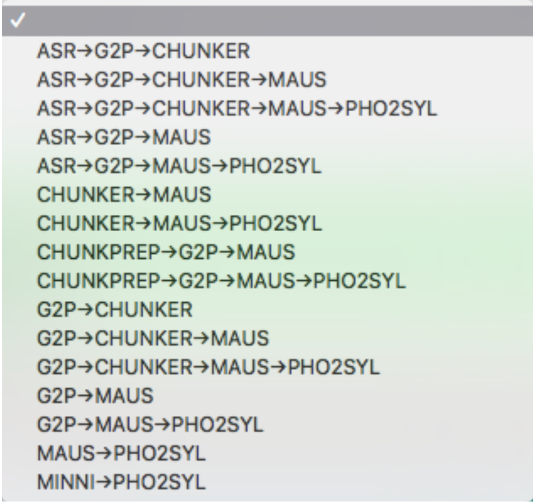
Pipeline services

Preconfigured sequences of tasks

- only one file upload needed
- default options are set
- expert options are still available, but ...
- notification with a download link via mail

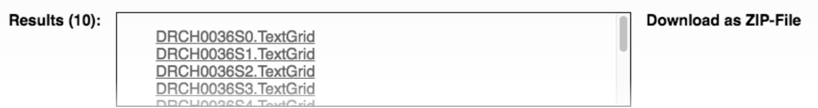
Pipeline services access new application areas, e. g. Oral History, qualitative sociology ...

Pipeline services



Pipeline services: results

Process result files in the browser or download them



DRCH0036S0.TextGrid
DRCH0036S1.TextGrid
DRCH0036S2.TextGrid
DRCH0036S3.TextGrid
DRCH0036S4.TextGrid

Several output formats available (BAS Partitur, AnnotJSON, TextGrid, CSV . . .)

Emu Web App – visualisation and editing

Emu WebApp: Motivation

Modern speech corpora are large and require collaborative organisation of work. This requires

- access to a speech database
- online and local mode of operation
- powerful visualisation of speech signals and annotations
- access to statistics package for analysis
- no software installation

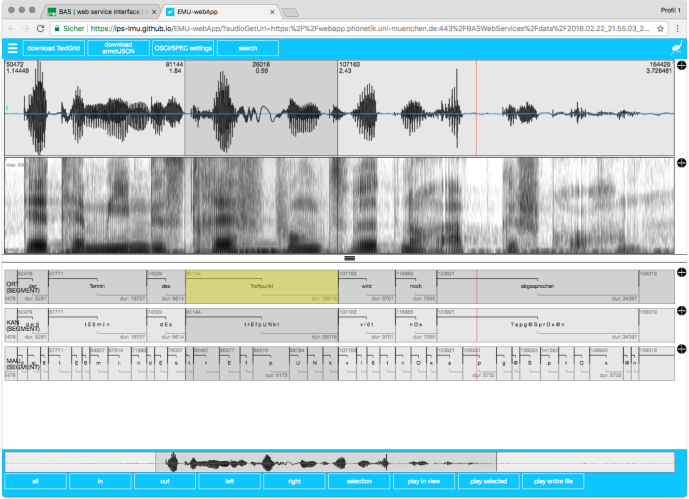
Emu WebApp: Motivation

Modern speech corpora are large and require collaborative organisation of work. This requires

- access to a speech database
- online and local mode of operation
- powerful visualisation of speech signals and annotations
- access to statistics package for analysis
- no software installation

Enter Emu WebApp by Raphael Winkelmann [WHJ17]

