Tools, strategies, and resources in corpus phonetics

Luke Annear and Emily Bagan
WiGL 19
March 15, 2025

About us

Luke Annear

- Speech language pathologist
- 3rd year PhD student in Language Sciences
- Extensive lab experience with forced alignment and pipelines for processing large phonetic corpora
- L1 phonological acquisition; Speech sound disorders; phonetics & phonology; laryngeal phonetics and phonology

Emily Bagan

- Speech language pathologist
- 3rd year PhD student in Communication Sciences & Disorders
- Project assistant in Dr. Margarita Kaushanskaya's Language Acquisition & Bilingualism lab
- Bilingual phonological acquisition & development; processing & learning; perception & production; speech sound disorders

Why this workshop?

- Collaborating on phonetics project involving acoustic analysis
  - Several rounds of creating, modifying, and organizing our data
- Experience creating tools for data processing pipelines
- Many of these tools are general and useful for everyone

Research Questions

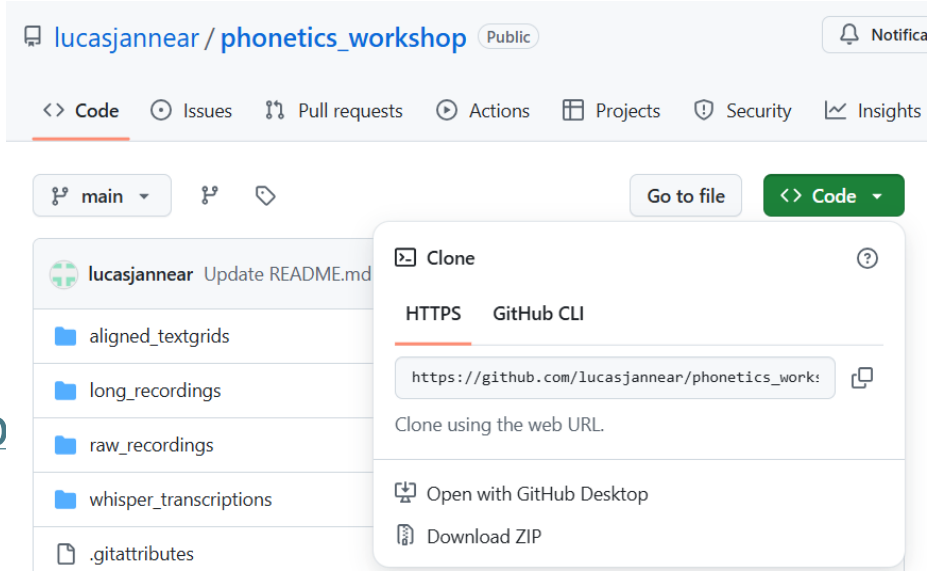
- **Voice onset time**
- Articulation rate
- Vowel formants/vowel space

What we've done

- Recorded two passages each
  - *Rainbow passage* (reading task)
  - *Frog where are you?* (spontaneous speech task)
- Created transcriptions of each recording
- Created phonetically-segmented Praat textgrids
- Created R Notebook for:
  - reading in textgrid data
  - querying textgrid data
  - analyzing textgrid data
- Github repository: https://github.com/lucasjannear/phonetics_workshop

Follow along…

- Download the directory: https://github.com/lucasjannear/phonetics_workshop
- Transcribe the files*
    - Perform checks on the transcription
- Force align the files*
- Modify the files with Praat scripts
- Read the data into the provided R document



*following along live may not be possible on these steps

Learning outcomes

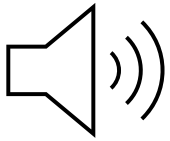
- Principles
  - Work done to answer a research question should facilitate work on future research questions
- Tools
  - *Whisper* for automated transcription
  - *Montreal Forced Aligner* for phone segmentation in Praat
  - Simple Praat scripts
- Methods
  - File organization and naming
  - Praat scripts to automate routine tasks
  - Reading textgrids in R, querying environments

Learning outcomes

- Creating…
- Managing…
- Modifying…
- Querying…

…a phonetic corpus

So you've collected some kind of speech/language sample...



File naming and organization

- Naming
  - What to encode?
    - speaker ID/number
    - speaker sex
    - speaker age
    - other variables?

File naming and organization

- Naming
  - What to encode?
    - speaker ID/number
    - speaker sex
    - speaker age
    - other variables?
- Organization
  - By task and then speaker?
    - `spontaneous_recordings/`
      - `speaker01.wav`
      - `speaker02.wav`

File naming and organization

- Naming
  - What to encode?
    - speaker ID/number
    - speaker sex
    - speaker age
    - other variables?
- Organization
  - By task and then speaker?
    - `spontaneous_recordings/`
      - `speaker01.wav`
      - `speaker02.wav`
  - By speaker and then task?
    - `speaker01_recordings/`
      - `speaker01_spontaneous.wav`
      - `speaker01_reading_passage.wav`

File Organization

- `phonetics_workshop/long_recordings/`
  - `f_01`
    - `f_01_frog_where_are_you.wav`
    - `f_01_rainbow_passage.wav`
  - `m_01`
    - `m_01_frog_where_are_you.wav`
    - `m_01_rainbow_passage.wav`

Transcription

- Automated transcription
  - Converting spoken word (.mp3/.mp4/.wav) into written text

Transcription

- Automated transcription
  - Converting spoken word (.mp3/.mp4/.wav) into written text
- Six model sizes (four are English-only)

Size	Parameters	English-only model	Multilingual model	Required VRAM	Relative speed
tiny	39 M	`tiny.en`	`tiny`	~1 GB	~10x
base	74 M	`base.en`	`base`	~1 GB	~7x
small	244 M	`small.en`	`small`	~2 GB	~4x
medium	769 M	`medium.en`	`medium`	~5 GB	~2x
large	1550 M	N/A	`large`	~10 GB	1x
turbo	809 M	N/A	`turbo`	~6 GB	~8x

Transcription

- Automated transcription
  - Converting spoken word (.mp3/.mp4/.wav) into written text
- Six model sizes (four are English-only)
- Identifies language in first :30
- Provides translations

Transcription

- Automated transcription
  - Converting spoken word (.mp3/.mp4/.wav) into written text
- Six model sizes (four are English-only)
- Identifies language in :30
- Provides translations
- Phrase level time stamps

