NLP-powered auditory data processing

at Winter Storm 2022

NLP resource for production data processing

Production data requires some work to use

- transcription

- onset detection

- coding

- segmentation …

=> Use off-the-shelf NLP tools minimize the trouble

Goal of this session

• The pipeline is specialized for online speeded cloze data

The purpose of the session is to:
• introduce the pipeline
• you know enough to modify it to suit your own data

Speeded cloze data

People

went

to

the

beach
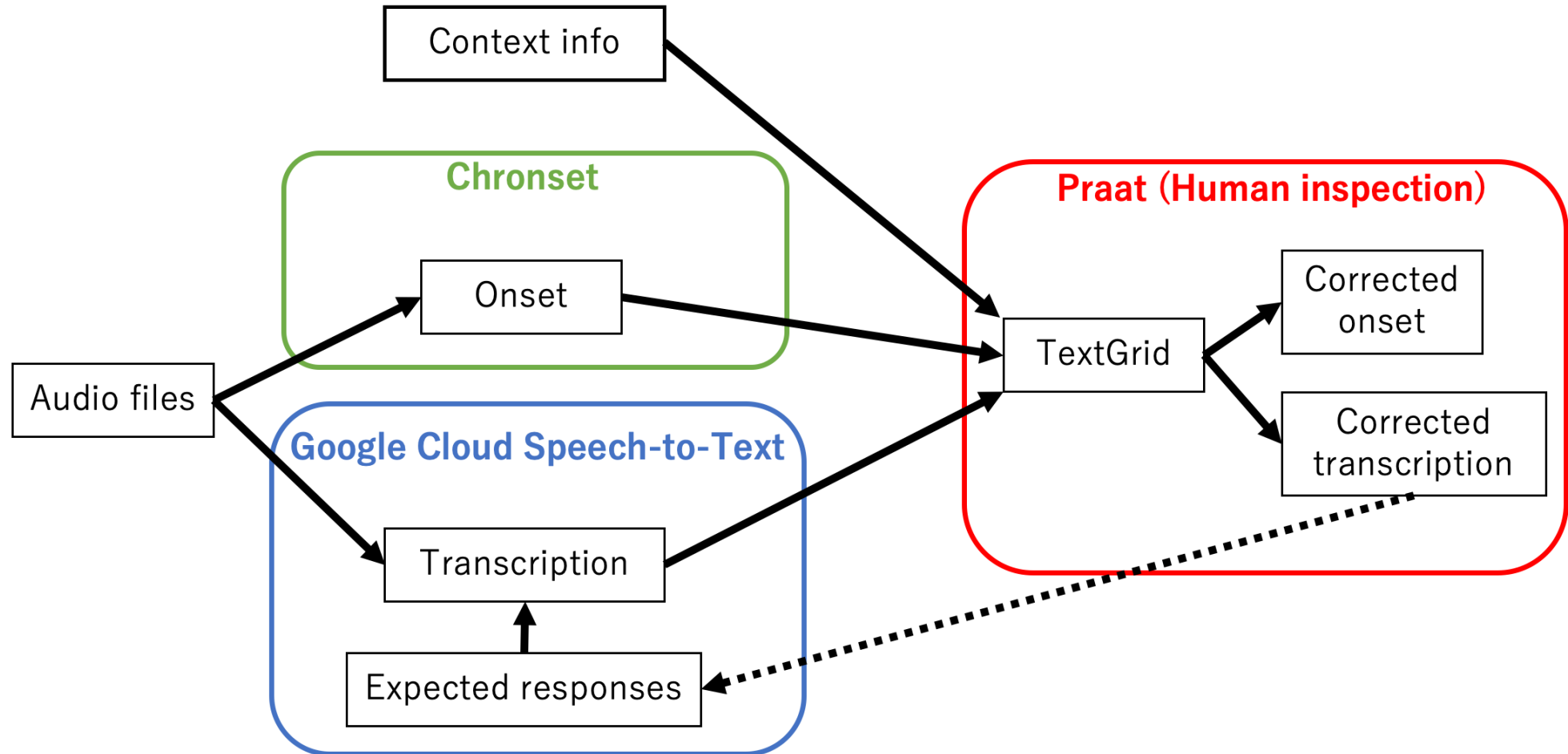
to

Speeded cloze data

The task:

- Partial sentences are presented

- A continuation is produced as quickly as possible

- Dependent measures are **the responses** and **its latency**

The responses are:

- Mostly single-word

- Open-ended, but constrained

- Collected online

Overview of the pipeline

Google Cloud Speech-to-Text

(https://cloud.google.com/speech-to-text)

- Neural network model for speech recognition

- Pretty accurate (82.29%)

