Lecture 21: Speech Data, Forced Alignment

LING 1340/2340: Data Science for Linguists

Na-Rae Han

Objectives

- Speech data
  - Speech corpora, datasets: TIMIT
  - PRAAT
  - Command-line conversion
- Popular speech data analysis tools
  - Forced aligners

TextGrid

- Praat was able to parse TIMIT's PHN file format (phone tier)
- Saving it out to a proper TextGrid file →
- However, Praat couldn't handle:
  - SA1.TXT (utterance tier)
  - SA1.WRD (word tier)

← How to get them into TextGrid?

There's a python library (or two) for that!

praat-textgrids 1.3.1

```
pip install praat-textgrids
```

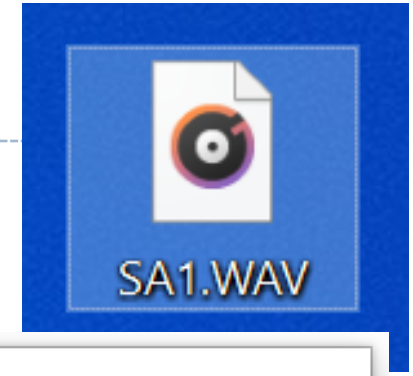
['pɹɑ.səl,maʊθ]

Parselmouth – Praat in Python, the Pythonic way

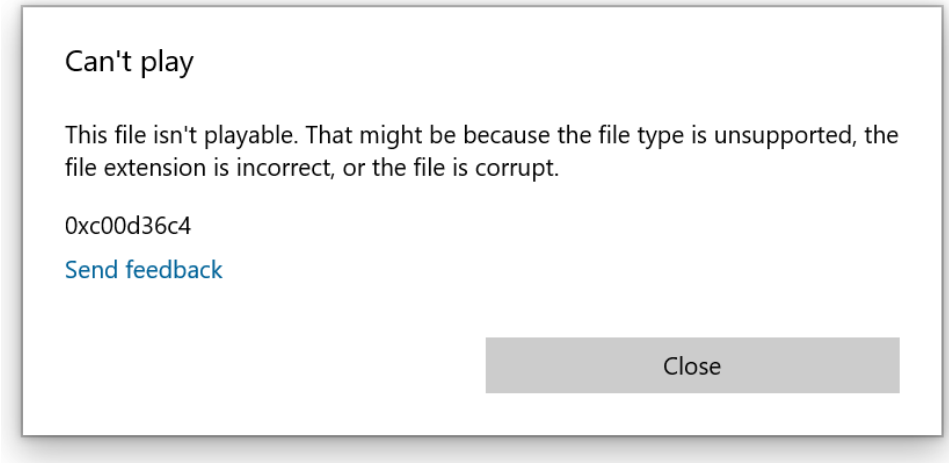
```
File type = "ooTextFile"
Object class = "TextGrid"

xmin = 0
xmax = 2.92
tiers? <exists>
size = 2
item []:
    item [1]:
        class = "IntervalTier"
        name = "phn"
        xmin = 0
        xmax = 2.92
        intervals: size = 37
        intervals [1]:
            xmin = 0
            xmax = 0.19062500000000002
            text = "h#"
        intervals [2]:
            xmin = 0.19062500000000002
            xmax = 0.2849375
            text = "sh"
        intervals [3]:
            xmin = 0.2849375
            xmax = 0.3576875
            text = "ix"
        intervals [4]:
            xmin = 0.3576875
            xmax = 0.415125
            text = "hv"
        intervals [5]:
            xmin = 0.415125
            xmax = 0.54825
            text = "eh"
        intervals [6]:
```

.WAV format?



- Also, even though PRAAT was able to open the .WAV files, Windows 10 cannot…



- These files are not really .WAV…
  - **SPHERE format**, normally with .SPH extension.
- How to convert to WAV?

Solution 1: Praat script

- Write a praat script
  - (Or, grab someone else's…)

```
# prep_audio_mfa.praat
# Written by E. Chodroff
# Oct 23 2018
# extract left channel and resample to 16 kHz for all wav files in a directory

### CHANGE ME!
# don't forget the slash at the end of the path
dir$ = "/Users/Eleanor/Desktop/align_input/"
###

Create Strings as file list: "files", dir$ + "*.wav"
nFiles = Get number of strings

for i from 1 to nFiles
          # read in WAV file
          selectObject: "Strings files"
          filename$ = Get string: i
          Read from file: dir$ + filename$

          # extract left channel
          Extract one channel: 1

          # resample to 16kHz with 50 point precision (default)
          Resample: 16000, 50

          # save WAV file
          Save as WAV file: dir$ + filename$

          # clean up
          select all
          minusObject: "Strings files"
          Remove
endfor
```

Solution 2: SoX + bash shell

- SoX: Sound eXchage
  - https://sourceforge.net/projects/sox/

```
sox <input-file> -b 16 -t wav <output-file>
```

```
narae@T480s MINGW64 ~/Desktop/FCJF0
$ alias sox="/d/util/sox-14.4.2/sox.exe"

narae@T480s MINGW64 ~/Desktop/FCJF0
$ ls
SA1.PHN  SA2.WAV     SI1657.PHN  SI648.WAV  SX217.PHN  SX307.WAV  SX397.PHN
SA1.TXT  SA2.WRD     SI1657.TXT  SI648.WRD  SX217.TXT  SX307.WRD  SX397.TXT
SA1.WAV  SI1027.PHN  SI1657.WAV  SX127.PHN  SX217.WAV  SX37.PHN   SX397.WAV
SA1.WRD  SI1027.TXT  SI1657.WRD  SX127.TXT  SX217.WRD  SX37.TXT   SX397.WRD
SA2.PHN  SI1027.WAV  SI648.PHN   SX127.WAV  SX307.PHN  SX37.WAV   true_wav/
SA2.TXT  SI1027.WRD  SI648.TXT   SX127.WRD  SX307.TXT  SX37.WRD

narae@T480s MINGW64 ~/Desktop/FCJF0
$ sox SA1.WAV -b 16 -t wav true_wav/SA1.wav

narae@T480s MINGW64 ~/Desktop/FCJF0
$ ls true_wav/
SA1.wav
```