Automated Transcription

Benefits

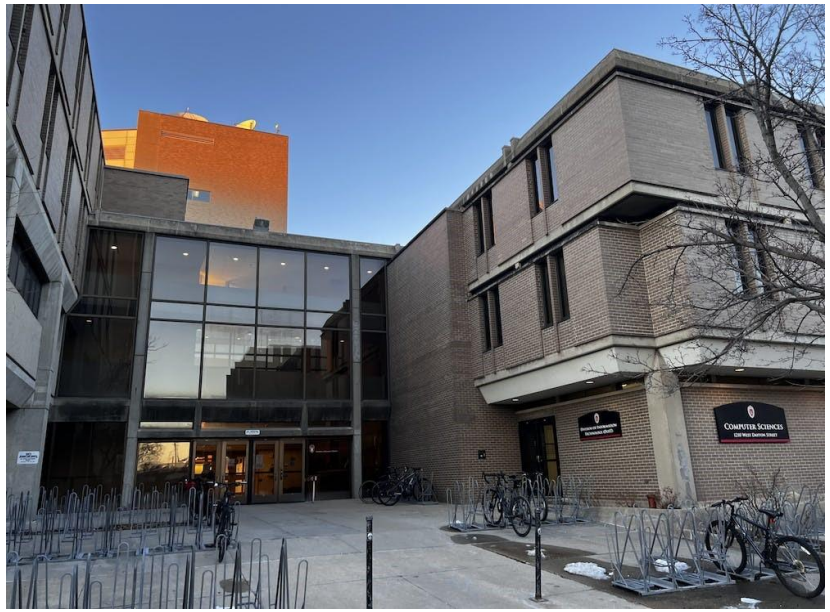
- Fast
- Language identification
- Translation

Drawbacks

- Accuracy
- Manual corrections
- Multiple speakers
- Multiple languages
- Trained on adult speech

Automated Transcription

- Helpful resources
  - Computer science department
    - Keep in mind data protection!
  - Center for high throughput computing (CHTC)



CHTC

Example

Example

The Rainbow Passage

When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow. Throughout the centuries people have explained the rainbow in various ways. Some have accepted it as a miracle without physical explanation. To the Hebrews it was a token that there would be no more universal floods. The Greeks used to imagine that it was a sign from the gods to foretell war or heavy rain. The Norsemen considered the rainbow as a bridge over which the gods passed from earth to their home in the sky. Others have tried to explain the phenomenon physically. Aristotle thought that the rainbow was caused by reflection of the sun's rays by the rain. Since then physicists have found that it is not reflection, but refraction by the raindrops which causes the rainbows. Many complicated ideas about the rainbow have been formed. The difference in the rainbow depends considerably upon the size of the drops, and the width of the colored band increases as the size of the drops increases. The actual primary rainbow observed is said to be the effect of superimposition of a number of bows. If the red of the second bow falls upon the green of the first, the result is to give a bow with an abnormally wide yellow band, since red and green light when mixed form yellow. This is a very common type of bow, one showing mainly red and yellow, with little or no green or blue.

- emily_rainbow.json
- emily_rainbow.srt
- emily_rainbow.tsv
- emily_rainbow.txt
- emily_rainbow.vtt
- emily_rainbow.wav
- luke_rainbow.json
- luke_rainbow.srt
- luke_rainbow.tsv
- luke_rainbow.txt
- luke_rainbow.vtt
- luke_rainbow.wav

emily_rainbow.txt - Notepad

File Edit Format View Help

```
When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow.
The rainbow is a division of white light into many beautiful colors.
These take the shape of a long, round arch, with its path high above and its two ends
apparently beyond the horizon.
There is, according to legend, a boiling pot of gold at one end.
People look, but no one ever finds it.
When a man looks for something beyond his reach, his friends say he is looking for the
pot of gold at the end of the rainbow.
Throughout the centuries, people have explained the rainbow in various ways.
Some have accepted it as a miracle without physical explanation.
To the Hebrews, it was a token that there would be no more universal floods.
The Greeks used to imagine that it was a sign from the gods to foretell war on heavy rain.
The Norsemen considered the rainbow as a bridge over which the gods passed from earth to their
home in the sky.
Others have tried to explain the phenomena physically.
Aristotle thought that the rainbow was caused by reflection of the sun's rays by the rain.
Since then, physicists have found that it is not reflection, but refraction, by the raindrops
which causes the rainbows.
Many complicated ideas about the rainbow have been formed.
The difference in the rainbow depends considerably upon the size of the drops, and the width
of the colored bands increases as the size of the drops increases.
The actu
of bows.
If the r
bow with
This is
or blue.
```

luke_rainbow.txt - Notepad

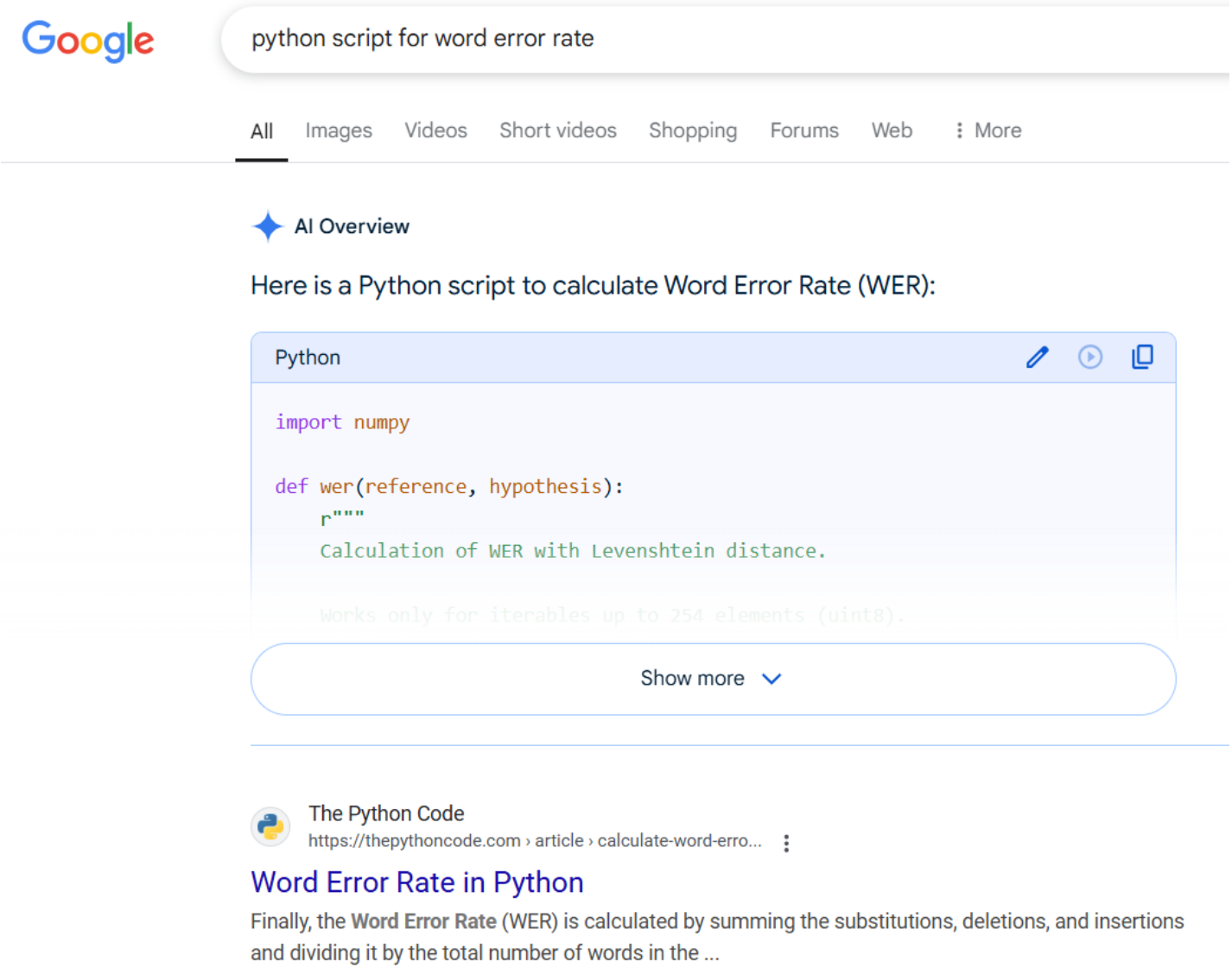
File Edit Format View Help

```
When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow.
The rainbow is a division of white light into many beautiful colors.
These take the shape of a long, round arch, with its path high above, and its two ends
apparently beyond the horizon.
There is, according to legend, a boiling pot of gold at one end.
People look, but no one ever finds it.
When a man looks for something beyond his reach, his friends say he is looking for the
pot of gold at the end of the rainbow.
Throughout the centuries, people have explained the rainbow in various ways.
Some have accepted it as a miracle without physical explanation.
To the Hebrews, it was a token that there would be no more universal floods.
The Greeks used to imagine that it was a sign from the gods to foretell war or heavy rain.
The Norsemen considered the rainbow as a bridge over which the gods passed, from earth to their
home in the sky.
Others have tried to explain the phenomenon physically.
Aristotle thought that the rainbow is caused by reflection of the sun's rays by the rain.
Since then, physicists have found that it is not reflection, but refraction, by the raindrops,
which causes the rainbows.
Many complicated ideas about the rainbow have been formed.
The difference in the rainbow depends considerably upon the size of the drops, and the width
of the colored bands increases as the size of the drops increases.
The actual primary rainbow observed is said to be the effect of superimposition of a number
of bows.
If the red of the second bow falls upon the green of the first, the result is to give a
bow with an abnormally wide yellow band, since red and green light when mixed form yellow.
This is a very common type of bow, one showing mainly red and yellow, with little or no green
or blue.
```

