Alzheimer's & Dementia
Diagnosis, Assessment
& Disease Monitoring

RESEARCH ARTICLE

Development of digital voice biomarkers and associations with cognition, cerebrospinal biomarkers, and neural representation in early Alzheimer's disease

Ihab Hajjar MD, MS[1,2] | Maureen Okafor MD, MPH[2] | Jinho D. Choi PhD[3] |
Elliot Moore II PhD[4] | Anees Abrol PhD[5] | Vince D. Calhoun PhD[5] |
Felicia C. Goldstein PhD[2]

[1]Department of Neurology, University of Texas Southwestern, Dallas, Texas, USA

[2]Department of Neurology, Emory University School of Medicine, Atlanta, Georgia, USA

[3]Department of Computer Science, Emory University, Atlanta, Georgia, USA

[4]School of Electrical & Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

[5]Tri-institutional Center for Translational Research in Neuroimaging and Data Science, Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, Georgia, USA

Correspondence
Ihab Hajjar MD, MS, Department of Neurology, University of Texas SOuthwestern, 5323 Harry Hines Blvd., Dallas, Texas 75390, USA.
Email: ihab.hajjar@utsouthwestern.edu

Abstract

Introduction: Advances in natural language processing (NLP), speech recognition, and machine learning (ML) allow the exploration of linguistic and acoustic changes previously difficult to measure. We developed processes for deriving lexical-semantic and acoustic measures as Alzheimer's disease (AD) digital voice biomarkers.

Methods: We collected connected speech, neuropsychological, neuroimaging, and cerebrospinal fluid (CSF) AD biomarker data from 92 cognitively unimpaired (40 Aβ+) and 114 impaired (63 Aβ+) participants. Acoustic and lexical-semantic features were derived from audio recordings using ML approaches.

Results: Lexical-semantic (area under the curve [AUC] = 0.80) and acoustic (AUC = 0.77) scores demonstrated higher diagnostic performance for detecting MCI compared to Boston Naming Test (AUC = 0.66). Only lexical-semantic scores detected amyloid-β status (p = 0.0003). Acoustic scores associated with hippocampal volume (p = 0.017) while lexical-semantic scores associated with CSF amyloid-β (p = 0.007). Both measures were significantly associated with 2-year disease progression.

Discussion: These preliminary findings suggest that derived digital biomarkers may identify cognitive impairment in preclinical and prodromal AD, and may predict disease progression.

KEYWORDS
Alzheimer's disease, diagnosis, digital biomarkers, lexical semantic, mild cognitive impairment, speech

Highlights

- This study derived lexical-semantic and acoustics features as Alzheimer's disease (AD) digital biomarkers.

- These features were derived from audio recordings using machine learning approaches.
- Voice biomarkers detected cognitive impairment and amyloid-β status in early stages of AD.
- Voice biomarkers may predict Alzheimer's disease progression.
- These markers significantly mapped to functional connectivity in AD-susceptible brain regions.

1 | BACKGROUND

Alzheimer's disease (AD) is characterized by progressive neuropathological changes that may begin decades before cognitive and functional symptoms appear. Consequently, efforts have focused on innovative tools or biomarkers for early identification of pre-dementia stages. The identification of in-vivo AD pathology using traditional cognitive tools has up to a 25% false positive rate when compared to autopsy confirmation.[1] Detailed neuropsychological assessments improve accuracy but may be time-consuming, costly, and may not have the specificity for in-vivo biomarker status.[2] In the Imaging Dementia Evidence for Amyloid Scanning study, 25% of those diagnosed with mild cognitive impairment (MCI) with AD and 15% of AD dementia had negative amyloid positron emission tomography (PET).[3] AD signature biomarkers, characterized by low amyloid-β-1-42 (Aβ42), and elevated total tau (t-tau) and phosphorylated (p-tau), enhance diagnostic accuracy and specificity, especially in early (preclinical and prodromal) AD.[4]

Subtle language features of connected speech may be detectable years before clinical presentation of cognitive decline. For example, in the Nun study, early low idea density and low grammatical complexity in autobiographies were associated with late-life dementia, post-mortem lower brain weight, greater cerebral atrophy, and neurofibrillary pathology.[5] However, it has been suggested that the relationship between idea density and AD risk in the Nun study may have been mediated by Apolipoprotein E allele status.[6] Evidence further suggests that early cognitive changes are underrecognized in primary care settings and hence, may deprive patients of knowledge concerning the need to plan for the future, and of clinician recommendations to engage in promising neuroprotective lifestyle intervention, and referral to clinical trials.[7-10] New brief digital-based screenings may offer additional support to primary care providers in flagging high risk individuals for further assessments.

