Is automatic speech-to-text transcription ready for use in psychological experiments?

Kirsten Ziman, Andrew C. Heusser, Paxton C. Fitzpatrick, Campbell E. Field, and Jeremy R. Manning
Psychological and Brain Sciences Department, Dartmouth College

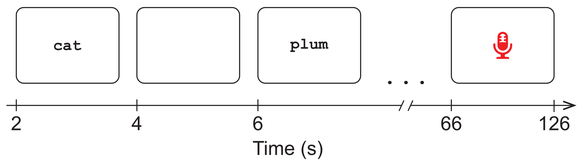
Contextual Dynamics Laboratory • www.context-lab.com

Introduction

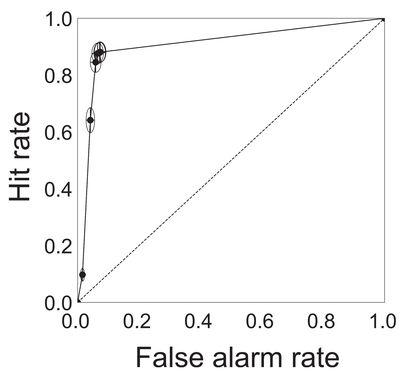
Verbal responses are a convenient and naturalistic way for participants to provide data in psychological experiments (Salzinger, 1959). However, audio recordings of verbal responses often require time-intensive processing.

Here, we evaluate the performance of a state-of-the-art speech recognition algorithm (Halpern et al., 2016) in transcribing audio data from a list-learning experiment into text. We compare transcripts made by human annotators to the computer-generated transcripts.