Ln 1, Col 1 | 100% | Windows (CRLF) | UTF-8

How reliable is whisper?

reading_passage.txt
automated_transcription.txt



How reliable is whisper?

Word Error Rate = Errors/Total Number of Words

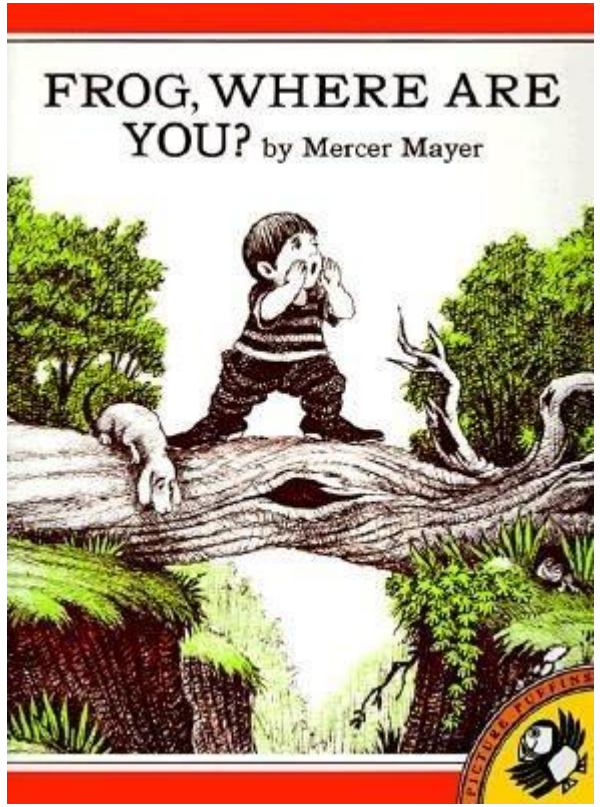
```
Correct: war
Missing in transcription: or
Extra in transcription: on
Correct: heavy
Correct: rain
```

3/329 = .009

```
Total number of discrepancies (unique errors): 6
```

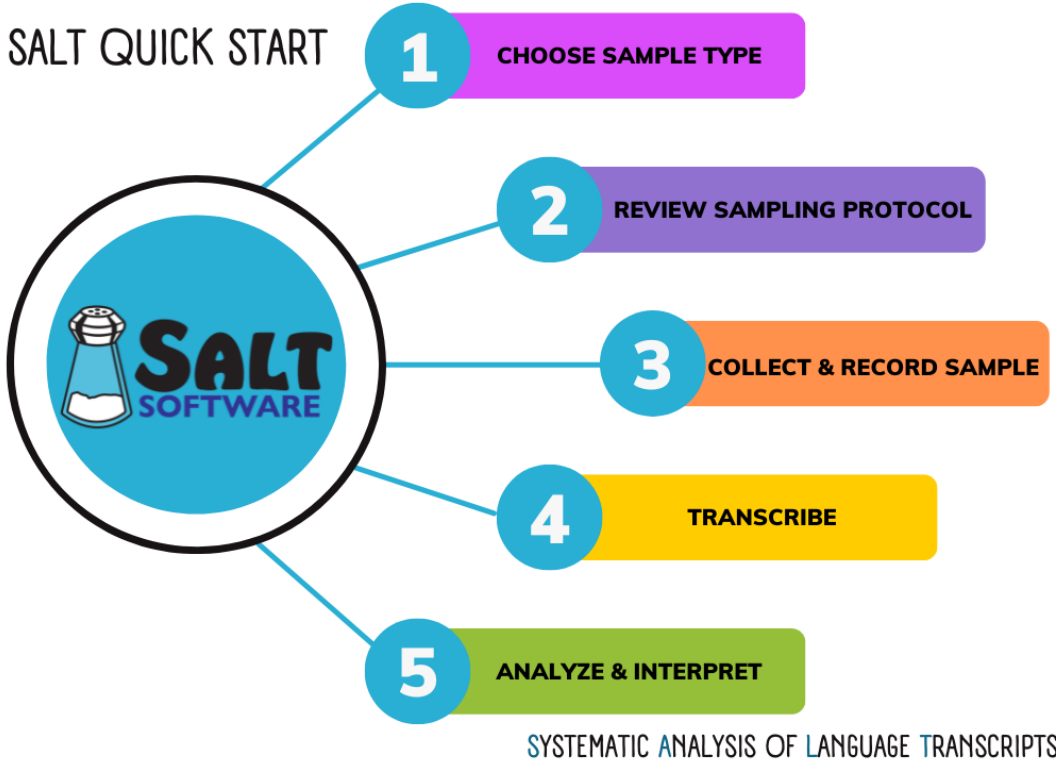
How reliable is whisper?