- Can be improved by simple scripts

- Requires Google Cloud Platform

Google Cloud Speech-to-Text

- Paid, but pretty cheap (w/ 90-day trial, a free hour for each month)
- One experiment costed $6.34

Feature	Standard models (all models except enhanced video and phone call)		Enhanced models (video, phone call)	
	0-60 Minutes	Over 60 Mins up to 1 Million Mins	0-60 Minutes	Over 60 Mins up to 1 Million Mins
Speech Recognition (without Data Logging - default)	Free	$0.006 / 15 seconds **	Free	$0.009 / 15 seconds **
Speech Recognition (with Data Logging opt-in)	Free	$0.004 / 15 seconds **	Free	$0.006 / 15 seconds **

*** Each request is rounded up to the nearest increment of 15 seconds.*

Google Cloud Speech-to-Text

- Has model adaptation feature that makes the recognizer sensitive to specific vocabulary

- Has models for many languages

- Can also be accessed through command line, Google Cloud UI console, etc

Other options

Google Speech Recognizer

- An old version of Google Cloud Speech-to-Text

- Free

- Less accurate than the latest one

- Probably because it does not have the keyword function

Chronset

An online onset detector run by BCBL trained to detect human voice

(https://www.bcbl.eu/databases/chronset)

CHRONSET

An Automated Tool for Detecting Speech Onset

The Chronset website allows users to upload a set of .wav files and receive a list of onset latency estimates via e-mail.

A full description of the Chronset algorithm is available in *Roux, F., Armstrong, B. C., & Carreiras, M. (2016).* **Chronset: an automated tool for detecting speech onset**. *Behavior Research Methods*

Usage

Run via website. To use the chronset website, create a .zip file that contains your .wav files (and *only* your wav files, without any subdirectory structure) and upload it to the website. Currently, Chronset requires that these files be readable using matlab's **wavread2** routine, which is able to read most standard .wav files. There have been some issues with using wav2read in the latest versions of matlab, so the program will try to use the audioread routine in these later versions as a fallback. This fallback routine has not been tested on as wide a range of systems as the wav2read functionality, however.

Note to MacOS users: Some methods of creating .zip files on this operating system will include a hidden folder within the zip file (_MACOSX) that also contains stubs of the .wav files in the main part of the .zip. This hidden folder must be deleted before upload, otherwise Chronset will attempt to process those files and yield an error.

If your recording is in stereo, the first channel will be used. Better performance can be achieved by using a sampling rate of 22kHz or higher (e.g., 44kHz) with the current version of Chronset. Once your files have been uploaded, you will receive a confirmation message indicating that they have been scheduled for processing.
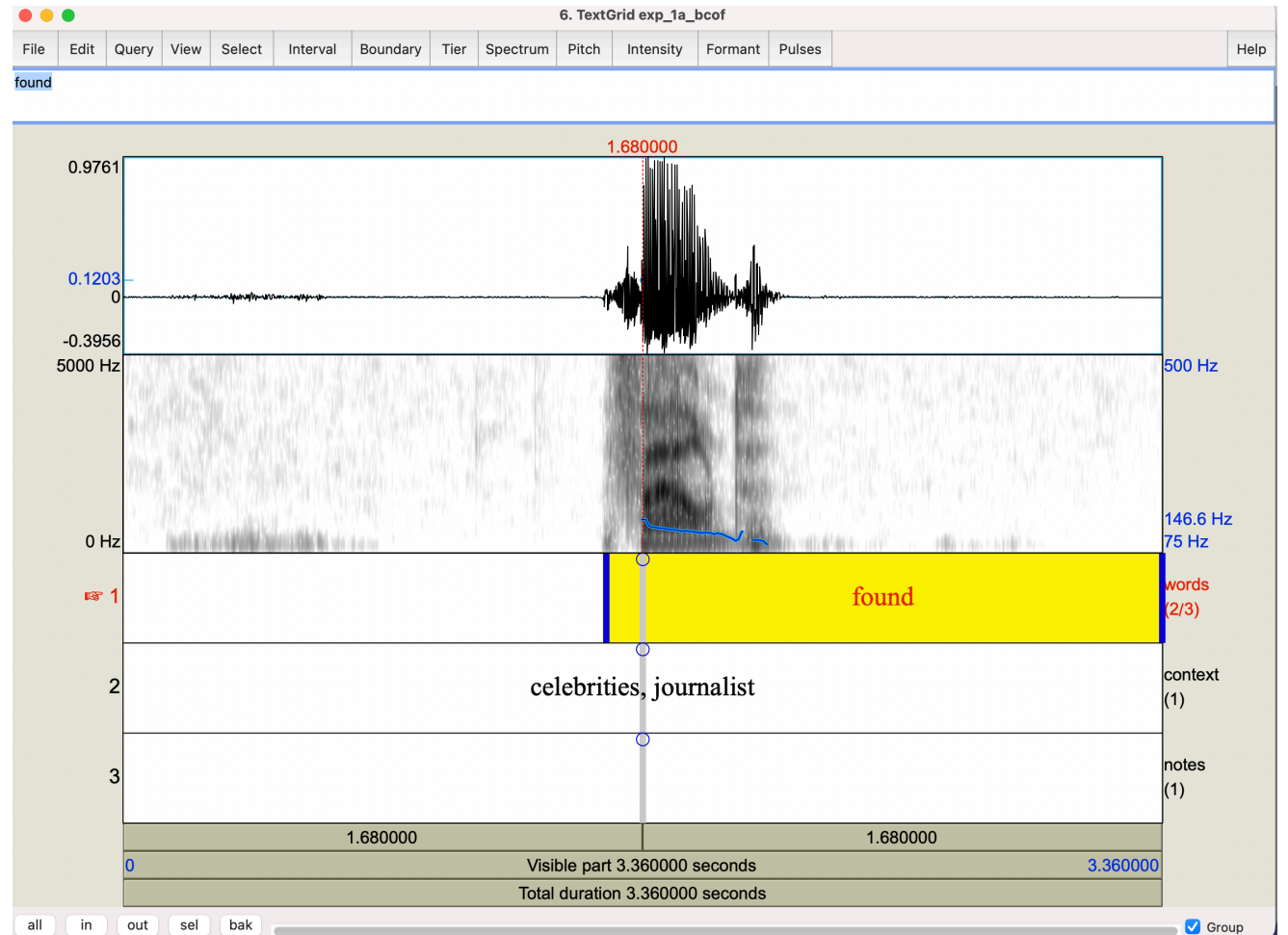
Run on your computer. Alternatively, you can download the Chronset source code (written in Matlab) and run it on your local computer. Click **here** to download the source code.

