Logical Memory Performance and Speech Metrics in the California Cognitive Assessment Battery

Kathleen Hall[1], Michael Blank[1], Kristin Geraci[1], Isabella Jaramillo[1], Omar Kahly[1], Miranda Miranda[1], Peter Pebler[1], David K Johnson[2], and David Woods[1],
[1]Neurobehavioral Systems, Inc, [2]UC Davis Alzheimer's Disease Center

Introduction

Logical Memory story recall tasks are frequently used to assess episodic memory and language abilities[1], and to assist in detecting mild cognitive impairment (MCI). Here we describe a logical memory encoding and recall task from the California Cognitive Assessment Battery (CCAB). Administration and scoring of this task are automated, and allow for the extraction of speech and language biomarkers that may be predictive of performance.

Methods

Participants: The CCAB Logical Memory task was administered to 1753 healthy adults. 54% of participants were female, with a mean age of 54.1 years, ± 17.3.

Technology: Participants were tested using a tablet computer with circumaural headphones and 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 scored using consensus automatic speech recognition (CASR) with performance scores displayed in real time. An examiner remotely monitored participant performance over audio and visual feeds.

Task: Participants heard a story presented over headphones, and were asked to immediately recall (IR) the story. After an interval of ~20 minutes, they were asked to recall the story again without additional story presentation.

Scoring: Participant responses were transcribed with consensus automatic speech recognition (CASR), with transcription accuracy rates > 98%. Transcripts were automatically scored for exact match and alternative match count against a set of 46 keywords.

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.
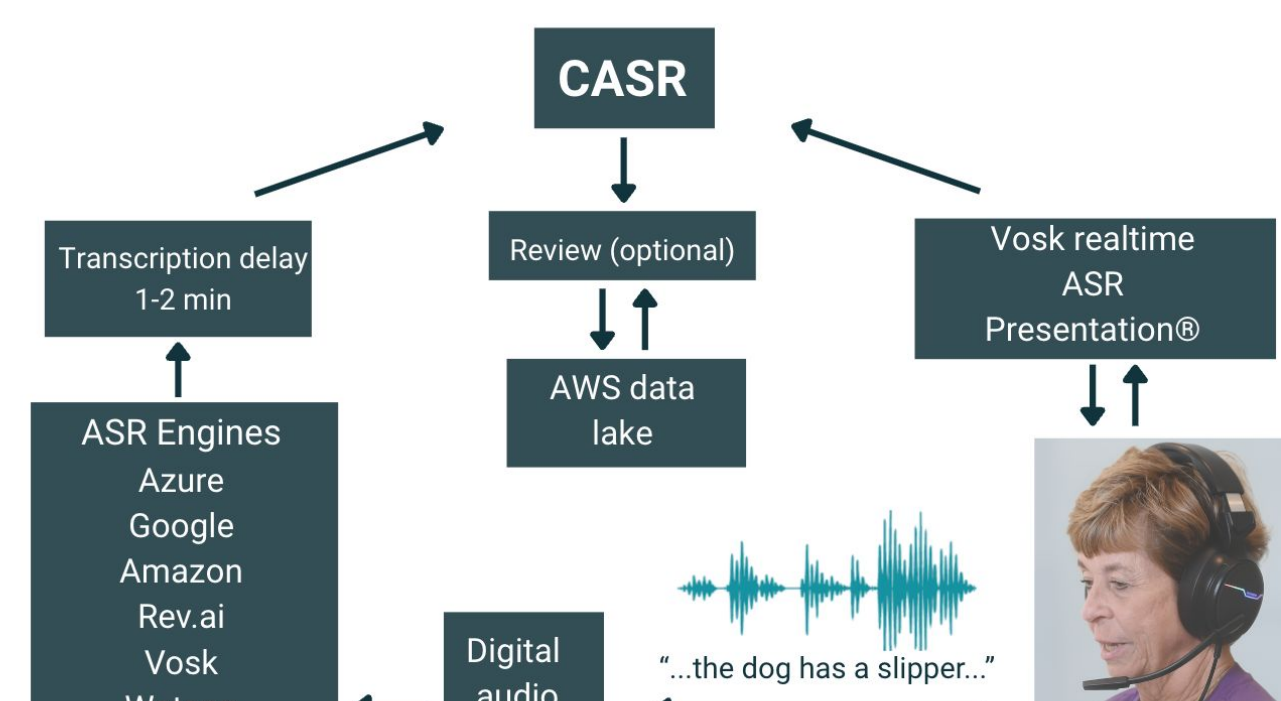
Intratest Reliability

Measure	Pearson's r	ICC
Total Match Count	.91	.9
LTAS	.83	.82
Jitter	.77	.77
Harmonicity	.86	.86
Speech Rate	.61	.55
Honore's Statistic	.55	.69
Brunet's Index	.74	.74
Word Entropy	.78	.77

Correlations with performance

Measure	Correlation with match count
LTAS	.08
Jitter	-.03
Harmonicity	.06
Speech Rate	.24
Honore's Statistic	.59
Brunet's Index	.63
Word Entropy	.7