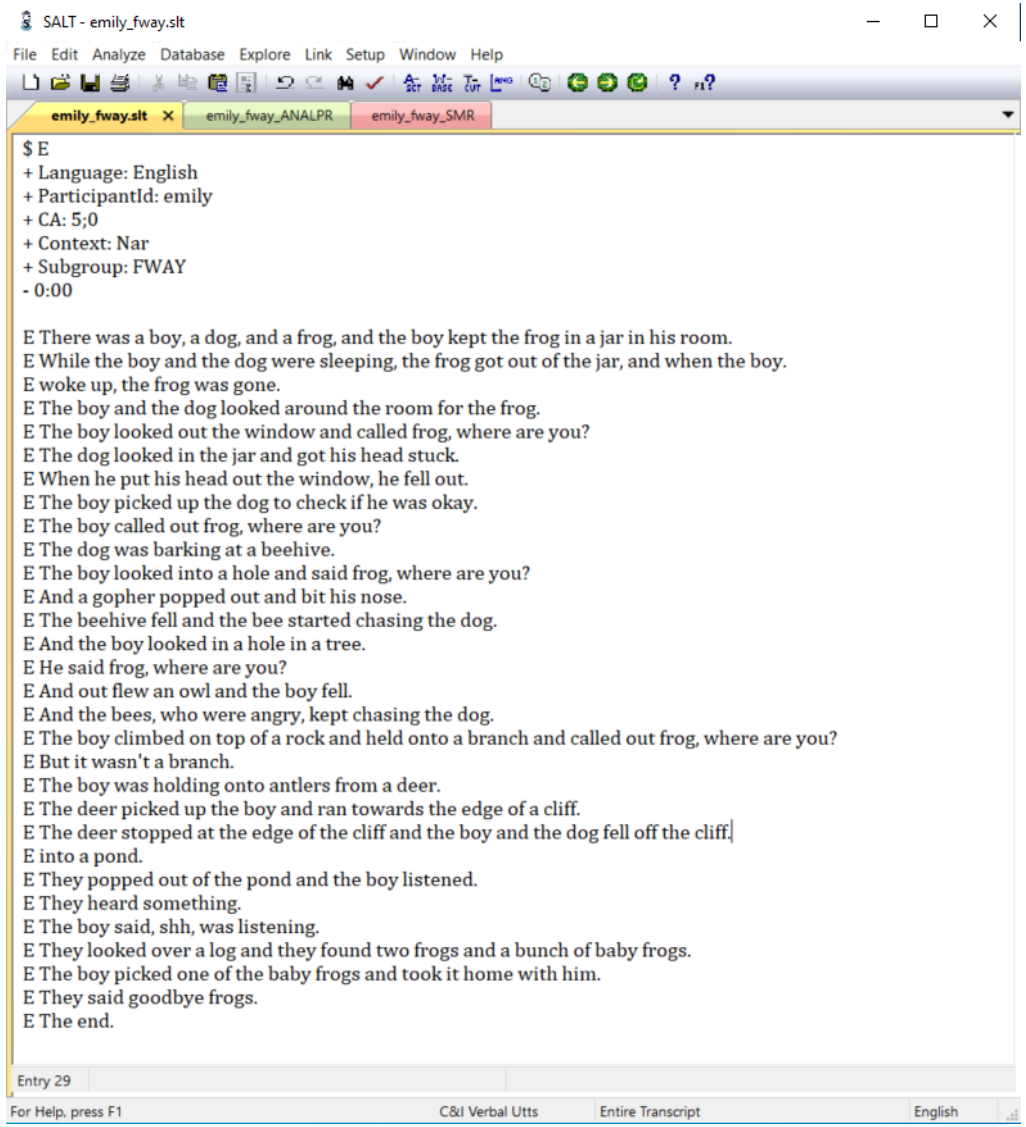
- 6 discrepancies
  - 5 corrections of revisions, fillers, false-starts
    - "**then the boy** – then the boy and his dog went in the backyard"
    - "the dog was jumping up **at a** – at a beehive hanging from a tree"
  - 1 missing word
    - "**And** the boy said I found my frog"



Now you've got transcribed samples ...

Systematic Analysis of Language Transcription





emily_fway

STANDARD MEASURES REPORT			
		E	***
TRANSCRIPT LENGTH			
	Total Utterances	30	0
#	Analysis Set (C&I Verbal Utts)	30	0
	All Words Including Mazes	310	0
	Elapsed Time	---	
INTELLIGIBILITY			
	% Intelligible Utterances	100%	---
	% Intelligible Words	100%	---
SYNTAX/MORPHOLOGY			
#	MLU in Words	10.33	---
#	MLU in Morphemes	10.33	---
#	Verbs/Utterance	1.90	---
SEMANTICS			
#	Number Total Words	310	0
#	Number Different Words	102	0
#	Type Token Ratio	0.33	---
#	Moving-Average TTR (100)	0.48	---
VERBAL FACILITY			
	Words/Minute	---	---
	Pauses Within Utterances	0	0
	Pauses Between Utterances	0	
	Pause Time as % of Total Time	---	
#	Maze Words as % of Total Words	0.0%	---
	Abandoned Utterances	0	0
ERRORS			
#	% Utterances with Errors	0.0%	---
	Number of Omissions	0	0
	Number of Error Codes	0	0

