

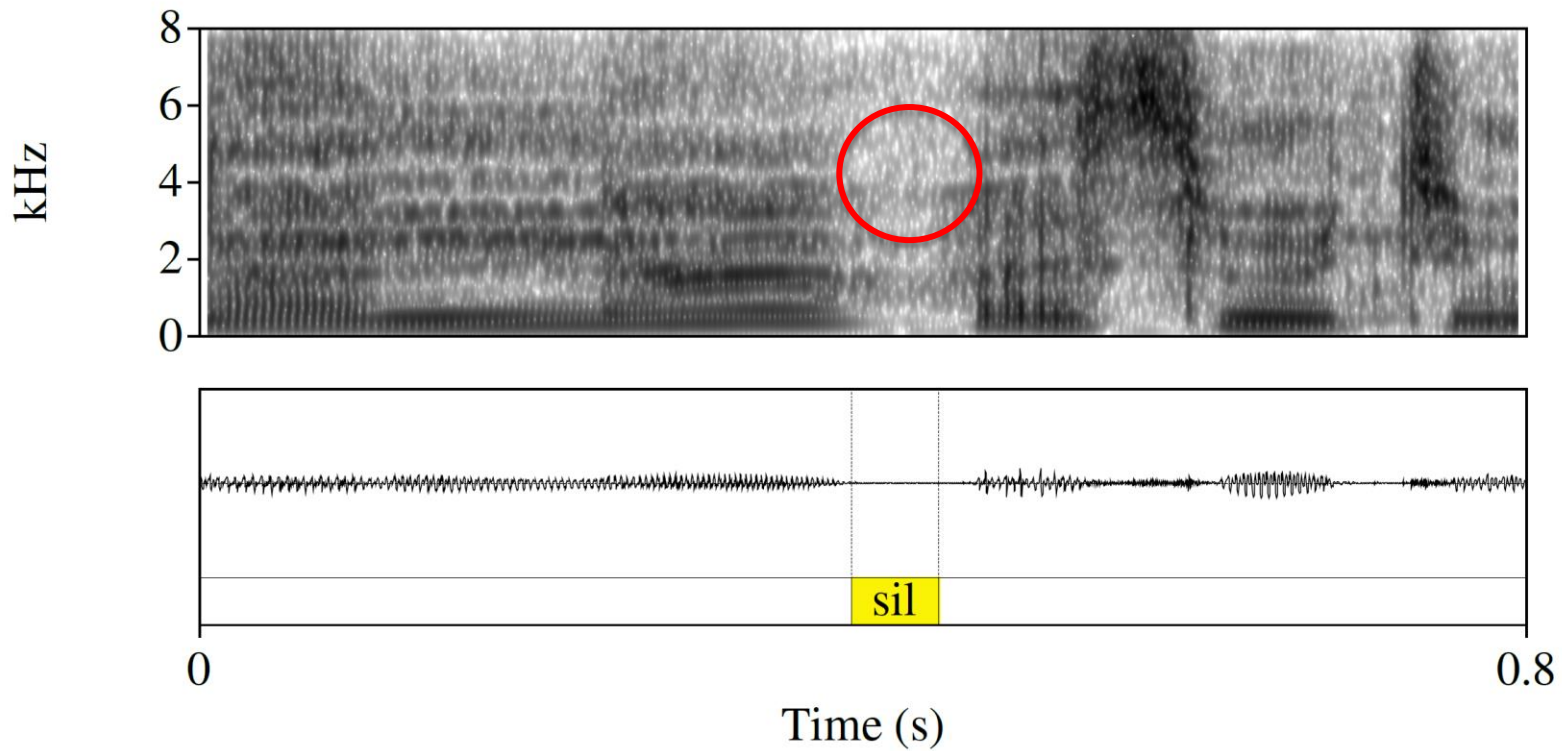
Corpus generation for research in pause-internal phonetic particles (PINTS)

