Digital Speech Markers From Brief Smartphone Recordings Identify Cognitive Impairment

Sam Murdock, BA[1], Erik Larsen, PhD[2], Olivia Murton, PhD[2], 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

INTRODUCTION

Computerized assessments using speech and Natural Language Processing (NLP) analysis offer a promising approach to accessible, scalable screening for cognitive impairment. Recent studies suggest that digital cognitive assessments and AI-NLP tools may be able to identify mild cognitive impairment and dementia[1-5], detect amyloid and tau accumulation[6], and predict future cognitive decline[7]. However, to date, few studies have explored the utility of digital speech screening outside the clinic setting and in populations with non-AD dementias.

This study evaluated the feasibility of using voice recordings collected via mobile app to detect cognitive impairment of varying etiologies in the at-home setting. Both NLP and prosodic speech analyses were compared to standardized clinical assessments.

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. 7 patients and 1 control participant were excluded from analyses because their audio samples failed quality control review. Patient syndromes 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 presumed due to AD (N=7) or FTLD (N=5), and cognitive impairment of unclear etiology (N=2).

Table 1. Demographic and clinical characteristics of participants

Cohort	Control (N = 19)	Patient (N = 27)
Age (years)	69.3 ± 4.5	66.3 ± 8.0
Sex (% female)	58	41
Education (years)	17.8 ± 1.6	17.1 ± 2.4
CDR Global (0-3)	0.0 ± 0.0	0.7 ± 0.3
MoCA Score (0-30)	-	20.1 ± 4.2

Table 1. Values represent means and standard deviations. CDR = Clinical Dementia Rating. MoCA = Montreal Cognitive Assessment.

Participants completed a battery of seven neuropsychological tests assessing language, memory, attention, and executive functioning administered on the Sonde One app. Assessments were completed at home on a mobile phone without direct support of study staff. Scores from each participant's baseline session were summed to create a composite score reflecting global cognitive function. Acoustic and prosodic features (e.g., speech rate, pause duration) were calculated from each participant's speech samples and aggregated across tasks.

To probe verbal memory further, we focused on participants' immediate and delayed recalls of a story task, which provided a naturalistic context for analyzing narrative recall. We applied a large language model-based algorithm to compute the Language Informative Index (LII)[3,4] 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.

ROC curve analysis was conducted to assess the accuracy of the NP composite score, LII scores, and prosodic feature scores in distinguishing MCI and mild dementia patients from cognitively healthy controls. In a secondary analysis, we investigated the relationship between MoCA scores and LII scores within patients.

RESULTS

Compared to control participants, patients took significantly longer to complete the neuropsychological battery; however, the mean session completion time was under twenty minutes for both groups. Patients displayed lower performance on both manually scored neuropsychological tasks and the automatically calculated LII metric. Extracted prosodic features also indicated that, on average, patients spoke more slowly and had longer pauses in speech than control participants.

Table 2. Summary of neuropsychological battery performance

Cohort	Control (N = 19)	Patient (N = 27)
Time to complete baseline session (minutes)	11.0 ± 1.8	17.0 ± 9.3 *
Baseline session NP composite score†	193.6 ± 23.1	100.0 ± 41.2 **
Story memory immediate recall, manual score (0-29)	20.6 ± 3.7	6.7 ± 5.3 **
Story memory delayed recall, manual score (0-29)	20.3 ± 4.3	6.1 ± 5.7 **
Story memory immediate recall, LII score (-1-1)	0.93 ± 0.04	0.61 ± 0.25 **
Story memory delayed recall, LII score (-1-1)	0.92 ± 0.06	0.57 ± 0.32 **
Aggregated speech rate (words/min)	120.6 ± 17.7	96.6 ± 22.6 **
Aggregated pause duration (seconds)	0.40 ± 0.11	0.72 ± 0.40 **

Table 2. Values represent means and standard deviations for neuropsychological and prosodic measures. LII = Language Informativeness Index. Values in parentheses in the first column indicate the range of possible scores.
** p < 0.05 after Bonferroni correction*
*** p < 0.005 after Bonferroni correction*
† NP composite score is comprised of aggregated scores on mental alternation (parts A and B), passage reading, category fluency, serial subtraction, and story memory (immediate and delayed recall) tasks.

