Consensus Automatic Speech Recognition (CASR) in Cognitive Testing

Timothy Herron, Kathleen Hall, Gabriel Sucich, Mike Blank, Kristi Geraci, Juliana Baldo and David Woods

Scoring verbal cognitive tests with automatic speech recognition (ASR) engines increases the efficiency of scoring and provides word timestamps that enable detailed temporal analyses of spoken responses. Here, we describe consensus ASR (CASR) procedures that incorporate multiple ASR engines to increase transcription and timing accuracy. Seven ASR engines produced automatic transcriptions of both speech database samples (GMU Speech Accent Archive [1] and NUS Auditory English Lexicon Project [2]) and verbal test responses of 41 subjects from the California Cognitive Assessment Battery (CCAB). A novel, modified ROVER algorithm [3] was used to mutually align the transcripts, and a Bayesian voting algorithm [4] produced the best transcript, mean word timestamps, and consensus levels. Word error rates (WER) gauged CASR accuracy against manually corrected transcripts. Database sentence WERs ranged from a mean of 25% (Windows10) to 9% (IBM) with CASR producing 6% with no significant gender or age effects. In CCAB test responses, for limited word response tests (e.g. digitspan) CASR WERs ranged from 3% to 1%, and for discursive speech (e.g. picture description), CASR WERs ranged from 7% to 5%. Finally, word start time ASR estimates for 1732 database words (arranged in phrases) ranged in mean deviations from true times from 220ms s.d. (Google) to 11ms s.d. (Rev.ai). CASR produced transcripts for verbal test responses accurate enough for estimating scores in most word response tests and provided discursive response transcripts facilitating quick manual correction.

References

[1] https://accent.gmu.edu

[2] https://inetapps.nus.edu.sg/aelp/

[3] S.Jalalvand, M.Negri, D.Falavigna, M.Matassoni & M.Turchi, Automatic quality estimation for ASR system combination, Computer Speech & Language, Vol. 47, January 2018, pp 214-239, DOI: 10.1016/j.csl.2017.06.003.

[4] L. Kuncheva & J.J. Rodríguez, A weighted voting framework for classifiers ensembles, Knowledge and Information Systems, 38:259–275, Feb 2014, DOI: 10.1007/s10115-012-0586-6