Mikey Elmers

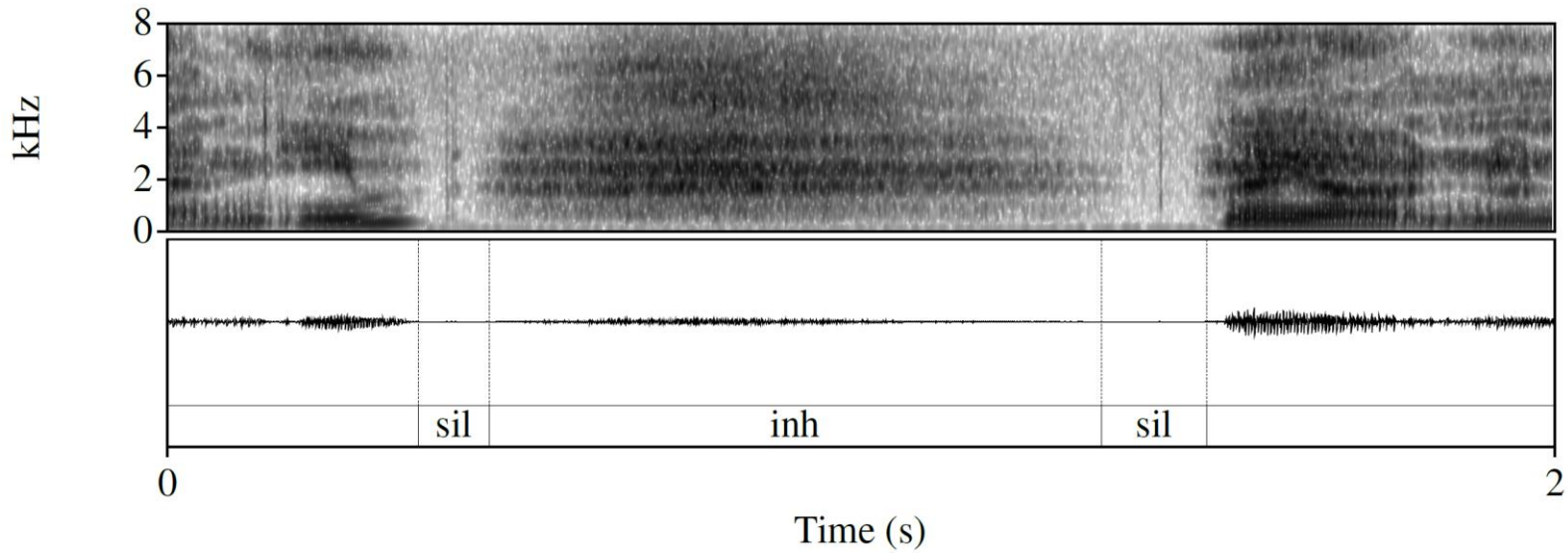
Phonetic Particles

- Investigating phonetic particles in speech pauses
 - Silent pauses
 - Breath noises (BrN)
 - Filler particles
 - Clicks
 - Laughter

