GailBot: An automated system for Jeffersonian transcription of conversation

Muhammad Umair, Tufts University, Department of Computer Science
Julia Mertens, Tufts University, Department of Psychology
Saul Albert, Loughborough University, School of Social Sciences and Humanities
JP de Ruiter, Tufts University, Departments of Psychology and Computer Science
Correspondence email: muhammad.umair@tufts.edu

Transcription is essential for studying language in interaction. Some automatic transcription systems reproduce the words and timings of different utterances. However, interlocutors use many other behaviors to express and perceive meaning: silence, pitch, in- and out-breaths, laughter, speech rate, and other paralinguistic features of talk. To represent an interaction accurately, a transcript must also represent these features.

Conversation Analysts annotate these features in so-called Jeffersonian transcriptions, named after the founding Conversation Analyst Gail Jefferson. Unfortunately, producing a Jeffersonian transcription requires extensive expertise and time; it takes an expert approximately 10 minutes to transcribe one minute of conversation. Consequently, most corpora do not contain annotations of paralinguistic features. Instead, scientists are forced to choose between using a larger, less detailed corpus or a small, fully-transcribed dataset.

To enable researchers to study larger corpora in Jeffersonian detail, we have developed GailBot. It automatically transcribes conversations, including annotations of laughter, overlaps, silences, and changes in speech rate. For examples of each of these features, see Excerpts 1-3. GailBot transcripts are very useful as a first draft, which substantially decreases the time and cost of developing conversation-ana- lytic transcripts.

GailBot works in three stages. First, it sends audio data to IBM's Watson for Speech- to-Text analysis. Watson returns words and timestamps. Depending on the quality of the audio, the background noise, and the Automated Speech Recognition algorithms, the Word Error Rate ranges from 0% to 22%. When GailBot receives the output from Watson, it calculates the time between every word, and organizes the talk into lines depending on silences (e.g., Excerpt 1, lines 2-4) and speaker transitions (e.g., Excerpt 2, lines 2-4). Any within-turn pauses that meet a duration threshold are marked in the transcript. Since this algorithm cannot identify Turn Construction Units or Transition Relevance Places, GailBot occasionally introduces turn construction errors. However, these errors are not common, and are easily corrected.

The third stage involves GailBot *modules*. These modules enable users to transcribe other features of the talk, depending on their needs. Currently GailBot has modules for transcribing laughter, overlapping talk, silences, and changes in speech rate. The laughter module uses TensorFlow to locate instances of laughter. The overlap module calculates the proportion of overlap in the first and second turns, and places overlap markers accordingly (e.g., Excerpt 3, lines 2-3). This module provides a better understanding of the sentential context, although the placement of the overlap marker is approximate. Finally, the speech rate module uses an outlier algorithm to identify slow or fast (e.g., Excerpt 2, line 2) speech. GailBot's final output is a transcript in one of the widely used CLAN, ELAN, or Excel formats.

GailBot is the first automated transcription system to include paralinguistic features of talk. In addition, GailBot was designed to be easily improved or customized depending on the researchers' use. GailBot can work with any available Speech-to-Text system. Further, the modules responsible for determining the paralinguistic features are independent of each other. Therefore, users can add or improve modules without interfering with the functionality of other modules. While GailBot alone cannot replace human transcription, it greatly facilities the transcription and analysis of conversation in all its richness.

Additional Materials

Excerpt 1: Turn construction and silences. Audio can be found at https://sites.tufts.edu/hilab/files/2020/06/excerpt-4-pilgrim.mp3

*SP1: Um (.) it's about like this (0.5) man (0.3) who it's like (.) he's like (.) pilgrim (0.5)
and he's (.) like
(2.5)
*SP1: kind of like an undercover agent (0.4) and he (0.5) does (.) is like (0.3) living out
like (.) in Europe and traveling in doing like

Excerpt 2: Increased speech rate. Audio can be found at https://sites.tufts.edu/hilab/files/2020/06/excerpt-6-spring-semester.mp3.

*SP2: Yeah we will perform in the parade of nations and like we will have our own show
case maybe next semester (0.3) >usually in the spring semester<
(0.4)
*SP1: Wow that's awesome how many people are in that

Excerpt 3: Overlap markers. Audio can be found at https://sites.tufts.edu/hilab/files/2020/06/excerpt-10-new-members.mp3

*SP2: Uhm so this semester we have some new members and the total number
would be like (0.5) about twenty I [think]
*SP1: ⌊Oh that's⌋ that's a nice (.) number

Table 1: Transcription symbols in excerpts 1-3

Symbol	Meaning
[ ], ⌊ ⌋	Words in the box are overlapped speech
(0.6)	600ms pause / gap
(.)	Micropause
>word<	Faster syllable rate