Subtle patterns that transcend sentence structure, word count, or grammatical features may be detected via natural language processing (NLP) using graph theory to derive semantic graphs from audio recordings of cognitively normal or impaired individuals. Automatic speech recognition (ASR) acoustically distinguishes MCI in symptomatic stages of AD using prosodic cues related to pitch (e.g., rate of vocal fold vibration during voiced segments of speech), voicing (e.g., percentage of speech produced utilizing vocal folds such as with vowel

sounds as opposed to unvoiced harsher sounds usually associated with consonants), and speaking rate and formant energy (e.g., spectral shape of energy in voiced sounds).[11] Advances in NLP, ASR, and machine learning (ML) offer an opportunity to explore highly complex linguistic and acoustic changes in connected speech in a non-obtrusive way. These assessments can be automated and used in clinical artificial intelligence. Although such methods can detect MCI,[12-15] the ability to detect brain biomarker status in AD has not been sufficiently explored. The latter has become more critical as newer therapies are targeting those with brain amyloid-β positivity.

Prior linguistic changes in AD and speech production have been mapped to specific brain regions for example, atrophy in hippocampal, entorhinal, and temporoparietal regions, and speech motor control networks.[16-18] These changes also map onto altered network connectivity[19,20] including the semantic control network (SCN).[21] Resting state functional MRI (rs-fMRI) investigations have revealed alterations of functional connectivity (FC) in the cognitive control network (CCN) linked to linguistic abilities.[22-24] Since digital voice biomarkers may reflect both linguistic and speech changes, it is unclear if these novel biomarkers will map to previously known linguistic or speech brain networks.

In this study, we describe a brief protocol using connected speech and related analyses in normal cognition and MCI. We combined ML with innovative NLP and ASR to derive digital voice biomarkers for AD including acoustic features as well as 'meta-semantic' features which are based on the lexical semantic characteristics of connected speech and that capture subtle patterns that transcend sentence structure and word count and other grammatical indices. We then investigated their association with neuropsychological measures and cerebrospinal fluid (CSF) biomarkers and mapped the significant voice biomarker features to potential underlying brain networks.

2 | METHODS

2.1 | Participant description

Data were collected from participants in the Brain Stress Hypertension and Aging Research Program (B-SHARP) at Emory University. Participants were ≥50 years, cognitively unimpaired (CU) or with MCI, and identified through a referral from Emory Goizueta Alzheimer's Disease

RESEARCH IN CONTEXT

1. **Systematic review**: The authors conducted literature searches using PubMed and reviewed published literature on digital Alzheimer's disease (AD) biomarkers. With recent advances in natural language processing, automatic speech recognition and machine learning, linguistic and acoustic changes in connected speech may detect cognitive impairment and brain biomarker status in early stages of AD, but has yet to be fully explored.

2. **Interpretation**: Our findings suggest that lexical-semantic biomarkers have significant value in the detection of amyloid-β status, and both lexical-semantic and acoustic biomarkers are sensitive to cognitive status, have higher diagnostic accuracy compared to Boston Naming Test, and may be sensitive in tracking AD progression in its early stages.

3. **Future directions**: A strong potential exists for advancing technologies with connected speech measures to develop digital voice biomarkers of AD. Use of these digital biomarkers may offer additional diagnostic benefits in clinical and research settings and warrants further investigation.

Research Center or through strategic local community partnerships. Participants underwent study screening and were excluded if they had a history of stroke in the past 3 years, a clinical diagnosis of dementia of any type, abnormal serum thyroid stimulating hormone or vitamin B12, or had no study informant defined as an individual who maintained regular contact with the participant ≥ 1 time per week. Emory Institutional Review Board approved the study protocol, and each participant provided a written informed consent.

2.2 | Cognitive assessments

We collected two separate cognitive batteries for: (i) cognitive categorization (collected at screening) and (ii) cognitive monitoring (collected at baseline and follow-up). The cognitive categorization battery was based on modified Petersen MCI criteria[25]: (a) subjective memory complaints; (b) Montreal Cognitive Assessment (MoCA) score <26; (c) Clinical Dementia Rating (CDR) score, memory subscale = 0.5; (d) Wechsler Memory Scale-Revised delayed Logical Memory subscale, and (e) Functional Activities Questionnaire (FAQ) score ≤ 7. Supplement Section 1.1 provides additional details of the cognitive criteria and scoring protocol.

Additional cognitive monitoring tests[26] included episodic memory (Hopkins Verbal Learning Test-Revised), executive functioning (Trail Making Test), confrontation naming (15-item Boston Naming Test, BNT) and timed phonemic Verbal Fluency Test. The BNT was specifi-

cally of interest for comparison with digital voice biomarkers since this was our main linguistic task.

2.3 | Audio recording

Participants underwent audio recordings during their baseline visits. Adequate near card visual acuity was confirmed and a test recording was performed to assess audio clarity prior to the start of each recording. We designed a study protocol (see Supplement) to capture connected speech using a picture description, non-structured natural speech and speech during the verbal fluency and confrontation naming tasks. The picture description task using the picture, "The Circus Procession" (Figure 1, public domain from Juvenile collection, 1888[27]), provided an ecological approximation of conversational abilities (i.e., spoken language production used in a spontaneous and continuous manner).[28–31] Each task was recorded for 1–2 min on an Apple device (iPod) and audio files stored on a secure server.

