

Introduction

Speech and language biomarkers (SLBs) are increasingly being studied and employed for early detection of cognitive decline. However, many commonly used speech tasks (picture description, logical memory recall) produce limited and predictable responses with formulaic lexical content and constrained structural variety, reducing sensitivity of higher-order linguistic measures.

We introduce an autobiographical event recall task in the California Cognitive Assessment Battery (CCAB) that elicits richer, more ecologically valid speech samples for automated language analysis. CCAB also collects data from a wide range of cognitive tasks, allowing for interpretation of speech and language biomarkers alongside more traditional measures of cognition.

Methods

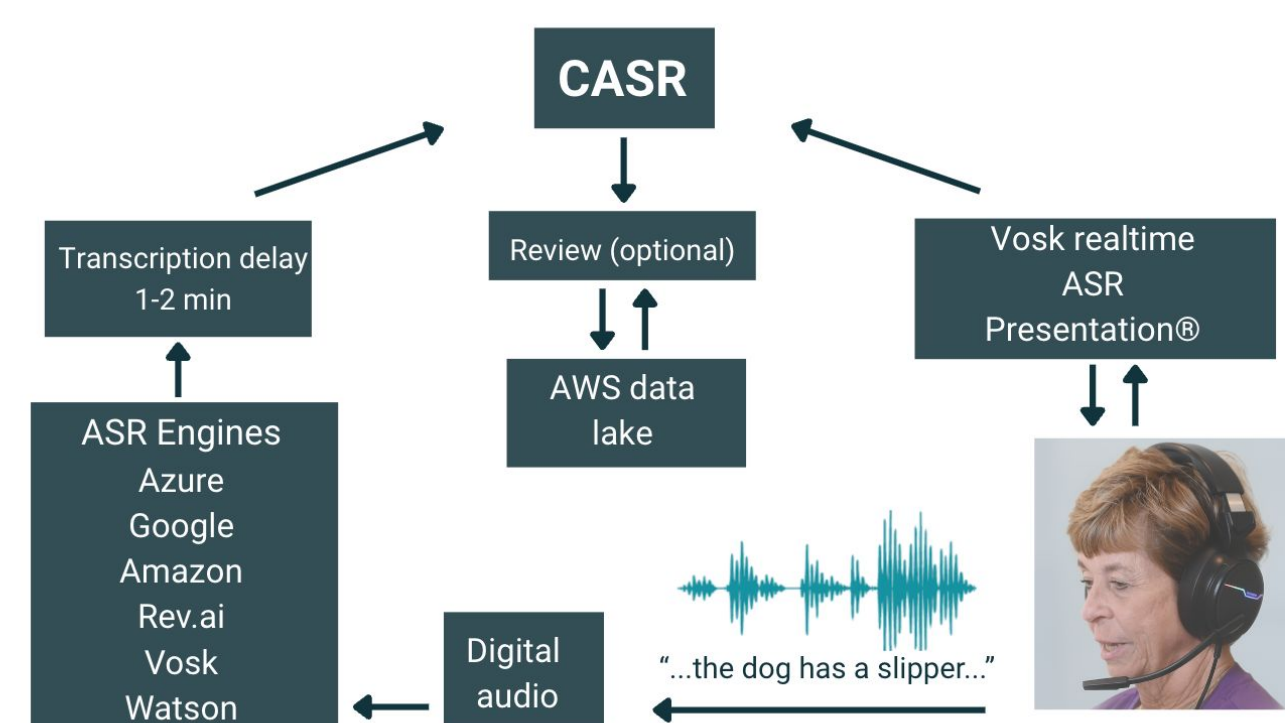
Participants: The CCAB 'Story Telling' event recall task was administered to 572 healthy adults during ongoing normative data collection. Participants completed a range of other cognitive tasks. 59.3% of participants were female, with a mean age of 42.5 years, \pm 14.4

Technology: Participants were tested in their own homes using a tablet computer with circumaural headphones and a head-mounted microphone. Instructions and stimuli were delivered using text-to-speech (TTS) with intensities adjusted to the participant's auditory threshold. Responses were automatically transcribed using consensus automatic speech recognition (CASR), with transcription accuracy exceeding 98%. An examiner remotely monitored participant performance over audio and visual feeds.