 # Calculations based on C&I Verbal Utts

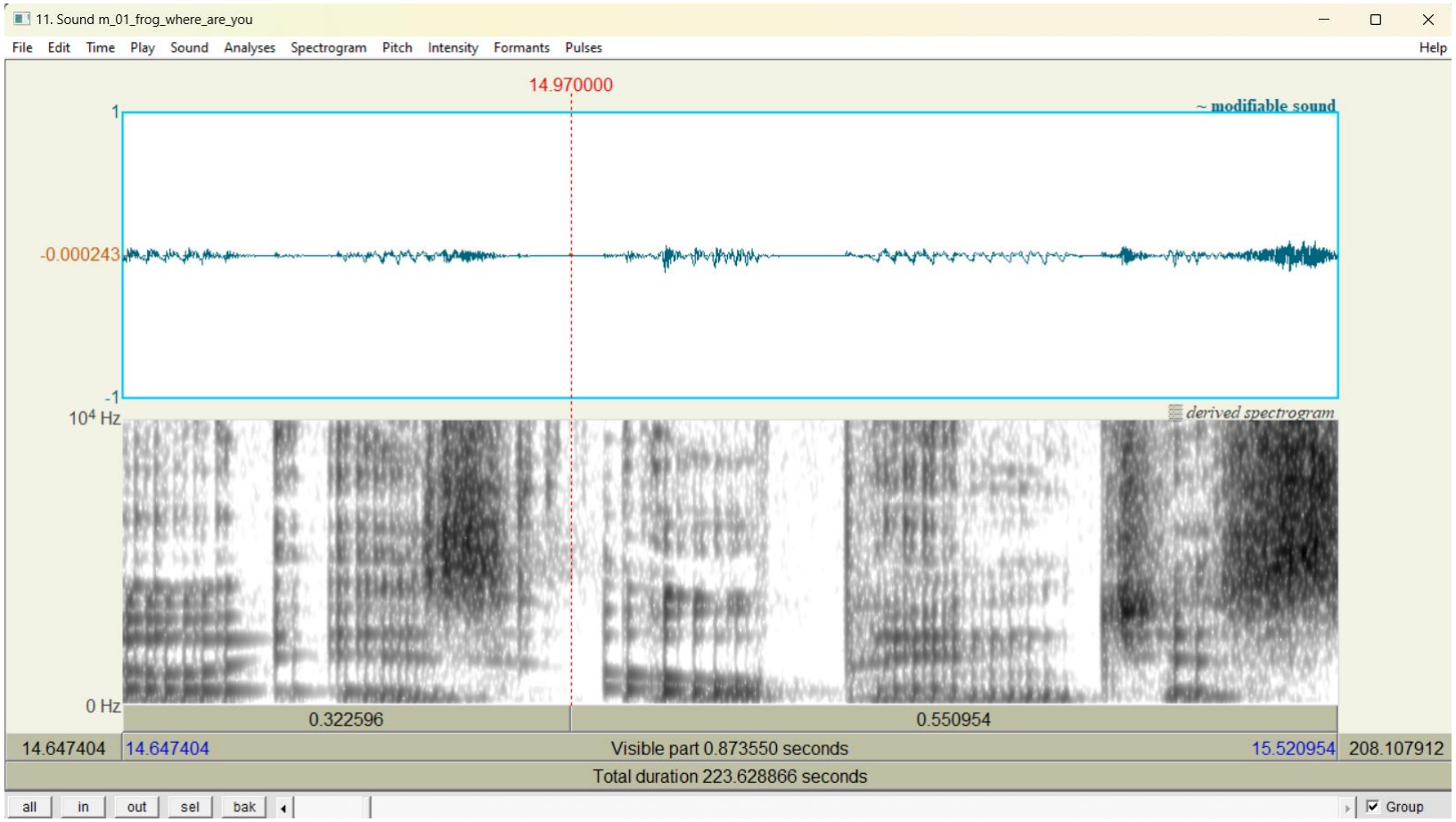
Check-in

- Questions about transcription

Check-in

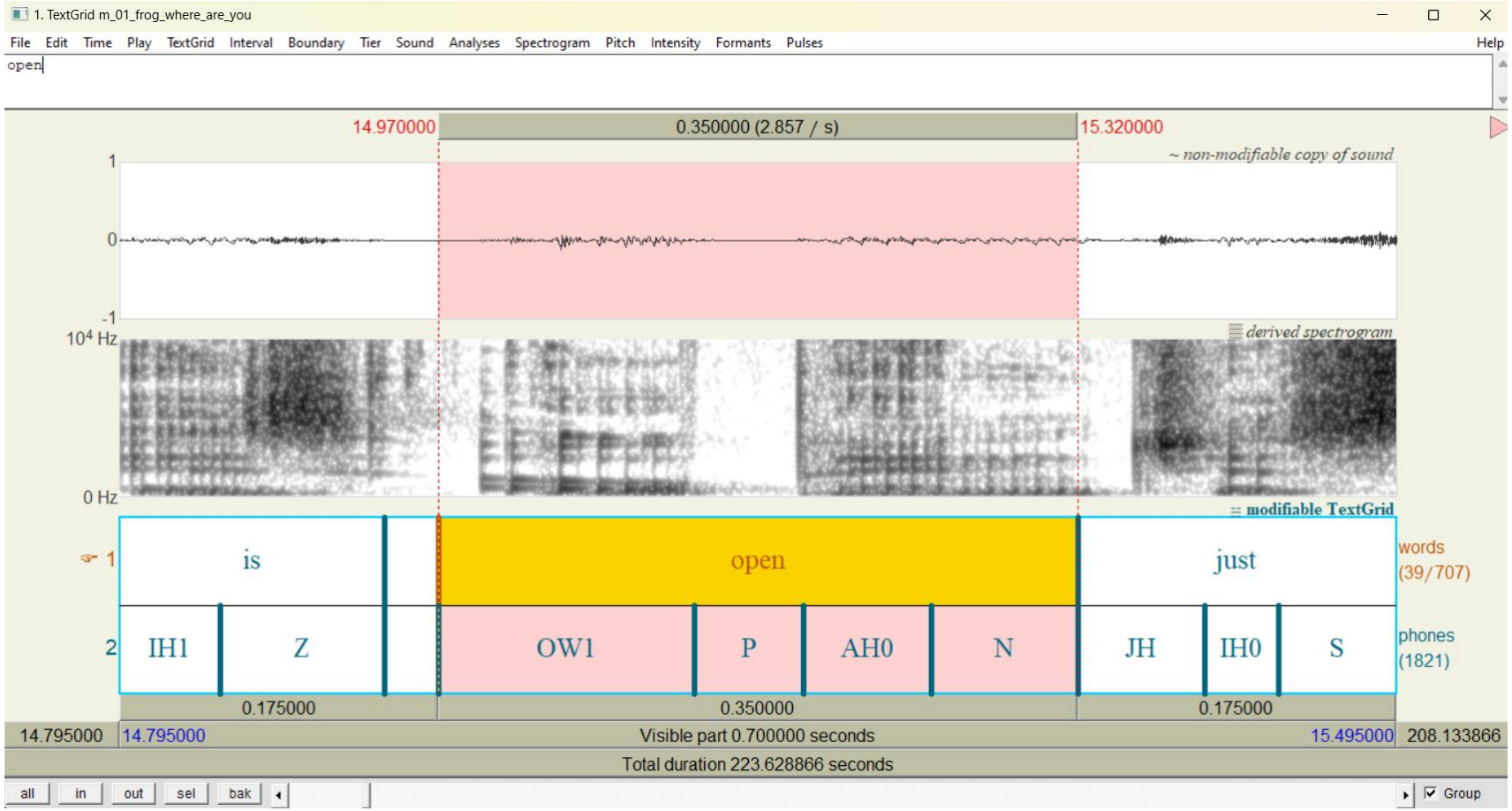
- Praat for phonetic analysis: https://www.fon.hum.uva.nl/praat/

Forced alignment



Forced alignment

Montreal Forced Aligner



Forced alignment

- Other languages: https://mfa-models.readthedocs.io/en/latest/dictionary/index.html#dictionary

Dictionaries G2P models Acoustic models Language models Ivector extractors More

Show 10 entries Columns Copy Excel PDF Search:

ID	Language	Dialect	Phoneset	License
Abkhaz CV dictionary v2_0_0	Abkhaz	N/A	XPF	CC BY 4.0
Arabic MFA dictionary v2_0_0	Arabic	N/A	MFA	CC BY 4.0
Armenian CV dictionary v2_0_0	Armenian	N/A	XPF	CC BY 4.0
Bashkir CV dictionary v2_0_0	Bashkir	N/A	XPF	CC BY 4.0
Basque CV dictionary v2_0_0	Basque	N/A	XPF	CC BY 4.0
Belarusian CV dictionary v2_0_0	Belarusian	N/A	XPF	CC BY 4.0
Bulgarian CV dictionary v2_0_0	Bulgarian	N/A	XPF	CC BY 4.0
Bulgarian MFA dictionary v2_0_0	Bulgarian	N/A	MFA	CC BY 4.0
Bulgarian MFA dictionary v2_0_0a	Bulgarian	N/A	MFA	CC BY 4.0
Bulgarian MFA dictionary v3_0_0	Bulgarian	N/A	MFA	CC BY 4.0

Showing 1 to 10 of 142 entries Previous 1 2 3 4 5 ... 15 Next

Corpus format

- `phonetics_workshop/long_recordings/`
  - `f_01`
    - `f_01_frog_where_are_you.wav`
    - `f_01_frog_where_are_you.txt`
    - `f_01_rainbow_passage.wav`
    - `f_01_rainbow_passage.txt`
  - `m_01`
    - `m_01_frog_where_are_you.wav`
    - `m_01_frog_where_are_you.txt`
    - `m_01_rainbow_passage.wav`
    - `m_01_rainbow_passage.txt`

Corpus format

- `phonetics_workshop/recordings/`
  - `f_01`
    - `f_01_frog_where_are_you.wav`
    - `f_01_frog_where_are_you.`~~`txt`~~`lab`
    - `f_01_rainbow_passage.wav`
    - `f_01_rainbow_passage.`~~`txt`~~`lab`
  - `m_01`
    - `m_01_frog_where_are_you.wav`
    - `m_01_frog_where_are_you.`~~`txt`~~`lab`
    - `m_01_rainbow_passage.wav`
    - `m_01_rainbow_passage.`~~`txt`~~`lab`

Corpus format

- `phonetics_workshop/recordings/`
  - `f_01`
    - `f_01_frog_where_are_you.wav`
    - `f_01_frog_where_are_you.lab`
    - `f_01_rainbow_passage.wav`
    - `f_01_rainbow_passage.lab`
  - `m_01`
    - `m_01_frog_where_are_you.wav`
    - `m_01_frog_where_are_you.lab`
    - `m_01_rainbow_passage.wav`
    - `m_01_rainbow_passage.lab`

Installation

- Instructions available at: https://montreal-forced-aligner.readthedocs.io/en/latest/installation.html

Installation

- Instructions available at: https://montreal-forced-aligner.readthedocs.io/en/latest/installation.html
- Forced-alignment and detailed corpus phonetics tutorial: https://eleanorchodroff.com/tutorial/index.html

Requirements

- Pronunciation dictionary and acoustic model

```
mfa model download acoustic english_us_arpa
mfa model download dictionary english_us_arpa
```

Pronunciation Dictionary

tomato	0.38	0.05	0.9	1.03	T AH0 M AA1 T OW2
tomato	0.99	0.14	1.31	0.92	T AH0 M EY1 T OW2

Phone segmentation with Forced Alignment

- Performing the forced alignment
  - `mfa align audio_directory dictionary_path model_path textgrid_directory`

```
(base) C:\Users\lucas>conda activate aligner3

(aligner3) C:\Users\lucas>mfa align --clean C:\Users\lucas\Documents\phonetics_workshop\long_recordings english_us_arpa
english_us_arpa C:\Users\lucas\Documents\PhD\projects\phonetics_workshop\long_recordings
```

Phone segmentation with Forced Alignment

- Performing the forced alignment
  - `mfa align audio_directory dictionary_path model_path textgrid_directory`

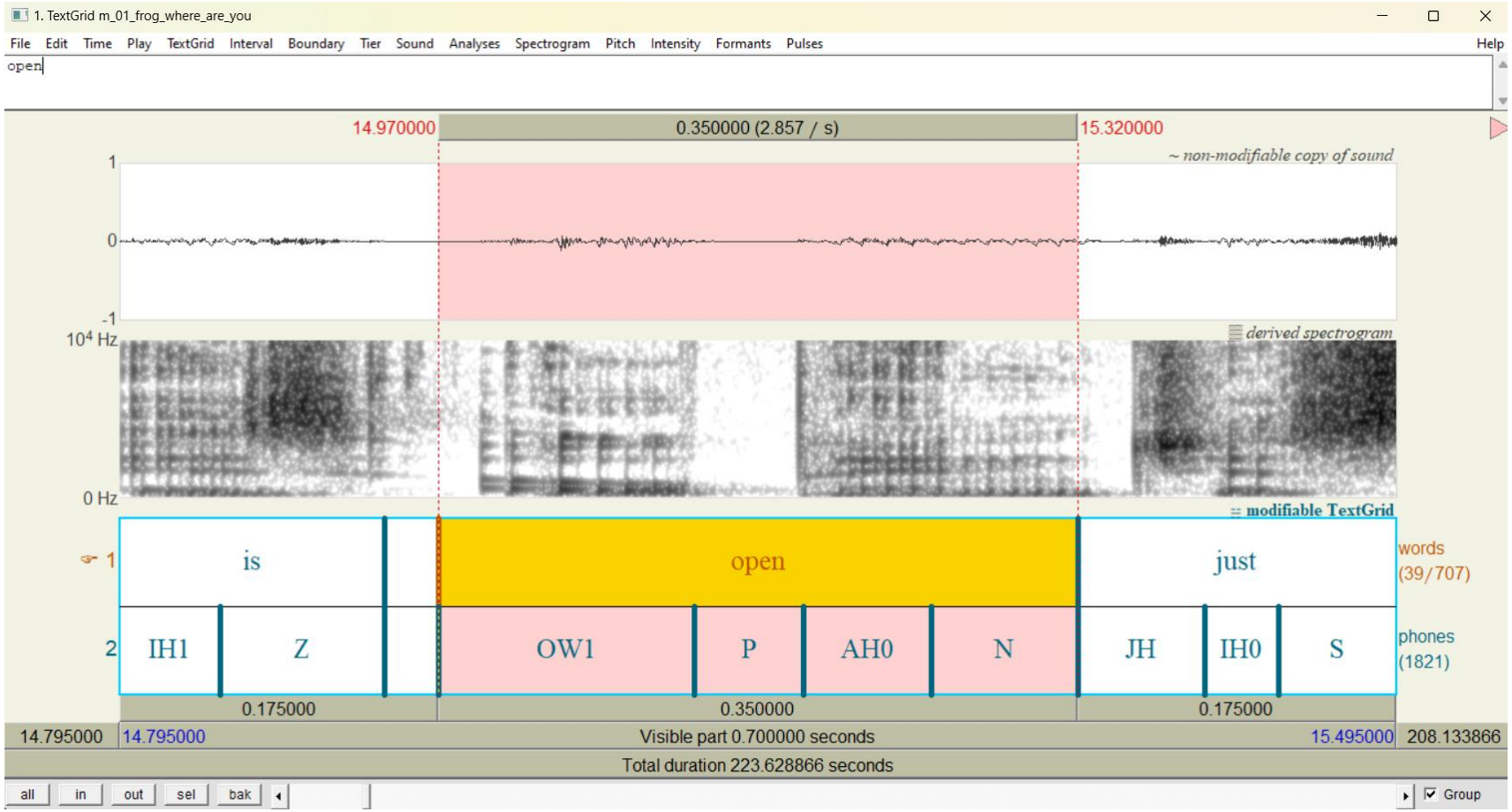
```
(base) C:\Users\lucas>conda activate aligner3

(aligner3) C:\Users\lucas>mfa align --clean C:\Users\lucas\Documents\phonetics_workshop\long_recordings english_us_arpa
english_us_arpa C:\Users\lucas\Documents\PhD\projects\phonetics_workshop\long_recordings
```