Free Recall Paradigm

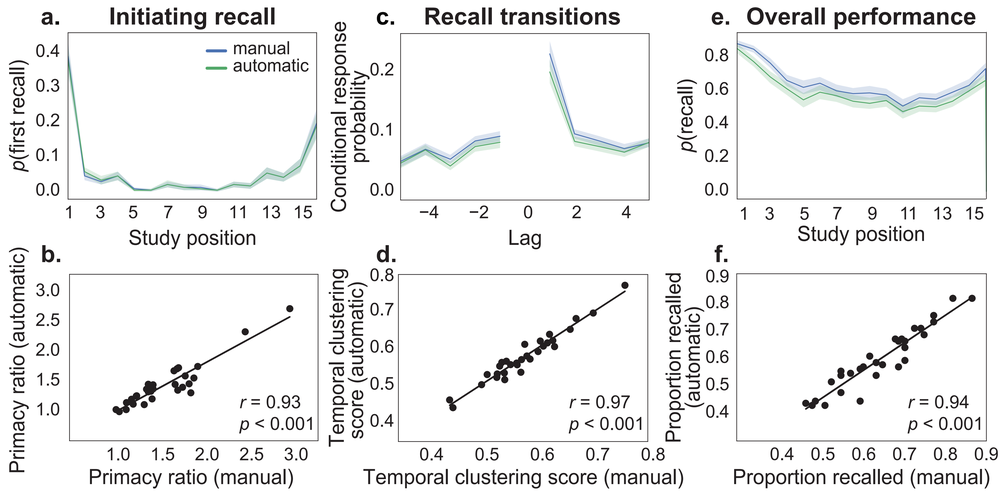


Automatic Transcription Accuracy



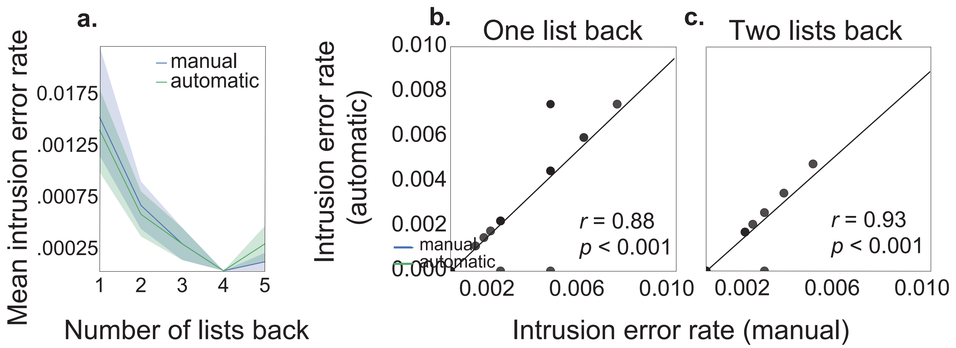
Receiving Operator Characteristic (ROC) Curve. False alarm rate and hit rate as a function of the speech-to-text engine's confidence ratings (evaluated on the interval [0, 1], in increments of 0.1). The curve reflects an average across a total of 240 lists studied by 30 participants. Error ellipses denote 95% confidence intervals (across

Recall Dynamics, Manual Versus Automatic Transcription



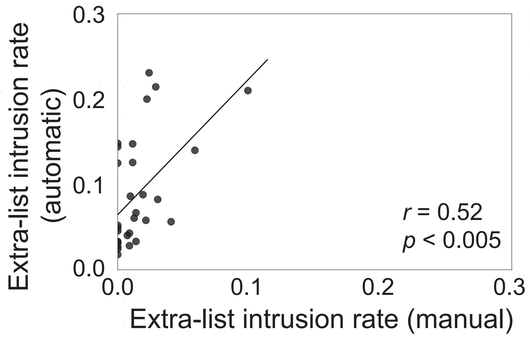
a-b. Initiating the recall sequence. **a.** Probability of recalling a word first as a function of its presentation position. **b.** Participant-wise average primacy effect. **c-d. Recall transitions.** **c.** Conditional probability of recalling a word as a function of its presentation position relative to the previously recalled word (lag). **d.** Degree of participant-wise average temporal clustering scores. **e-f. Recall probabilities.** **e.** Probability of recalling a word as a function of presentation position. **f.** Average participant-wise proportion of recalled words.

Prior List Intrusions Errors



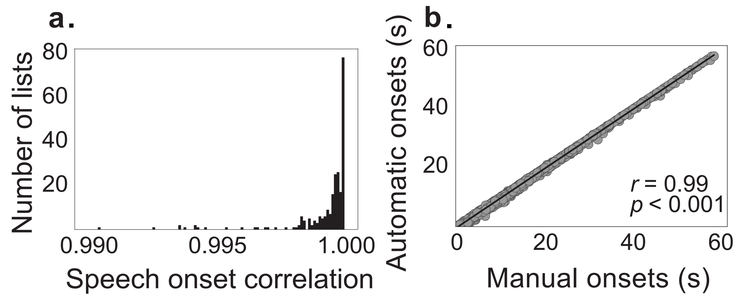
Prior List Intrusions. a. Average prior list intrusion error rates by list back (1-5). **b-c.** Average proportion of prior list intrusions (participant-wise), for one (b) and two (c) lists back.

Error Analyses, Extra List Intrusions



Extra List Intrusions. Comparison of participant-wise average proportion of etra-list intrusion errors from human- and computer-generated transcriptions.

Recall Onset Time Correlations



Recall onset times. a. Within-list correlations between automatically- and manually-recorded recall onset times. **b.** Onset times for individual recalls across all subjects, as recorded manually and automatically.

Future Use

Our findings indicate that automatic transcripts captures many of the key behavioral dynamics in free recall. Rapid verbal response transcription opens the door to on-the-fly data analysis and real-time data-based modulation of subsequent trials.

Open Source Tools. We have developed three open-source packages to streamline the automatic transcription process, analyze free recall data, and intuitively visualize and interpret results.

AutoFR
QUAIL
HyperTools

www.github.com/ContextLab

Ziman K., Heusser A.C., Fitzpatrick P.C., Field C.E., Manning J.R. (2018). Is automatic speech-to-text transcription ready for use in psychological experiments? *Behavioral Research Methods*, doi.org/10.3758/s13428-018-1037-4.

Acknowledgements. This work was supported in part by NSF EPSCoR Award Number 1632738.