Feasibility of detecting cognitive impairment using Natural Language Processing-AI analysis of digital voice recordings of a brief neuropsychological battery obtained by a smartphone

Sam Murdock, BA[1], Erik Larsen, PhD[2], Olivia Murton, PhD[2], Lex Goldstein, BA[1], Megan Quimby, MS, CCC-SLP[1], Neguine Rezaii, MD[3], Brad Dickerson, MD[3]*, and Bonnie Wong, PhD[3,4]*

[1]Frontotemporal Disorders Unit, Massachusetts General Hospital, Boston, MA, USA; [2]Sonde Health, Boston, MA, USA; [3]Frontotemporal Disorders Unit and Massachusetts Alzheimer's Disease Research Center, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; [4]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; *Indicates senior authorship

Contact: Sam Murdock, smurdock2@mgh.harvard.edu

BACKGROUND

- Mobile-based cognitive assessments and Natural Language Processing (NLP) analyses offer a promising low-burden approach to screening for cognitive impairment.
- Recent studies suggest that digital cognitive assessments and AI-NLP tools can detect mild cognitive impairment and dementia[1,2], predict amyloid and tau accumulation[3,4], and predict future cognitive decline.[4,5]
- This study evaluated the feasibility of using a mobile app to obtain voice-based neuropsychological testing and detect cognitive impairment through an NLP-AI analysis of memory performance.

METHODS

- 34 participants with MCI or mild dementia (CDR ≤ 1) and 20 cognitively healthy controls completed a brief battery of neuropsychological tests using the Sonde One mobile app. Due to technical issues that rendered some participant responses un-scorable, 8 participants were excluded from analysis.
- Participants included all variants of primary progressive aphasia (PPA, N=15), primary progressive apraxia of speech (PPAOS, N=2), posterior cortical atrophy (PCA, N=3), amnestic syndromes likely due to AD (N=7) or FTLD (N=5), and cognitive impairment of unclear etiology (N=2).
- Participants completed a battery of neuropsychological tests assessing language, memory, and executive functioning administered on the Sonde One app. Assessments were completed at home either independently or with caregiver support.

Sonde One Cognitive Assessment User Interface

Figure 1. Example from the Sonde One mobile app neuropsychological battery showing the story immediate recall task.

- Neuropsychological test scores from each participant's baseline session were summed to create a composite score reflecting general cognitive function.
- To probe verbal memory more specifically, we focused on participants' immediate and delayed recalls of a story task, which provided a naturalistic context for analyzing narrative recall. Manual guidelines were developed to measure story content recall.
- We also applied a large language model-based algorithm to compute a Language Informative Index (LII)[6] for responses to the story recall task (Story Recall-LII). LII quantifies the semantic similarity between participant responses and the source text, emphasizing content recall over exact wording. This approach enabled a sensitive, content-based measure of verbal memory not captured by standard scoring.
- ROC curve analysis was conducted to assess the accuracy of the composite score and Story Recall-LII in distinguishing MCI and mild dementia patients from cognitively healthy controls. For a subset of 26 participants who completed the Montreal Cognitive Assessment (MoCA) within 6 months of study enrollment, post-hoc analyses were conducted to investigate the relationship of MoCA scores with the NP composite and the Story Recall-LII scores.

RESULTS

Participant Demographics and Baseline Cognitive Performance

	Control (N=19)	Patient (N=27)
Age (years, Mean ± SD)	69.3 ± 4.5	66.3 ± 8.0
% female	58	41
Years of education		
12 - 15	n/a	3 (11%)
16 - 18	9 (47%)	13 (48%)
19 +	5 (26%)	4 (15%)
Unknown	5 (26%)	7 (26%)
% of mobile app tasks completed at baseline session (Mean ± SD)	93 ± 15	77 ± 26
Time to complete baseline session (mins, Mean ± SD)	11.0 ± 1.8	17.0 ± 9.3
Baseline NP composite score (Mean ± SD)	193.6 ± 23.1	100.0 ± 41.2
Story immediate recall (manual scoring, Mean ± SD)	20.6 ± 3.7	6.7 ± 5.3
Story delayed recall (manual scoring, Mean ± SD)	20.3 ± 4.3	6.1 ± 5.7
Story recall – LII (Immediate) (Mean ± SD, range 0 (least similar) - 1 (most similar))	0.93 ± 0.04	0.61 ± 0.25
Story recall – LII (Delayed) (Mean ± SD, range 0 (least similar) - 1 (most similar))	0.92 ± 0.06	0.57 ± 0.32
MoCA score (Mean ± SD) **	29.0 ± 1.0 (N = 5)	20.6 ± 4.5 (N = 21)

*Table 1. The NP composite includes scores from the following tasks: category fluency, mental alternation (parts A and B), passage reading, serial subtraction, and story recall paraphrase scores (immediate and delayed). **Note: MoCA scores were available only for a subset of participants who completed the assessment through other research protocols.*