Advantages of Chronset

- Pretty accurate
- Very robust to noise
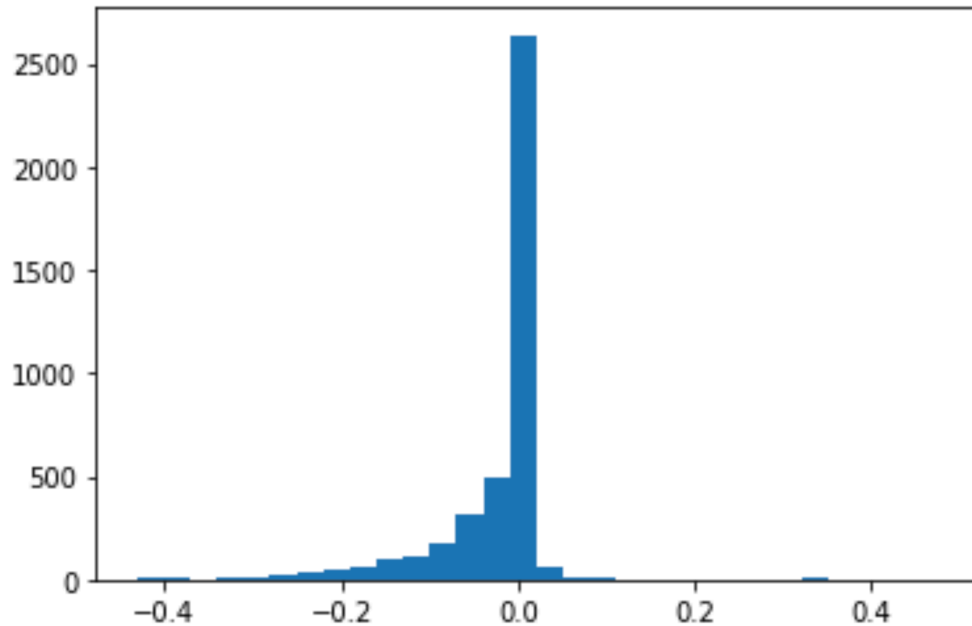
Difference < 0.1s: 89.61%
Difference < 0.02s: 67.76%

Shortcomings of Chronset

- Does not handle closures well

- Vulnerable to speech noises

- Require uploading of the data

- Takes long time
  500 files with 3.3 second recordings take 1-2 hours

- And the processing time depends on the length of the line

Other options

Intensity detection

- very vulnerable to any kind of noise
- works worse for heterogeneous data (like online-collected data)

Forced aligners

- Not accurate for onset

Github repository: https://github.com/Masato-Nakamura-3/transcription_pipline

Chronset: https://www.bcbl.eu/databases/chronset

Google Cloud Speech-to-Text: https://cloud.google.com/speech-to-text