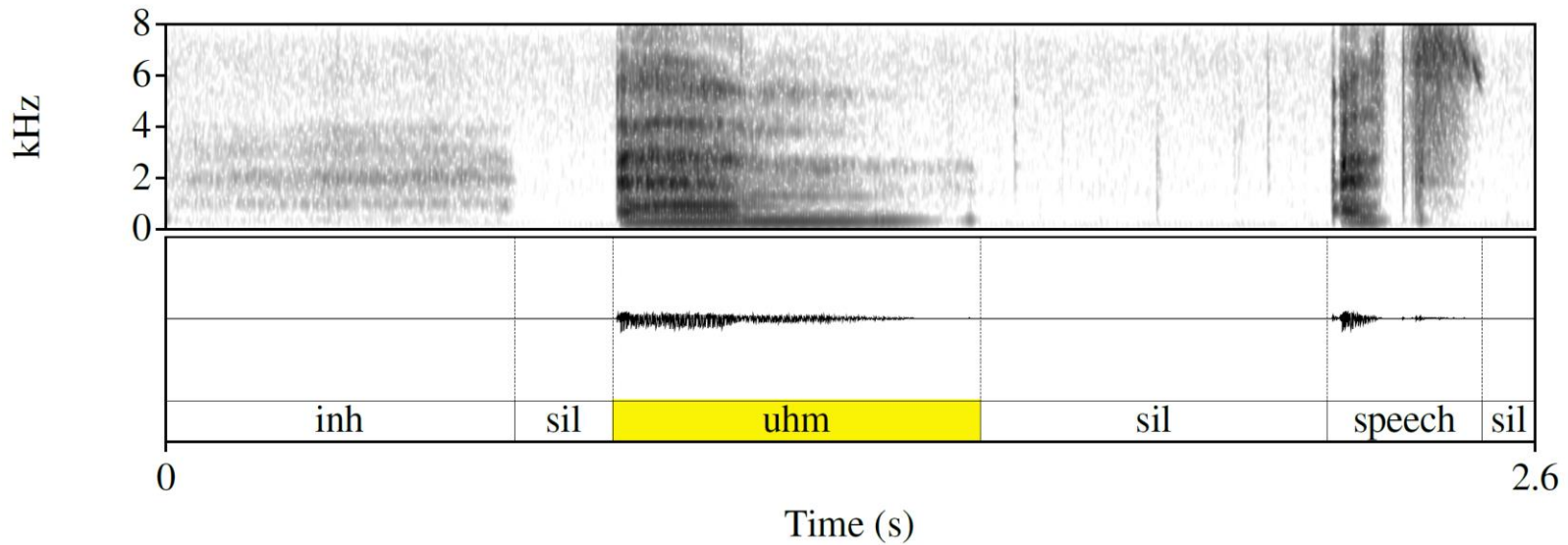
Silent Pauses



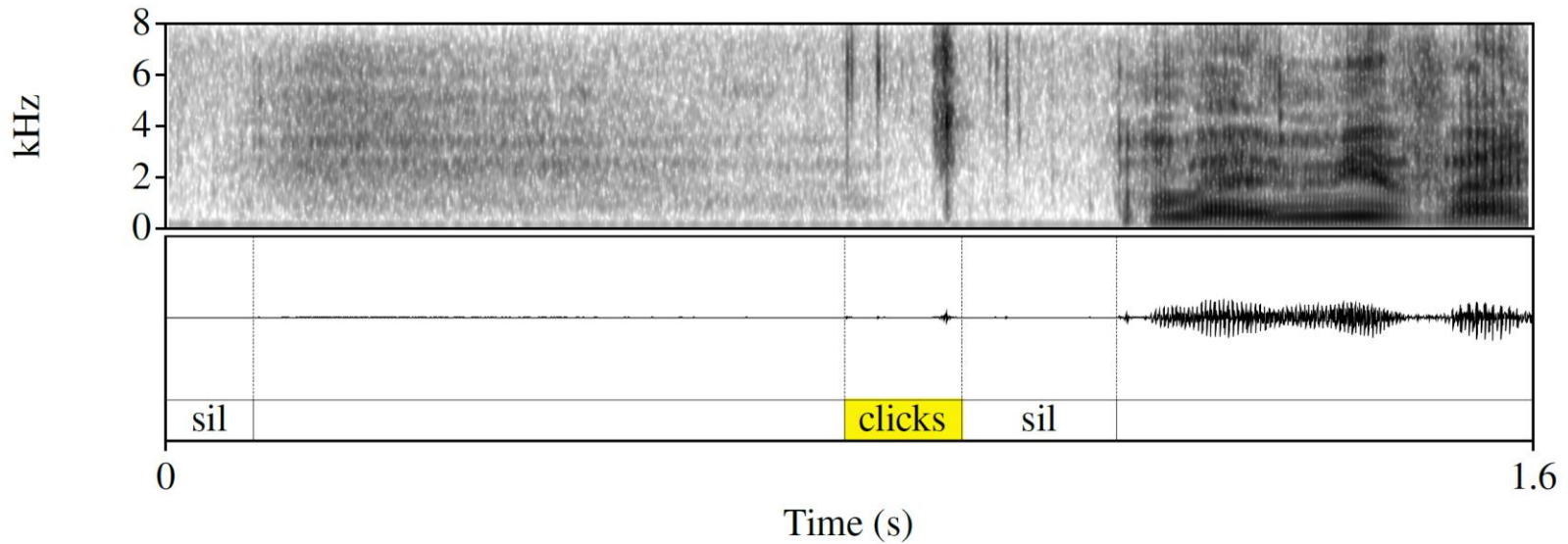
Breath Noises



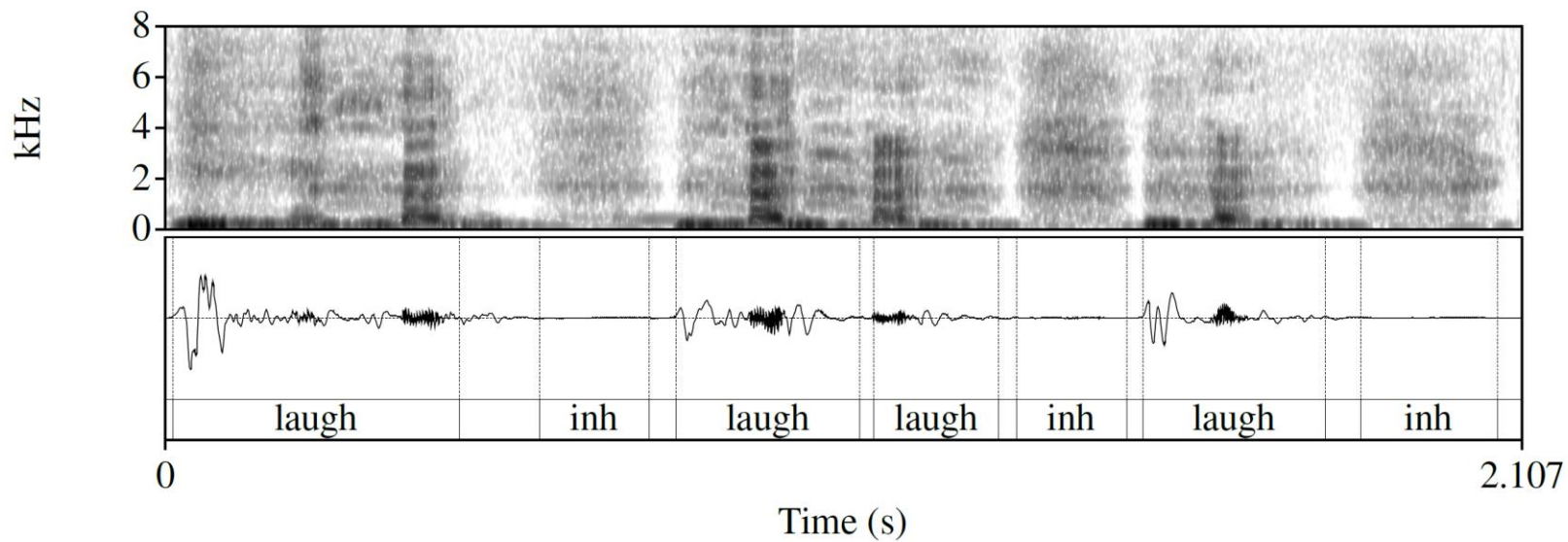
Filler Particles



Clicks



Laughter



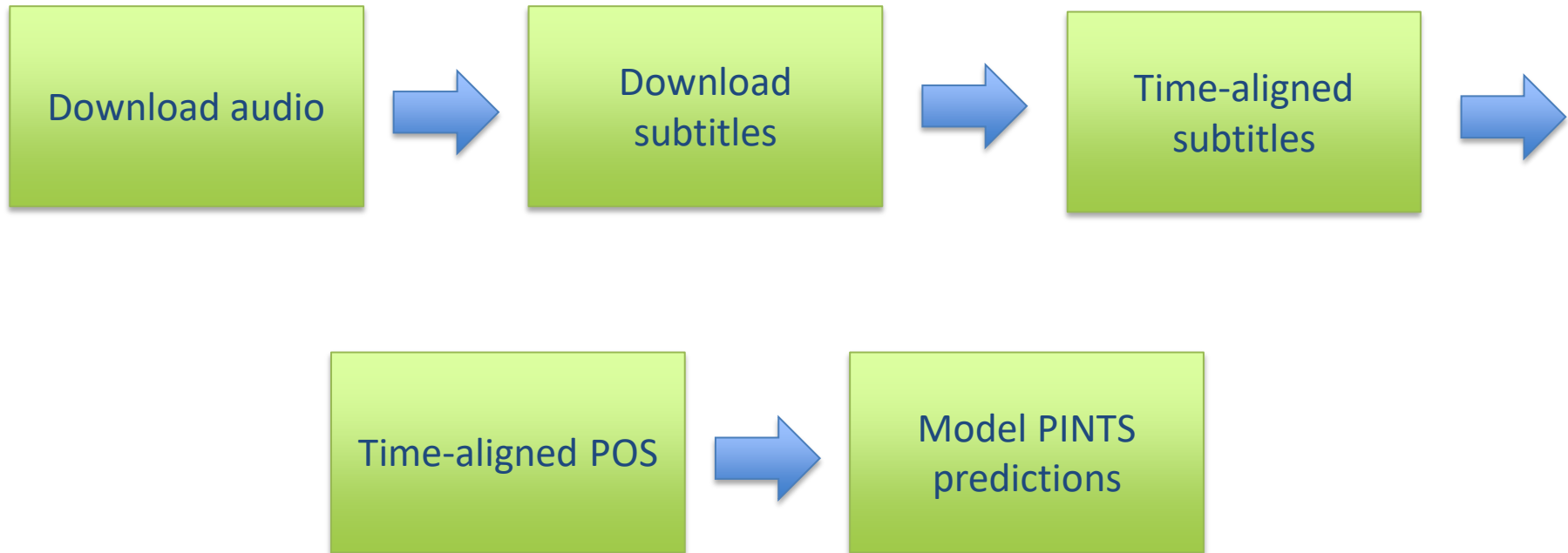
PINTS TTS

- Improve TTS via the modeling and inclusion of PINTS
 - Location
 - Duration
 - Frequency
 - Variety
- In order to train the TTS system we need a large amount of (ideally) annotated training data

Purpose

- Primarily focused on:
 - Creating automated pipeline
 - Scrape audio data from YouTube
 - Annotations
 - Use annotated sound files with PINTS information for machine learning (ML) models
- Create a corpus that satisfies the following:
 - Consistent
 - High quality