ROC Analysis of Patient / Control Classification

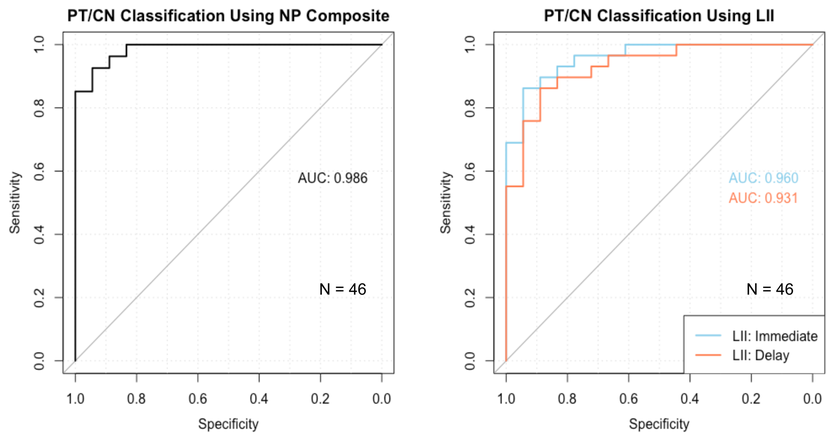

Figure 2. ROC area under the curve (AUC) was above 0.90 for both the Neuropsychological Composite and Story Recall-LII scores when classifying patients from cognitively healthy controls.

RESULTS (CONT.)

Neuropsychological Composite and Story Recall-LII (Immediate) vs MoCA

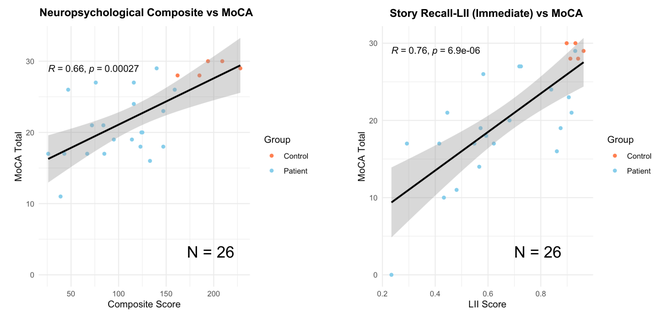

Figure 3. For a subset of participants who completed the Montreal Cognitive Assessment (MoCA) within six months of study enrollment, post-hoc analyses revealed a significant relationship (p < *0.001) between MoCA and both composite NP scores and Story-Recall LII scores.*

CONCLUSIONS

- While the NP composite and Story Recall-LII both discriminated patients from controls with high sensitivity and specificity, Story Recall-LII alone may be a powerful tool for assessing verbal memory performance.
- Both NP composite and Story Recall-LII scores were closely aligned with performance on a standardized cognitive assessments (MoCA), but Story Recall-LII (immediate) demonstrated a stronger and more robust relationship. This suggests that the LII may capture aspects of cognitive function that are more closely aligned with global cognition, potentially making it a sensitive marker for distinguishing cognitive status.
- Analyses are ongoing to investigate the discriminative power of LII with other neuropsychological subtests of the NP composite and the potential for using NLP-AI tools to differentiate among dementia syndromes and severity.
- NLP-AI approaches to cognitive assessment have the potential to transform screening by offering more efficient, low-burden methods for detecting cognitive impairment at scale.

REFERENCES & ACKNOWLEDGEMENTS

[1]Amini S, Hao B, Zhang L, Song M, Gupta A, Karjadi C, et al. Automated detection of mild cognitive impairment and dementia from voice recordings: A natural language processing approach. Alzheimers Dement 2023;19:946–55. https://doi.org/10.1002/alz.12721.

[2]Nyongesa CA, Hogarth M, Pa J. Artificial intelligence-driven natural language processing for identifying linguistic patterns in Alzheimer's disease and mild cognitive impairment: A study of lexical, syntactic, and cohesive features of speech through picture description tasks. J Alzheimers Dis 2025;106:120–38. https://doi.org/10.1177/13872877251339756.

[3]Balabin H, Tamm B, Spruyt L, Dusart N, Kabouche I, Eycken E, et al. Natural language processing-based classification of early Alzheimer's disease from connected speech. Alzheimers Dement 2025;21:e14530. https://doi.org/10.1002/alz.14530.

[4]Vanderlip, C. R., Stark, C. E. L., & Alzheimer's Disease Neuroimaging Initiative. (2024). Digital cognitive assessments as low-burden markers for predicting future cognitive decline and tau accumulation across the Alzheimer's spectrum. *Alzheimer's & Dementia*, *20*(10), 6881–6895. https://doi.org/10.1002/alz.14154

[5]Amini S, Hao B, Yang J, Karjadi C, Kolachalama VB, Au R, et al. Prediction of Alzheimer's disease progression within 6 years using speech: A novel approach leveraging language models. Alzheimers Dement 2024;20:5262–70. https://doi.org/10.1002/alz.13886.

[6]Bayat S, Sanati M, Mohammad-Panahi M, Khodadadi A, Ghasimi M, Rezaee S, et al. Language abnormalities in Alzheimer's disease indicate reduced informativeness. Ann Clin Transl Neurol 2024;11:2946–57. https://doi.org/10.1002/acn3.52205.

Funding: The research reported in this poster was supported by an award from the Massachusetts AI and Technology Council for Connected Care in Aging & Alzheimer's Disease, Year 1 Pilot Projects.