- Troubleshooting
  - `mfa align --clean audio_directory...`
  - Gets rid of files from previous runs

Output

- phonetics_workshop/recordings/
  - f_01
    - f_01_frog_where_are_you.wav
    - f_01_frog_where_are_you.lab
    - **f_01_frog_where_are_you.textgrid**
    - f_01_rainbow_passage.wav
    - f_01_rainbow_passage.lab
    - **f_01_rainbow_passage.textgrid**
  - m_01
    - m_01_frog_where_are_you.wav
    - m_01_frog_where_are_you.lab
    - **m_01_frog_where_are_you.textgrid**
    - m_01_rainbow_passage.wav
    - m_01_rainbow_passage.lab
    - **m_01_rainbow_passage.textgrid**

Output


1. TextGrid m_01_frog_where_are_you
File Edit Time Play TextGrid Interval Boundary Tier Sound Analyses Spectrogram Pitch Intensity Formants Pulses Help
open
14.970000
0.350000 (2.857 / s)
15.320000
~ non-modifiable copy of sound
derived spectrogram
modifiable TextGrid
is
open
just
words (39/707)
IH1
Z
OW1
P
AH0
N
JH
IH0
S
phones (1821)
0.175000
0.350000
0.175000
14.795000
14.795000
Visible part 0.700000 seconds
15.495000
208.133866
Total duration 223.628866 seconds
all in out sel bak
Group

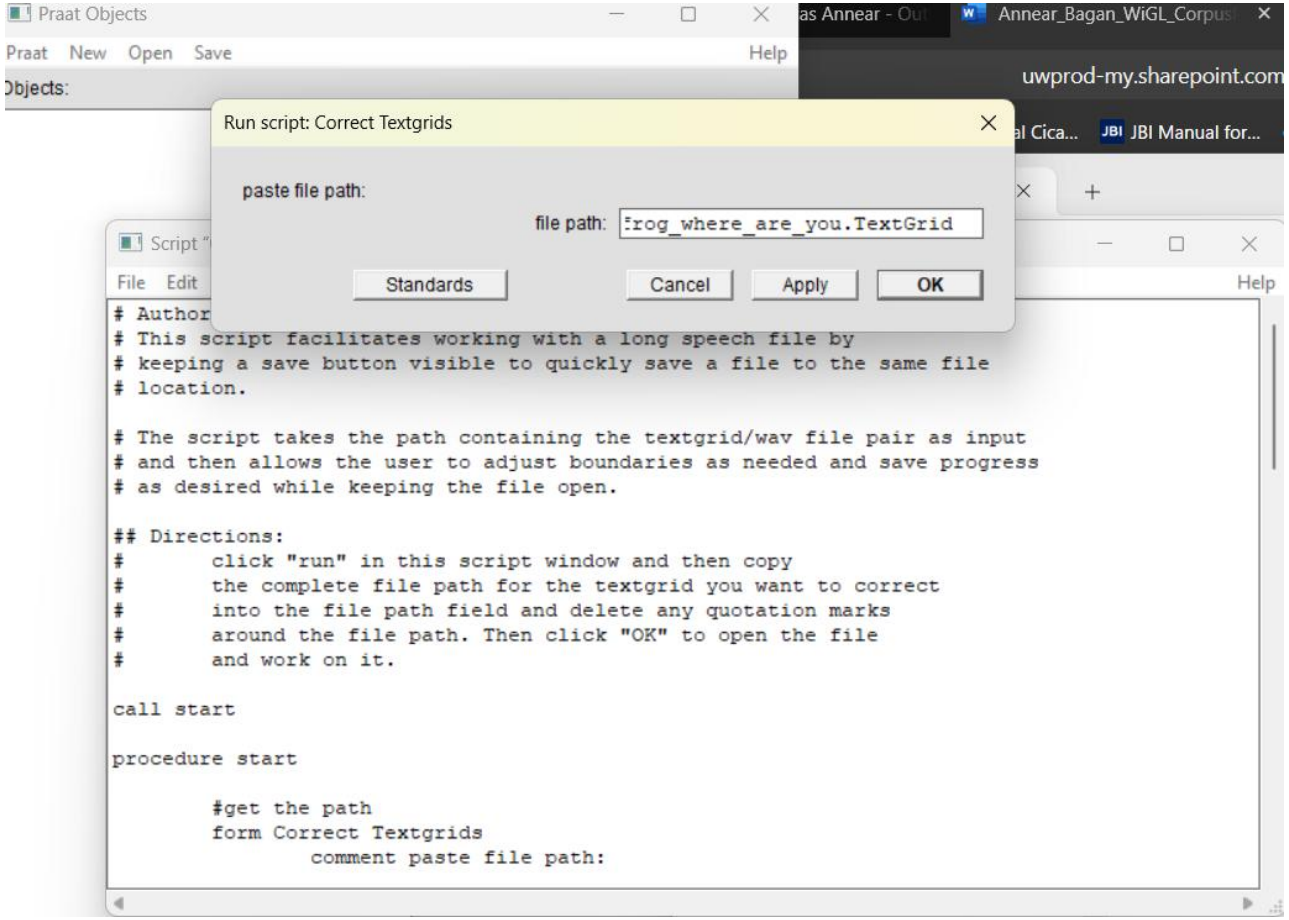
Modification

- `correct_textgrids_long_file.praat`
  - Use this script to save your work periodically as you work on a long file