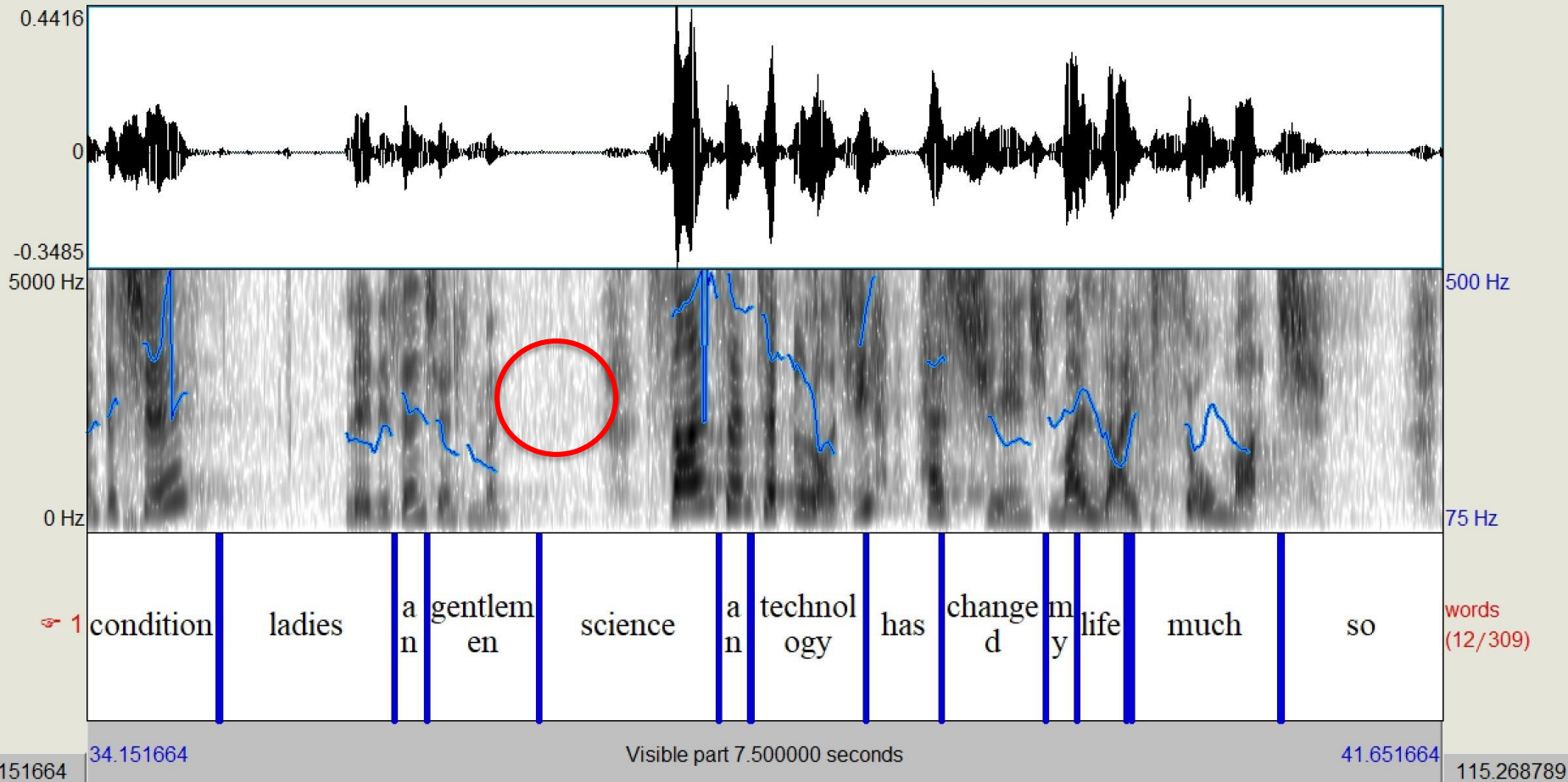
Pipeline



Pipeline Example: Words

- Impromptu speech audio
 - Single speaker
 - Semi-spontaneous situation
 - ~ 2 minutes
- Subtitles from YouTube
 - Subtitle file format is video text tracks (VTT)
 - Convert VTT subtitles -> plain text file (TXT) format
 - Generate a Praat TextGrid