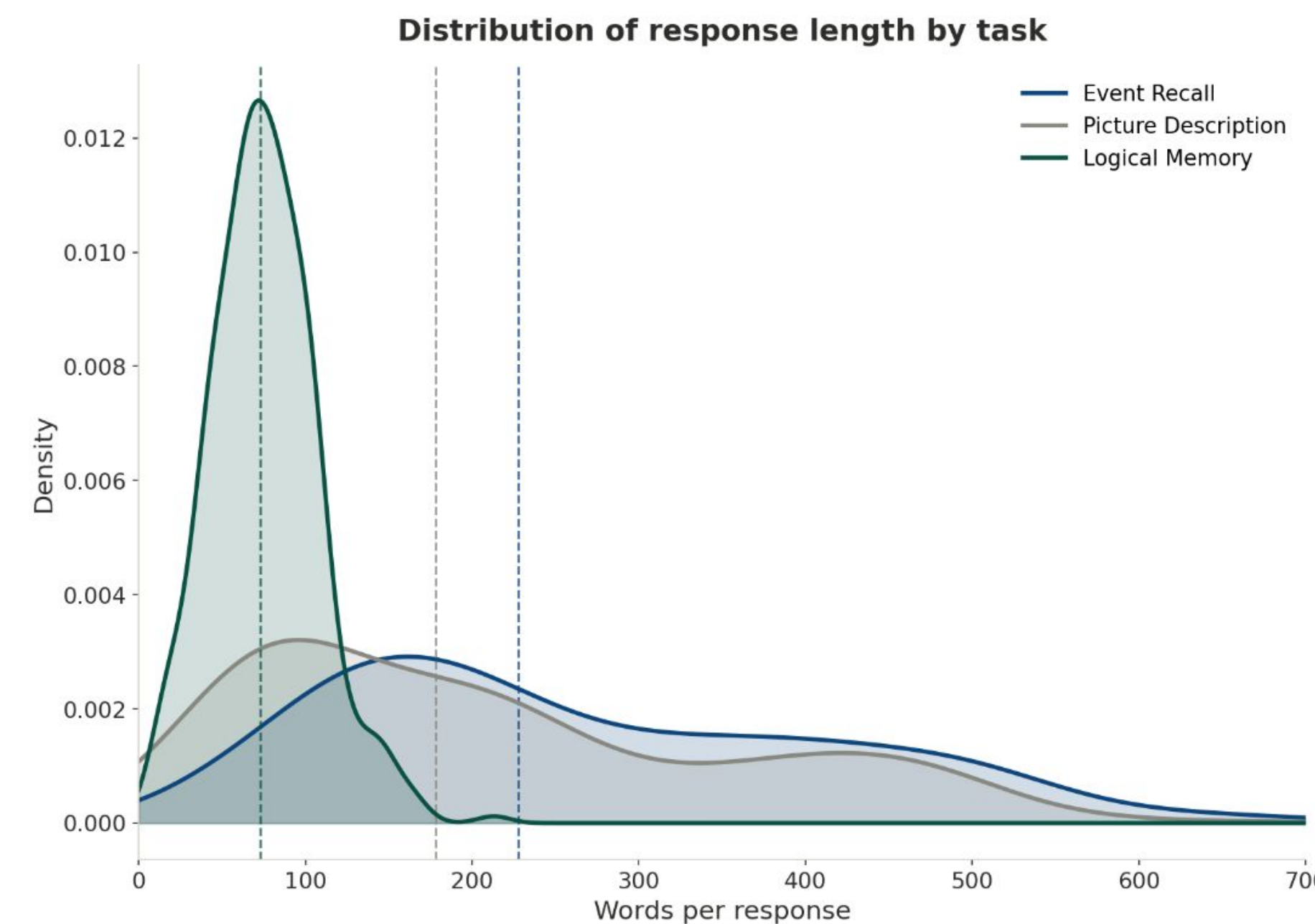
Task: Participants were asked to verbally describe a significant personal event from the prior two weeks. Instructions were automatically presented, and responses were recorded. Participants were tested in their homes with remote examiner proctoring.

Speech Analysis: Speech metrics were automatically extracted using response audio files and transcripts. Measures included acoustic variables, timing information, lexical and semantic indices, and syntactic complexity information. Metrics were generated using Praat and NLTK.

Consensus ASR pipeline: digital audio files are passed to multiple ASR engines, and then enter a weighted voting algorithm (CASR). Realtime ASR informs both CASR and realtime results reporting. Transcripts are optionally reviewed and stored in AWS before entering scoring pipeline.



Results



Speech measures across tasks (mean)

Measure	Event Recall	Picture Desc.	Logical Memory
Word count	267	211	75
Speech rate (words/sec)	2.80	2.58	2.38
Noun ratio	0.22	0.31	0.25
Vocabulary richness (Honoré's)	651	635	574
Word entropy	4.35	4.08	3.67

Means shown; measures independent of punctuation/parsing. All comparisons $p < .001$.

Summary

The CCAB event recall task elicits ecologically valid speech optimized for automated language analysis (m=267 words vs. picture description m=211 and logical memory recall m=75 words).

Longer, more diverse and more variable responses improve the interpretability of linguistic measures.

Because lexical, semantic, and syntactic measures are sensitive to sample length, the added length and diversity of event recall responses yields more stable, reliable estimates per subject than shorter speech tasks.

Integration within CCAB allows speech-derived measures to be interpreted alongside established cognitive outcomes.

Discussion

Event recall produced significantly longer and more linguistically diverse speech than picture description or logical memory paradigms.

Greater response length and variability in responses creates an opportunity to investigate structural differences, as well as lexical and semantic speech and language biomarkers.

Other cognitive tasks that generate verbal responses constrain linguistic variability in responses due to methodological design. The simple task design of event recall and CCAB infrastructure support scalable cognitive screening and longitudinal monitoring.

Normative data collection is ongoing. Future work will incorporate longitudinal data and expand analyses to additional speech and language biomarker measures.

References

- [1] Woods et al. (2024). The California cognitive assessment battery (CCAB). *Frontiers in Human Neuroscience*
- [2] Petti et al. (2020). A systematic literature review of automatic Alzheimer's disease detection from speech and language. *Journal of the American Medical Informatics Association*, 27(11), 1784-1797
- [3] Fraser et al. (2015). Linguistic features predict Alzheimer's disease in narrative speech. *Journal of Alzheimer's disease*, 49(2), 407-422.
- [4] Luz et al. (2021). Alzheimer's dementia recognition through spontaneous speech. *Frontiers in computer science*, 3, 780169.

Contact us

ccab@neurobs.com for reprints
ccabresearch.com
neurobs.com



Visit us at booth 1417! ALZHEIMER'S ASSOCIATION

Supported by NIA R44AG062076
NIA R44AG080951