Fig. 1: AUC-ROC curves demonstrate capacity to discriminate between patients and controls

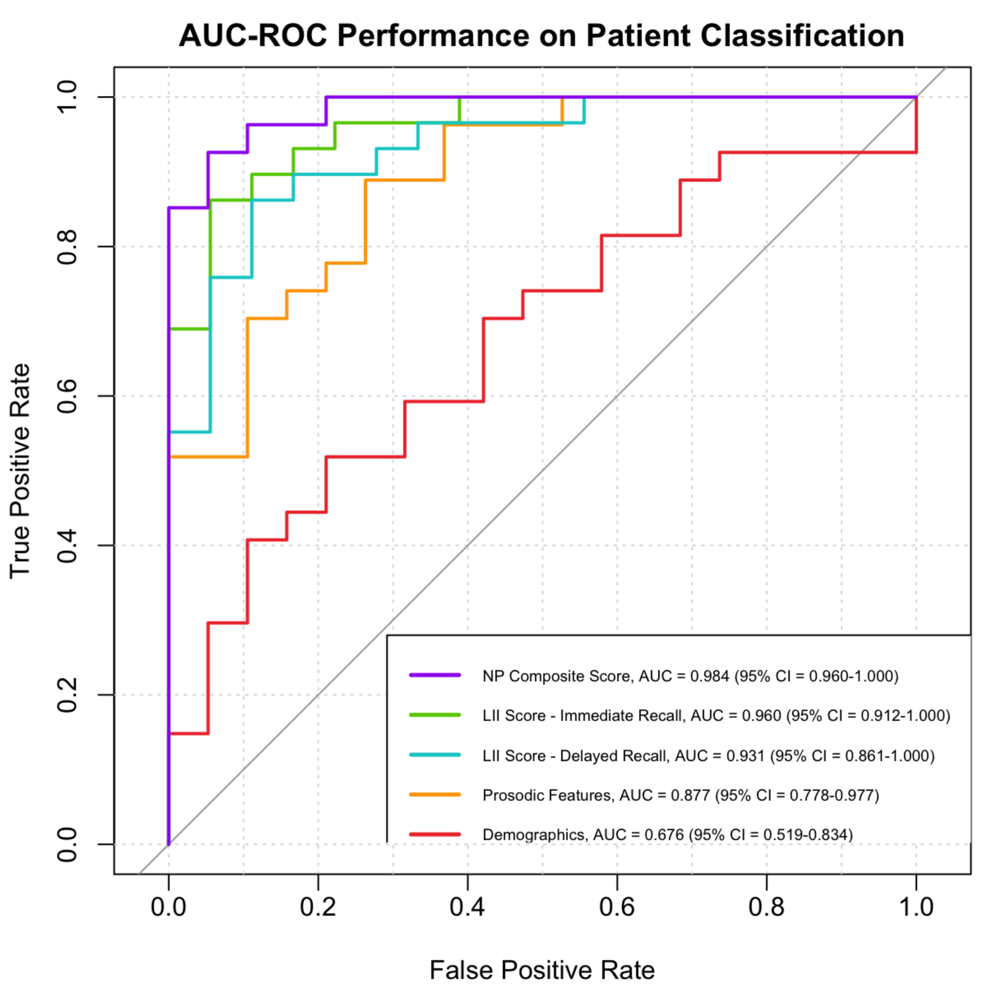

Figure 1. LII scores and prosodic features have high classification performance, comparable with the neuropsychological composite score.

ROC analysis demonstrated high classification performance for the neuropsychological composite score (AUC = 0.984, 95% CI = 0.960-1.000) and for LII scores in both the immediate (AUC = 0.960, 95% CI = 0.912-1.000) and delayed (AUC = 0.931, 95% CI = 0.861-1.000) story recall conditions.