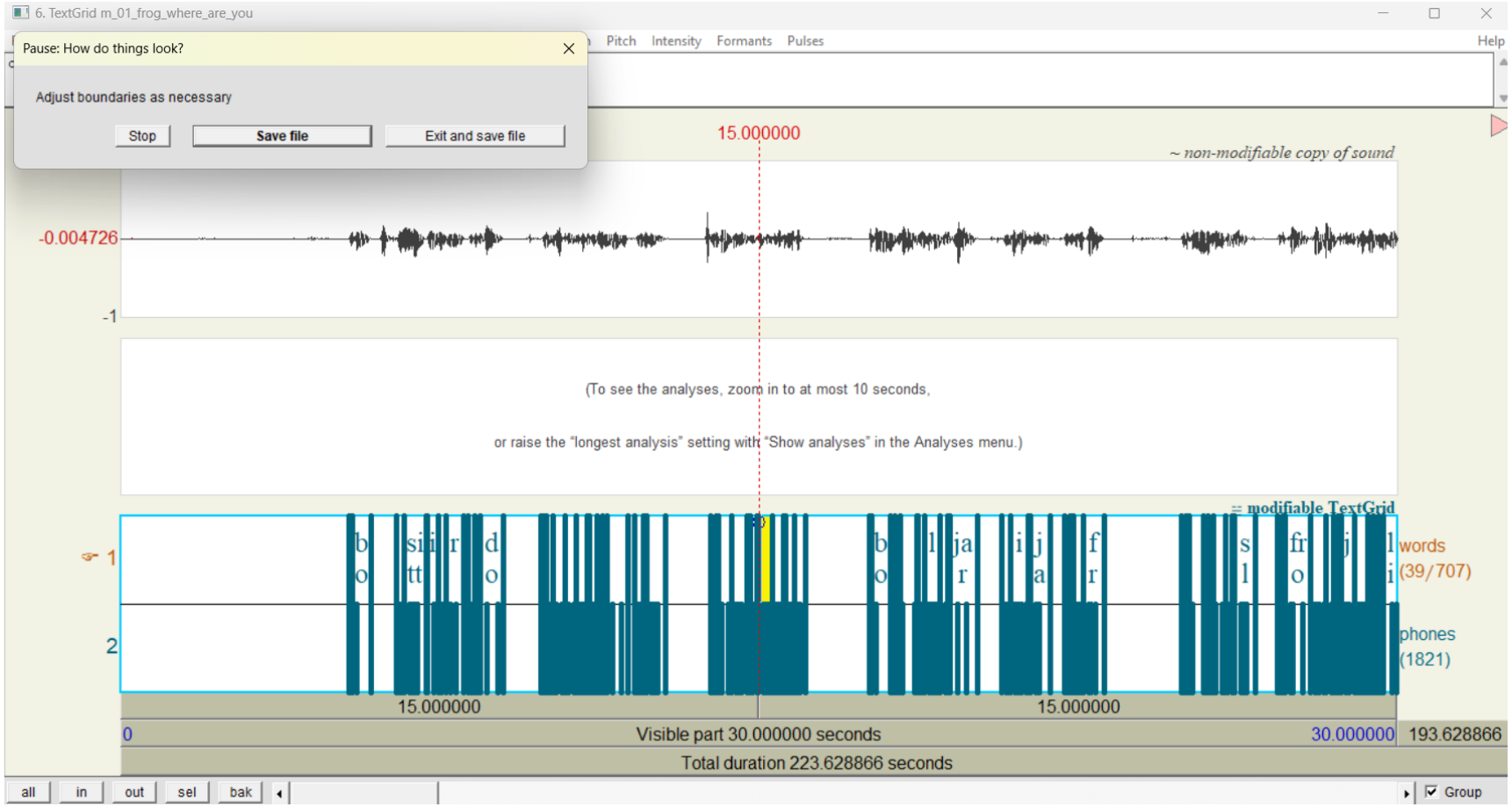
Modification

- `correct_textgrids_long_file.praat`
  - Use this script to save your work periodically as you work on a long file
- `correct_textgrids_directory.praat`
  - Use this script to work through many shorter audio files and textgrids
  - Saves your progress in the list and allows you to resume where you left off.

Modification



Modification



Check-in

- Questions on Forced-alignment

Querying

- How can you ask questions about what you have so far
- R Notebook:
  - `annear_bagan_phonetics_workshop.Rmd`

Demonstrate in Notebook

annear_bagan_phonetics_workshop.Rmd | data

Knit on Save | Knit | Run | Publish

Source | Visual | Normal | Format | Insert | Table | Outlin

Read data from textgrids

This stage uses the `readtextgrid` package from TJ Mahr. For documentation see:
https://github.com/tjmahr/readtextgrid

```
{r Load data from network}
# define the primary directory for textgrids
data_dir <- "C:/Users/lucas/Documents/PhD/projects/phonetics_workshop
/long_recordings"

# create a dataframe called "data" that will contain information from the textgrids
data <- file.path(data_dir) |>
  # Get the speaker folders
  fs::dir_ls(type = "dir") |>
  # Look through each folder for textgrids
  fs::dir_ls(regexp = "/*.TextGrid") |>
  # apply read_textgrid() to each TextGrid file
  future_lapply(read_textgrid) |>
  bind_rows() |>
  # create new columns based on info in filenames
  mutate(speaker = substr(file, 1, 4),
         task = str_sub(basename(file), 6),
         sex = substr(file, 1, 1)) |>
  # pivot the data frame so that each sound is structured as part of a word
```

Learning outcomes

- Creating
  - Using Whisper for Transcription
  - Montreal Forced Aligner for phonetic segmentation in Praat

Learning outcomes

- Creating
  - Using *Whisper* for Transcription
  - *Montreal Forced Aligner* for phonetic segmentation in Praat
- Managing
  - File naming and organization

Learning outcomes

- Creating
  - Using *Whisper* for Transcription
  - *Montreal Forced Aligner* for phonetic segmentation in Praat
- Managing
  - File naming and organization
- Modifying
  - Using Praat scripts to reduce saving and naming errors
  - Hand-correcting force-aligned TextGrids

Learning outcomes

- Creating
  - Using *Whisper* for Transcription
  - *Montreal Forced Aligner* for phonetic segmentation in Praat
- Managing
  - File naming and organization
- Modifying
  - Using Praat scripts to reduce saving and naming errors
  - Hand-correcting force-aligned TextGrids
- Querying
  - R for inspecting and validating data.

Final check-in/questions

Thank you!

- lucas.annear@wisc.edu
- ebagan@wisc.edu

Links to resources

- Workshop repository: https://github.com/lucasjannear/phonetics_workshop
- Whisper: https://github.com/openai/whisper
- Montreal Forced Aligner: https://montreal-forced-aligner.readthedocs.io/en/latest/
- Power Toys: https://learn.microsoft.com/en-us/windows/powertoys/install
- Praat: https://www.fon.hum.uva.nl/praat/
- readtextgrid package: https://github.com/tjmahr/readtextgrid
- Eleanor Chodroff corpus phonetics tutorial: https://eleanorchodroff.com/tutorial/index.html