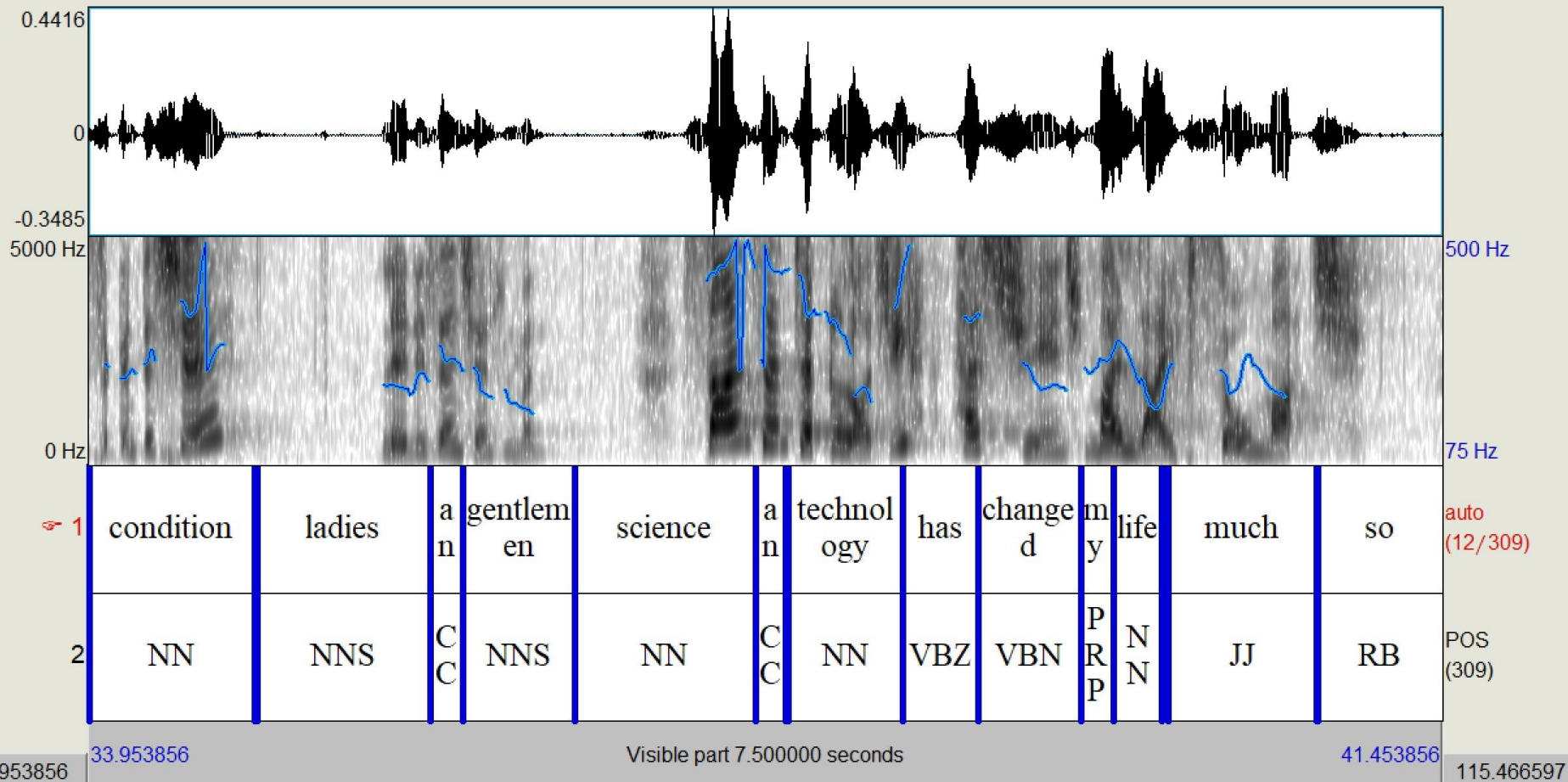
Pipeline Example: Words



Pipeline Example: POS

- Generating time-aligned parts of speech (POS)
 - Gather list of words used from previous module in pipeline
 - Use natural language toolkit (NLTK) to generate POS
 - Convert to TXT format
 - Generate a Praat TextGrid