The prosodic feature model also performed well, achieving an AUC of 0.891 (95% CI = 0.800-0.982). The demographic model had poor classification performance, achieving an AUC of 0.676 (95% CI = 0.519-0.834).

In post-hoc analyses, we probed the variability of speech rate and pause duration across individual tasks. Both metrics varied across tasks in similar patterns for patients and controls, possibly reflecting overall task difficulty.

We observed slower speech with longer pauses in the category fluency and serial subtraction tasks for both groups. Conversely, we observed faster and more fluent speech in the mental alternation part A task.

Fig. 2: Speech rate varies predictably with task type

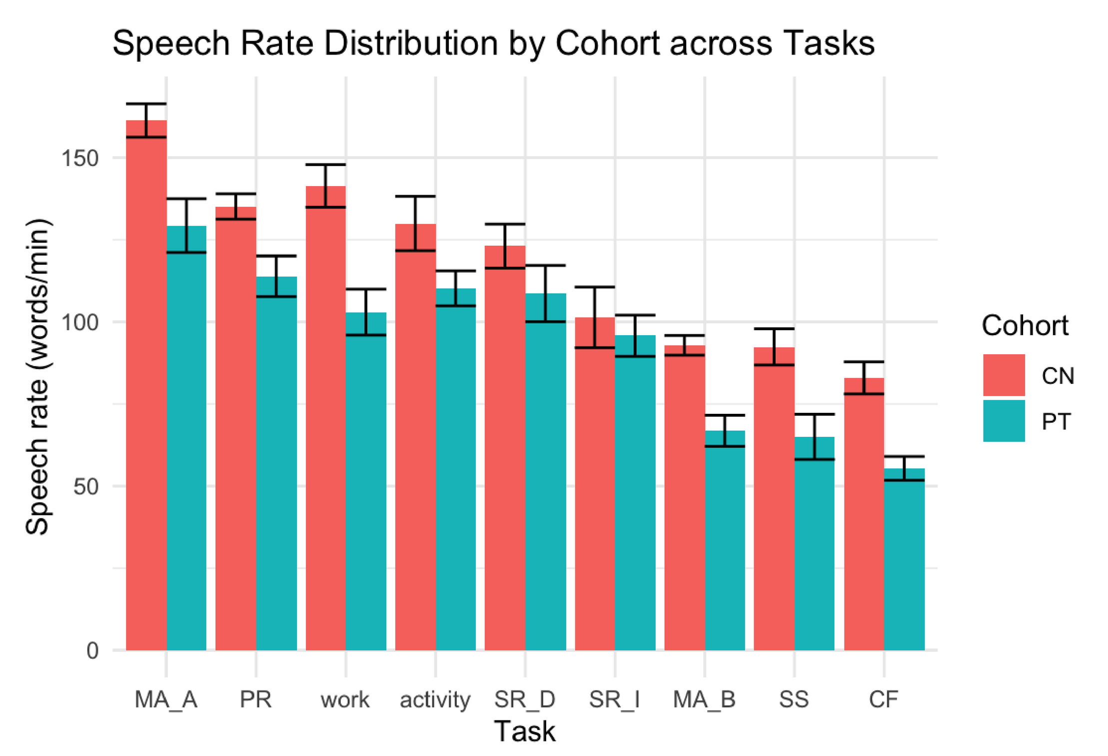

Figure 2. Patients and controls displayed similar trends in speech rate, possibly reflecting task difficulty. MA = Mental Alternation, PR = Passage Reading, SR = Story Recall, SS = Serial Subtraction, CF = Category Fluency.

RESULTS (CONT.)

Fig. 3: LII scores display positive correlation with MoCA scores among patients.