Emu WebApp



Check segmentations in the browser and correct them

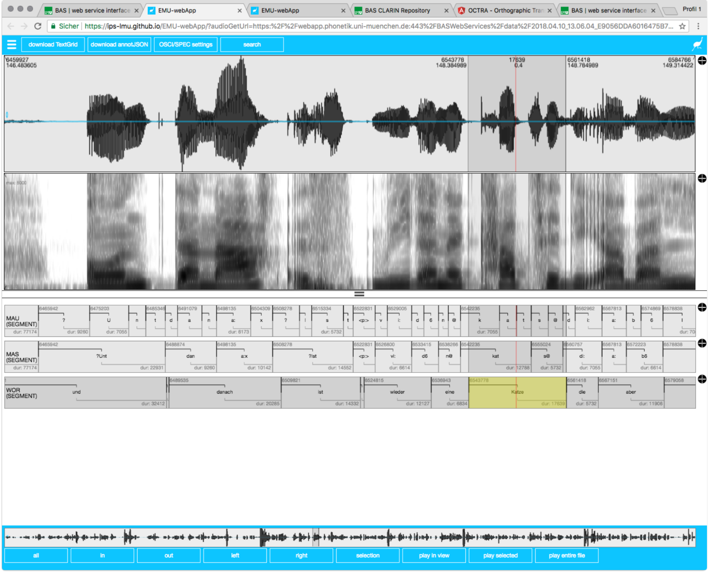
Finally, a bit of magic. . .

'Magic' web service

1. upload audio files
2. select 'Magic' service
3. wait...
4. download Emu database

Watch the demo!

Some time later. . .


BAS | web service interface
EMU-webApp
EMU-webApp
BAS CLARIN Repository
OCTRA - Orthographic Tran
BAS | web service interface
Profil 1
Sicher
https://ips-lmu.github.io/EMU-webApp/?audioGetUrl=https:%2F%2Fwebapp.phonetik.uni-muenchen.de:443%2FBASWebServices%2Fdata%2F2018.04.10_13.06.04_E9056DDA6016475B7...
download TextGrid
download annotJSON
OSCI/SPEC settings
search
6459927
146.483605
6543778
148.384989
17639
0.4
6561418
148.784989
6584766
149.314422
MAU (SEGMENT)
dur: 77174
? U n t d a n a: x ? I s t <p:> v i: d 6 n @ k a t s @ d i: a: b 6 l
MAS (SEGMENT)
dur: 77174
?Unt dan a:x ?Ist <p:> vi: d6 n@ kat s@ di: a: b6
WOR (SEGMENT)
und danach ist wieder eine Katze die aber
dur: 32412
dur: 20285
dur: 14332
dur: 12127
dur: 6834
dur: 17639
dur: 5732
dur: 11906
all
in
out
left
right
selection
play in view
play selected
play entire file

Summary

BAS web services are available today

- free access
- convenient pipeline services
- new services, e. g. speech recognition

The quality of the services depends on

- signal quality
- feedback to the BAS developers

Famous last words

Tool and service development is scientific work!

- both for the application field
- and (media)informatics

Support this work by publication and citation!

T. Kisler, U. Reichel, and F. Schiel.
Multilingual processing of speech via web services.
Computer Speech and Language, 45:326–347, 2017.

K. Kvale.
Segmentation and Labelling of Speech.
PhD thesis, Norwegian Institute of Technology, Trondheim, 1993.

Julian Pömp and Christoph Draxler.
OCTRA – A configurable browser-based editor for orthographic transcription.
In *Proceedings Phonetik und Phonologie*, pages 145–148, Berlin, 2017.

Nina Poerner and Florian Schiel.
An automatic chunk segmentation tool for long transcribed speech recordings.
In *Proceedings Phonetik und Phonologie*, pages 144–146, Munich, 2016.

Raphael Winkelmann, Jonathan Harrington, and Klaus Jänsch.
Emu-SDMS: Advanced Speech Database Management and Analysis in R.
Computer Speech and Language, 2017.

J. Williams, I. Melamed, T. Alonso, B. Hollister, and J. Wilpon.
Crowd-sourcing for difficult transcription of speech.
In *Proceedings IEEE Workshop on Automatic Speech Recognition and Unterstanding (ASRU 2011))*, 2011.