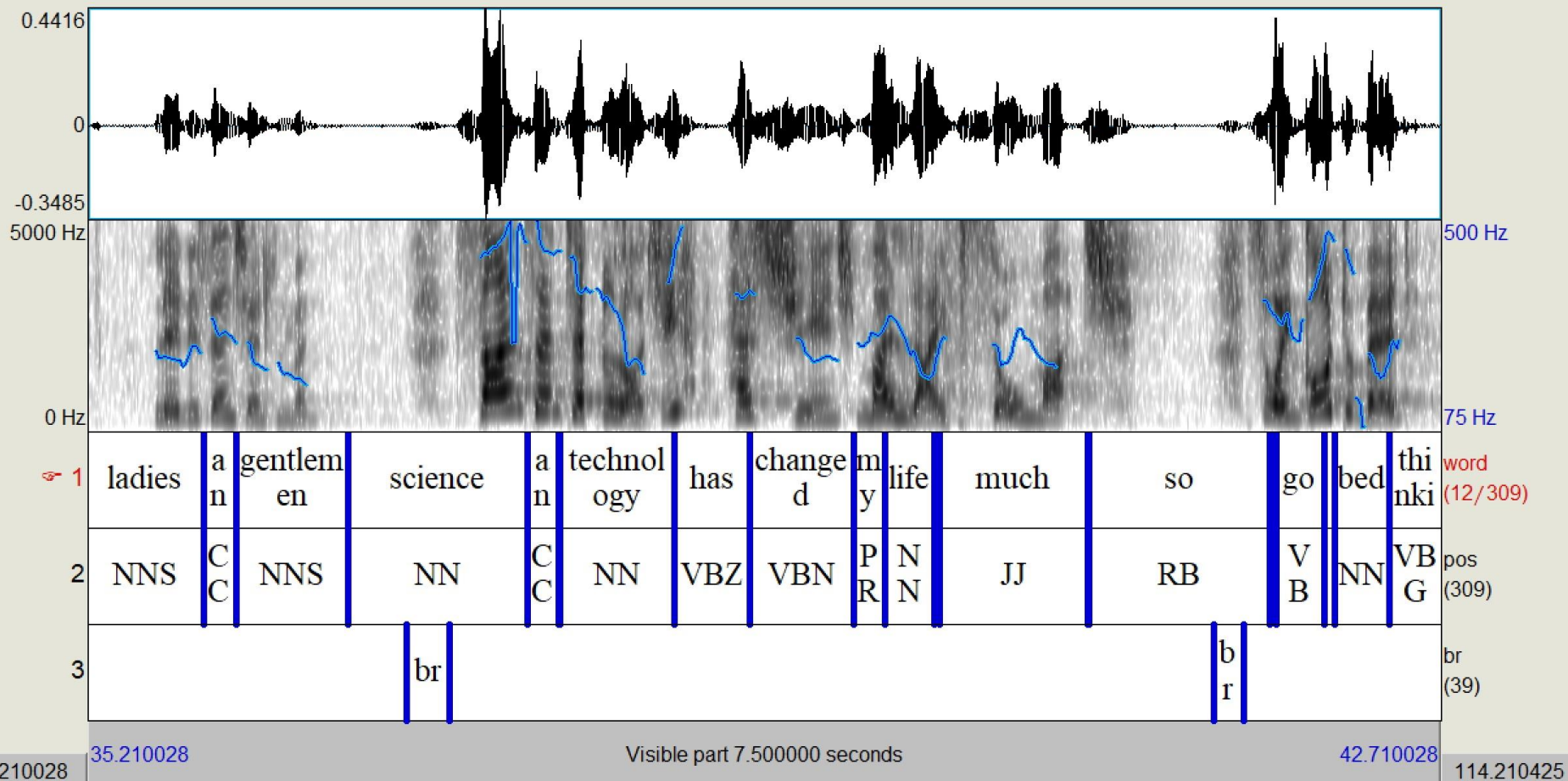
Pipeline Example: POS



Pipeline Example: Manual BrN

- Breath noise annotation
 - Normally not part of the pipeline
 - Hand-labeled
 - Used for comparing against model predictions

Pipeline Example: Manual BrN



Pipeline Example: Predicted BrN

- Corpus for training model
 - Buckeye corpus
 - Annotated for BrN
 - ~ 10 minutes
- Model information
 - Input: mel frequency cepstral coefficients (MFCC)
 - Target: BrN or non-BrN
 - Neural network (NN) for binary classification