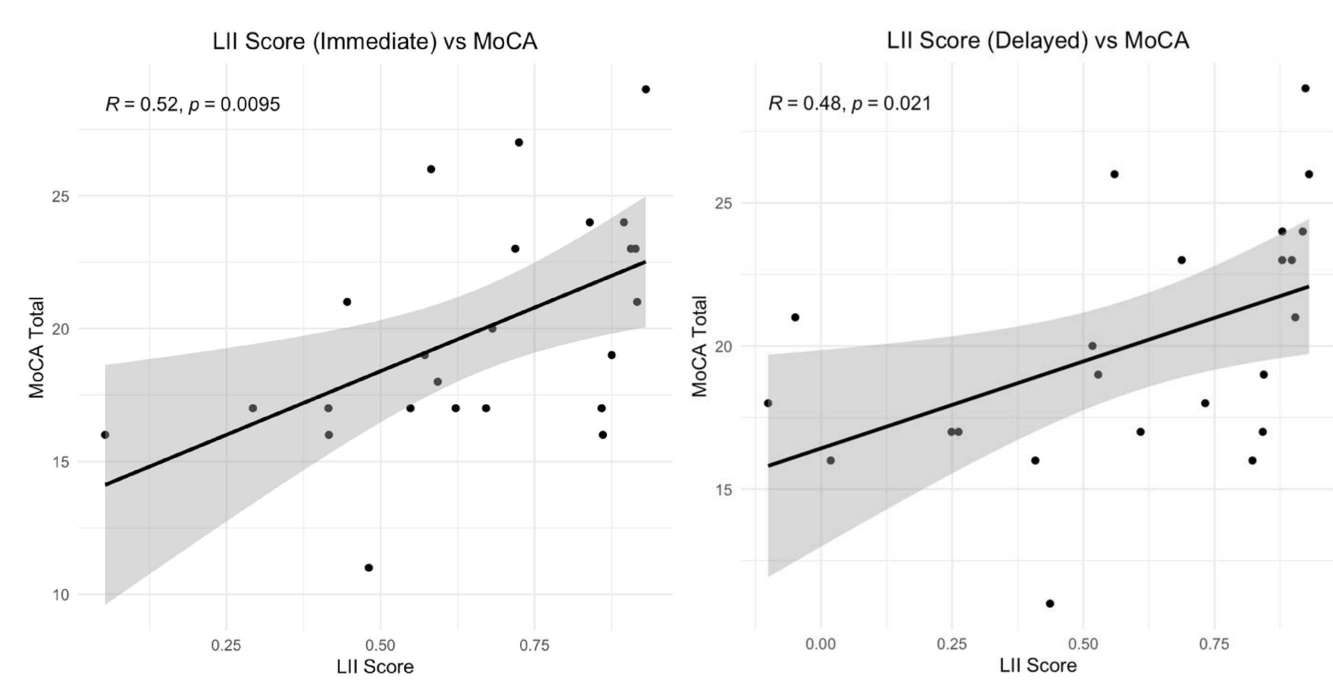

Figure 3. Both the immediate and delayed recall LII scores track with performance on the MoCA, a commonly used screening tool for assessing global cognition.

Among cognitively impaired participants, the LII score demonstrated a strong correlation with MoCA score for both the story recall immediate (R = 0.52, p < 0.01) and delayed (R = 0.48, p = 0.02) tasks. Composite neuropsychological scores were also strongly correlated with MoCA scores (R = 0.44, p = 0.02). For prosodic features, higher MoCA scores were very weakly correlated with faster speech rates (R = 0.15, p = 0.48) and shorter pause durations (R = 0.06, p = .78).

CONCLUSIONS

Evidence from this study suggests that both semantic and prosodic speech features derived from at-home audio recordings may be useful in screening for cognitive impairment. However, the generalizability of these findings is limited by small sample size and a heterogeneous patient cohort biased toward PPA.

Across patients with a variety of syndromes and etiologies, classification using LII scores and prosodic speech features had high performance, on par with manual neuropsychological tests. LII scores for both immediate and delayed story recall showed a robust relationship with patient performance on a standardized global cognitive assessment (MoCA).

Analyses are ongoing to investigate the discriminative power of LII and prosodic features within specific subtests 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]Rezaii N, Wong B, Aisen P, et al. Voiceprints of cognitive impairment: analyzing digital voice for early detection of Alzheimer's and related dementias. *NPJ Dement*. 2025;1(1):35. doi:10.1038/s44400-025-00040-0

[4]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.

[5]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.

[6]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

[7]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.

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.