TITLE: Consensus Automatic Speech Recognition (CASR) in the California Cognitive Assessment Battery (CCAB)

The neuropsychological assessment landscape has seen an increasing number of digital assessment tools, however, these tools often fail to include tests with verbal responses due to the inaccuracy of automatic speech recognition (ASR). Here we describe the speech transcription pipeline of the California Cognitive Assessment Battery (CCAB), which uses consensus ASR (CASR) to produce transcripts with near-human accuracy, beyond what is possible with any single engine. We also present confidence metrics to estimate transcription accuracy and guide manual correction using our Transcript Review Tool (TRT).

The speech pipeline used in CCAB tests passes digital recordings of verbal responses through six cloud-based ASR engines (e.g. Google). Individual transcripts are then combined to produce a ‘consensus’ transcript based on the concordance of engines, historical performance, and engine confidence values. Transcription also produces a CASR confidence metric that reflects the probability of mistranscription, as well as timestamped word boundaries. If needed, ‘consensus’ transcripts can be manually corrected using the Transcript Review Tool, which enables the review of all words or just those below a predefined CASR confidence.

CASR transcripts were generated from the responses of over 300 CCAB participants who underwent 3 days of CCAB testing. These transcripts were reviewed and corrected for accuracy by trained reviewers, for a total of >.5M words. Analyses show that CASR transcription accuracy surpassed 99% for tests with limited response sets (e.g., digit span, verbal list learning, face-name binding) and exceeded 98% for discursive speech tests (e.g., picture description, logical memory). Accuracies for individual ASR engines ranged from ~88% to ~96%.

CASR transcription is more accurate than that of any single ASR engine. When combined with the TRT, "consensus" ASR can produce error-free, timestamped transcripts that enable the detailed analysis of verbal responses from individuals at risk of cognitive decline or neuropsychological disorders.