Pipeline Example: Predicted BrN

- Modeling breath noise predictions
 - Convert time to frames
 - MFCCs as input

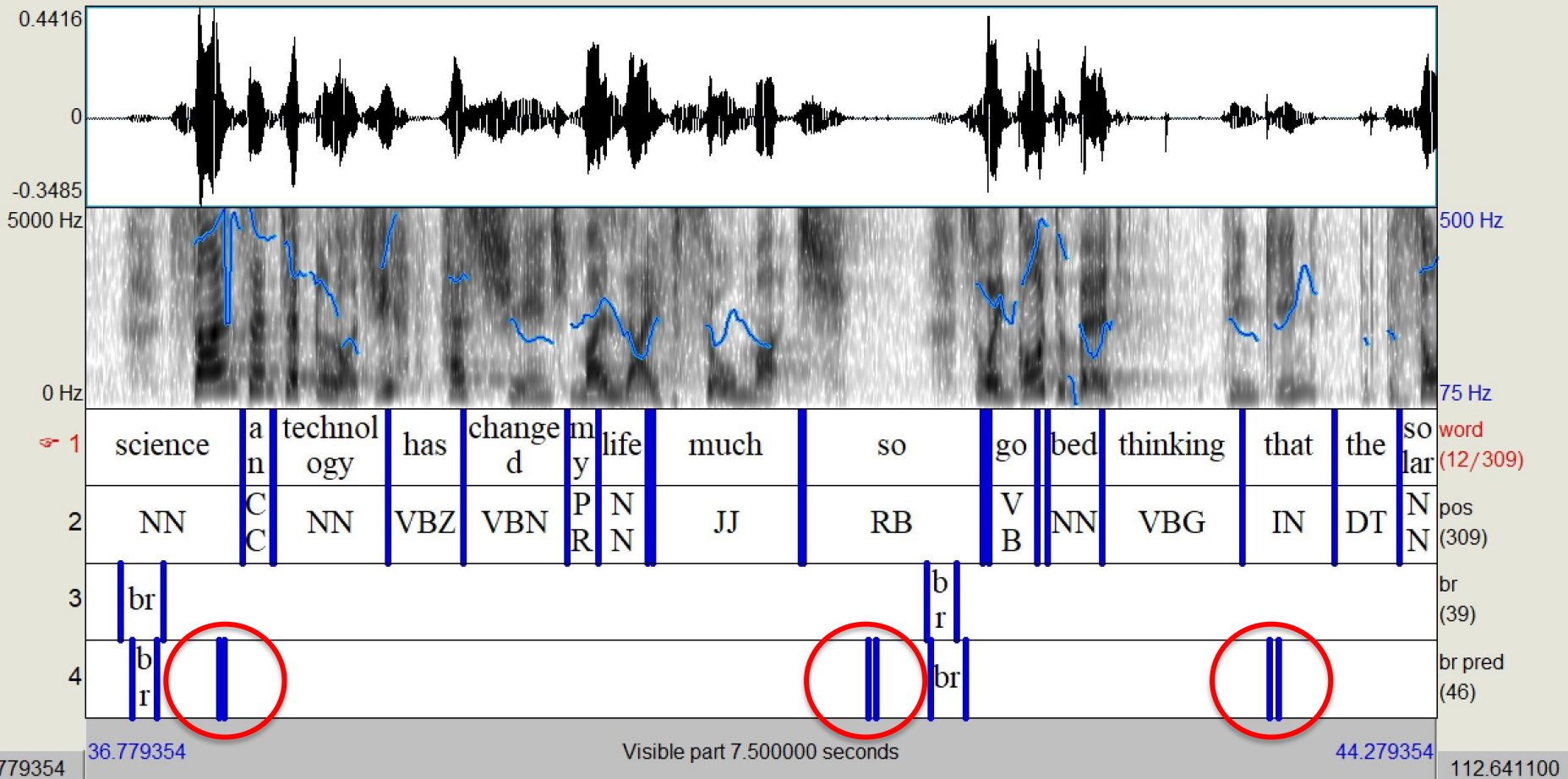
```
"mfcc": [  
  [  
    -582.6719360351562,  
    0.8451012372970581,  
    0.8422243595123291,  
    0.8374413847923279,  
    0.8307602405548096,  
    0.8222096562385559,  
    0.8118045330047607,  
    0.799580454826355,  
    0.7855702638626099,  
    0.7698125839233398,  
    0.7523539066314697,  
    0.7332457900047302,  
    0.7125375270843506  
  ],  
  [  
    -582.8423461914062,  
    0.6043773293495178,  
    0.6023807525634766,  
    0.5990595817565918,  
    0.5944247841835022,  
    0.5884836912155151,  
    0.5812586545944214,  
    0.572767972946167,  
    0.5630406141281128,  
    0.5520878434181213,  
    0.5399560928344727,  
    0.5266737937927246,  
    0.5122793316841125  
  ],  
],
```


Pipeline Example: Predicted BrN

- Modeling breath noise predictions
 - Frame length -> Time duration
 - Convert to TXT format
 - Create Praat TextGrid

0.02
28.77
28.79 br
34.55
34.57 br
37.04
37.18 br
37.52
37.55 br
41.12
41.17 br
41.47
41.66 br
43.35
43.4 br
45.21
45.28 br
51.13
51.15 br
52.87
53.01 br
57.98
58.03 br
68.92
69.03 br
75.02
75.05 br
76.09
76.12 br
79.11
79.13 br

Pipeline Example: Predicted BrN



Legal Issues

- Audio Scraping
 - Allowed to scrape audio from YouTube but not allowed to distribute it
 - Audio content must be monitored for personal information
- Distribution
 - Limited distribution for scientific research only

Conclusions

- Pipeline
 - Able to download audio from YouTube
 - Generate time-aligned words and POS
 - Label breath noises from model predictions
- Scale
 - Limited distribution for scientific research only

Reference

- Bird, Steven, Edward Loper and Ewan Klein (2009), *Natural Language Processing with Python*. O'Reilly Media Inc.
- Boersma, Paul & Weenink, David (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.53, retrieved 8 September 2021 from <http://www.praat.org/>
- Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. (2007) Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).

PINTS Website

Thank you!

<http://pauseparticles.org/>