converting a single file

Solution 2: SoX + bash shell

```
narae@T480s MINGW64 ~/Desktop/FCJF0
$ for x in *WAV
> do
> sox $x -b 16 -t wav true_wav/$x
> echo $x finished
> done
SA1.WAV finished
SA2.WAV finished
SI1027.WAV finished
SI1657.WAV finished
SI648.WAV finished
SX127.WAV finished
SX217.WAV finished
SX307.WAV finished
SX37.WAV finished
SX397.WAV finished

narae@T480s MINGW64 ~/Desktop/FCJF0
$ ls true_wav/
SA1.wav   SI1027.WAV  SI648.WAV  SX217.WAV  SX37.WAV
SA2.WAV   SI1657.WAV  SX127.WAV  SX307.WAV  SX397.WAV
```

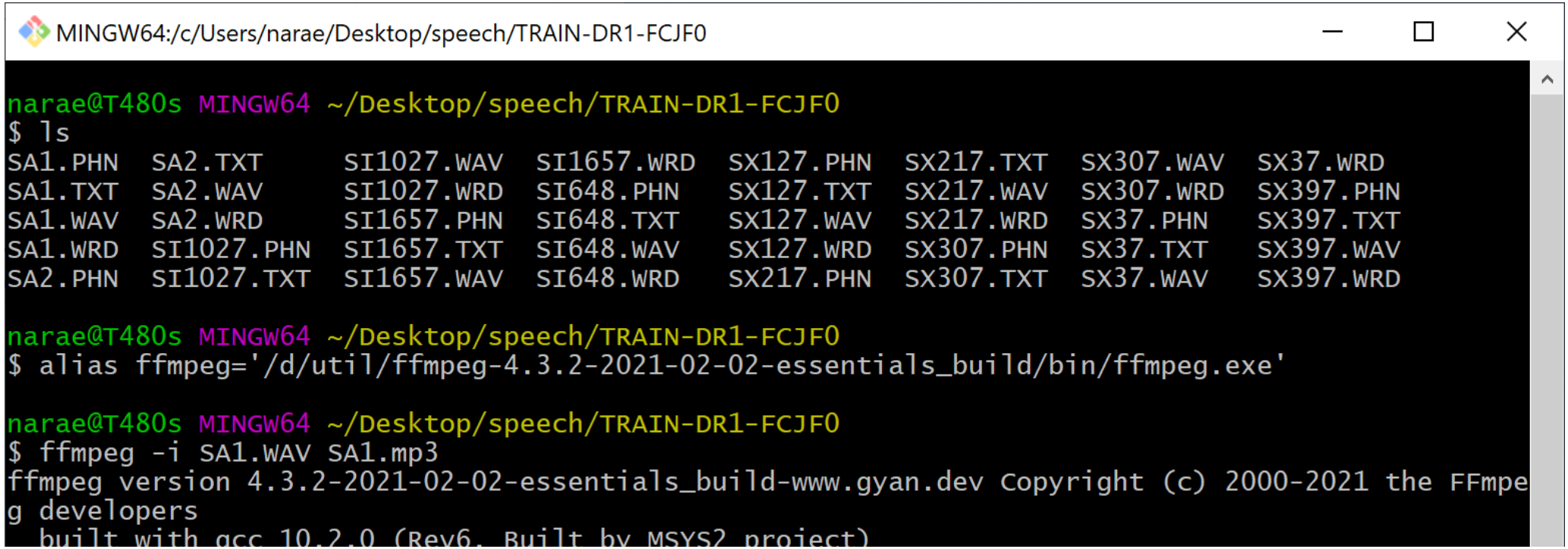
for loop in bash!

```
for x in *.WAV
do
sox $x -b 16 -t wav true_wav/$x
echo $x finished
done
```

Declared as `x`, subsequent references as `$x`

Command-line conversion to mp3: with ffmpeg

- `ffmpeg -i input.wav output.mp3`



Again, how to do this with *every* wav file…?

General-purpose audio/video manipulation software

- Audacity
  - Open-source audio software

- SoX
  - Sound eXchange; audio format conversion tool

Powerful command-line tools!!

- FFmpeg
  - For recording and converting audio/video data

https://musicinformationretrieval.com/sox_and_ffmpeg.html

Popular speech data analysis tools for linguists (1)

Multimodal (audio + video):

- ELAN multimodal annotator (Wittenberg et al. 2006)

Some tools are online:

- NORM: the Vowel Normalization and Plotting Suite
- DARLA: Dartmouth Linguistic Automation
  - ← You upload an audio file and a transcript file, the site will process them and email you the results, etc!

Popular speech data analysis tools for linguists (2)

- **Forced alignment** is a technique to take an orthographic transcription of an audio file and generate a time-aligned version using a pronunciation dictionary to look up phones for words.
- Forced aligners
  - Penn Phonetics Lab Forced Aligner (Yuan & Liberman 2009) → legacy, became FAVE-align
  - FAVE-align (Rosenfelder et al. 2011)
  - Montreal Forced Aligner (McAuliffe et al. 2017) ← we'll take a look

Wrapping up

- Work on your project!
- Project presentations dates/presenters fixed… check yours!