Results
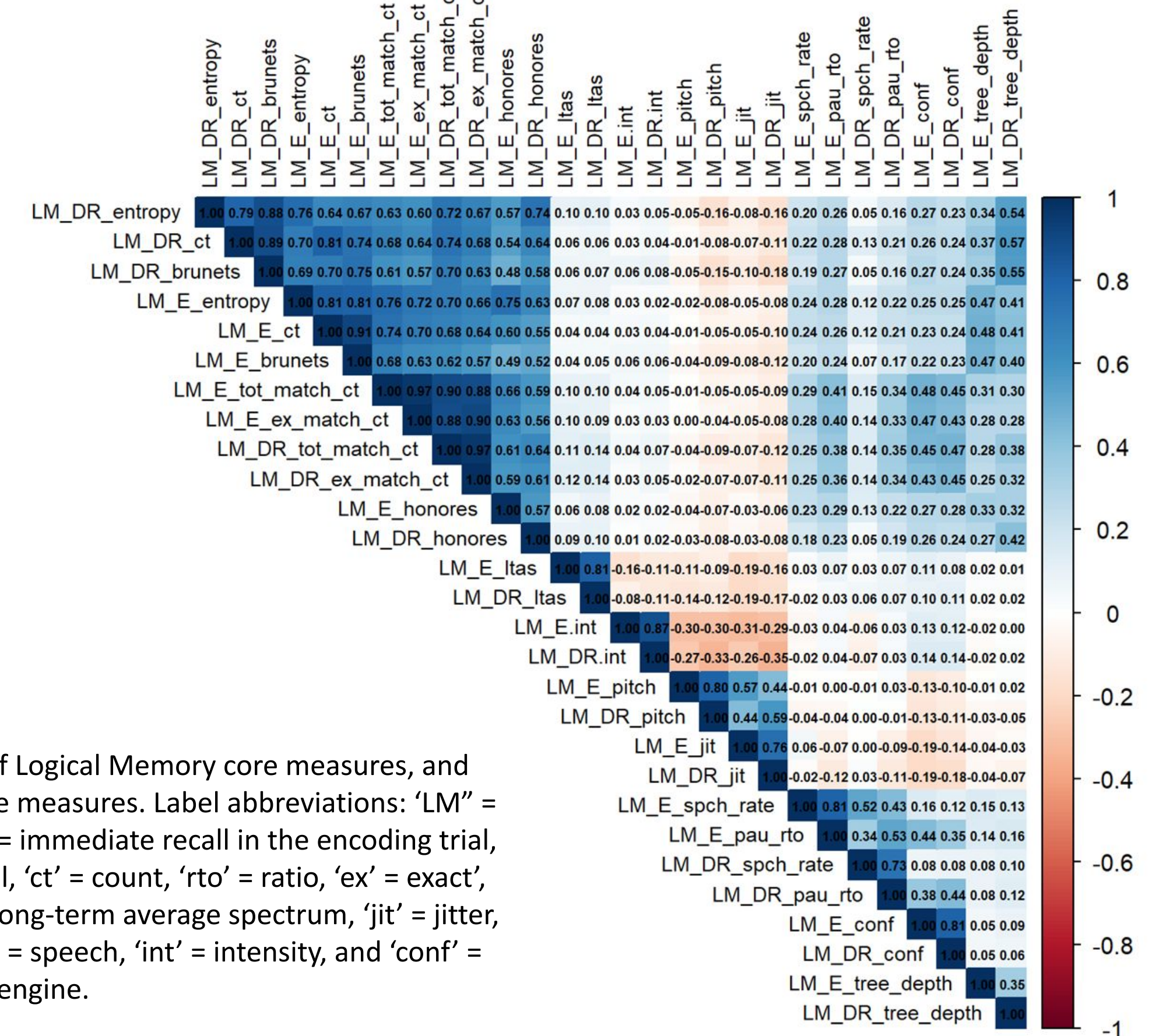
Correlation matrix of Logical Memory core measures, and speech and language measures. Label abbreviations: 'LM' = Logical Memory, 'E' = immediate recall in the encoding trial, 'DR' = Delayed Recall, 'ct' = count, 'rto' = ratio, 'ex' = exact', 'tot' = total, 'ltas' = long-term average spectrum, 'jit' = jitter, 'pau' = pause, 'spch' = speech, 'int' = intensity, and 'conf' = confidence of CASR engine.

Summary

- The core measure of 'match count' displayed excellent intratest reliability. Test retest values from a subset of the population (n=900) who completed 2 or more sessions were also very good (r=.78, ICC = .55)

- Acoustic parameters (LTAS, jitter) were largely independent of core performance metrics and demonstrated excellent intratest reliability.

- Timing metrics such as speech rate showed a moderate positive relationship with recall match count and good intratest reliability.

- Lexical and Semantic measures such as Brunet's Index and word entropy showed strong positive correlations with core performance metrics and high intratest reliability

Discussion

Results from the Logical Memory test in the CCAB demonstrate that verbal responses are captured with a high degree of accuracy, and core measures demonstrate excellent reliability.

Speech and language measures related to core performance measures, such as match count, to differing degrees. Timing measures and metrics of lexical diversity and semantic predictability had high reliability, and furthermore correlated strongly with overall performance measures. These measure show promise as predictors of performance and cognitive decline.

In contrast to some recent research[2], acoustic variables did not appear to correlate with performance, but did display high reliability. Future research will explore additional linguistic variables such as syntactic complexity.

References

[1] Storandt, M., & Hill, R. D. (1989). Very mild senile dementia of the Alzheimer type: II. Psychometric test performance. Archives of Neurology, 46(4), 383-386.
[2] Ivanova, O., Martínez-Nicolás, I., & Meilán, J. J. G. (2024). Speech changes in old age: methodological considerations for speech-based discrimination of healthy ageing and Alzheimer's disease. International Journal of Language & Communication Disorders, 59(1), 13-37.

Contact us

kat_hall@neurobs.com for reprints
ccabresearch.com
neurobs.com

Visit us at booth 646!

ALZHEIMER'S ASSOCIATION
AAIC 25