2.4 | Acoustic and lexical-semantic analyses

We derived acoustic features using a standardized approach recommended in the Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research.[32] GeMAPS represents a common baseline for evaluating efficacy of specific acoustic features in relation to various speech studies. For this analysis, multiple measures related to speech motor control possibly affected by neurodegeneration were included. Our acoustic analysis workflow was conducted on the full audio file including conversational recordings and cognitive assessments.[33] A list of features is provided in the online supplement (Table S1) and their derivations are described in Eyben et al.[34] In addition, we derived an equivalent sound level measure which adjusts for the variability in audio recording settings.

Our NLP approach has been published[35] and is described in the online supplement (Figure S1). JC conducted NLP analysis on 100 MCI and CU participants selected randomly from the sample but matched on MoCA to examine the ability of the developed digital biomarkers in distinguishing cognitive status and amyloid-β status in the specific situations where commonly used screening tasks like MoCA may overlap and hence, alone may be insensitive to group differences. Comparisons of derived lexical-semantic indices (described in Supplement Section 1.2.5, Figure S2) were conducted between cognitive and CSF biomarker status.

2.5 | ML approach

Due to differences in the nature of NLP and acoustic data, we implemented two ML approaches. For the acoustic analysis, we first conducted a feature selection step to obtain features that provided the most information differentiating CU versus MCI from the full sample.[36–38] We compared three feature selection approaches using

FIGURE 1 Picture Description Task using The Circus Procession (public domain from Juvenile collection, 1888[27]).

a classification ML modeling (Table S2). The model with the highest accuracy was used to derive a model-based digital *acoustic score* for each participant, which was then advanced into the statistical analyses. For the NLP analysis, a classification prediction ML model was also employed using logistic regression (LR), neural network (1 and 2 hidden layers),[35] and an ensemble model. The approaches with the highest accuracy/performance were then used to derive *lexical-semantic scores* advanced into the statistical analyses. Feature selection results and ML modeling accuracy scores are provided in the online supplement (Supplement Sections 1.2.5 and 1.2.6, Figure S1, Tables S3-S4).

2.6 | AD Biomarker and brain MRI measurements

Following a fast, baseline CSF samples were collected via lumbar puncture using 24G Sprotte atraumatic spinal needles and sterile polypropylene tubes (Falcon Fisher Scientific UNSPSC Code 41121703), separated into 0.5 cc aliquots, stored at -80°C, and shipped for analysis by the Biomarker Research Laboratory, University of Pennsylvania. CSF biomarkers (Aβ42, total tau, and p-tau$_{181}$) were measured using the Multiplex xMAP Luminex platform (Luminex Corp, Austin, Texas, USA) with Innogenetics (INNO-BIA AlzBio3; Ghent, Belgium) immunoassay reagents.

Brain MRI data were collected using a 3.0 Tesla Trio MRI scanner (Siemens Medical Solutions, Malvern, Pennsylvania, USA). High-resolution, T1-weighted images, hippocampal volume and other volumetric measurements were collected.[39] Left and right hippocampal volumes were obtained and combined to derive the total hippocampal volume. Intra-cranial volume was derived for adjusted measurements.[4] rs-fMRI data were acquired to assess FC between brain regions.[40]

See Supplement for details of MRI assessment and pre-processing (Section 1.3).

2.7 | Statistical analyses

We compared demographic, clinical, imaging and CSF biomarkers in CU and MCI using Student *t*-test or Chi-square test. We assessed the discriminatory/diagnostic ability of digital voice biomarkers (acoustic and lexical-semantic) in detecting cognitive status (two groups) and CSF biomarker status (two groups) using receiver operating characteristic (ROC) curves and area under the curve (AUC) for the derived scores.[41] For reference, we provide similar analyses for our main language task (BNT) as well as the other cognitive tests. We then tested performances of the voice measures in detecting CSF Aβ42 status in CU and the MCI separately (four groups in total). We also assessed associations of derived digital voice biomarkers with other AD indicators including hippocampal volume and CSF tau (both as continuous variables) and evaluated their associations with disease progression assessed as the change in CDR Sum of Boxes (CDR-SOB) over the subsequent 2 years using regression analysis with residual distribution check for model fit. Statistical analyses were conducted using SAS v9.4 (Cary, North Carolina, USA) and ML performed using Waikato Environment for Knowledge Analysis (WEKA; v3.8.4, Hamilton, New Zealand).[42]

3 | RESULTS

At the time of this analysis, we had screened 300 participants and 91 were excluded (25 refused LP and/or MRI, three had dementia, 12 had

TABLE 1 Overall comparisons of demographic, neuropsychological, and biomarker characteristics by cognitive status

Characteristics	CU Mean ± SD* (n = 92)	MCI Mean ± SD* (n = 114)	p-value
Age (years)	63.2 ± 7.3	64.9 ± 7.2	0.09
Sex, no. (%)			0.19
Female	60 (65.9%)	65 (57.0%)	
Male	31 (34.1%)	49 (43.0%)	
Race, no. (%)			0.0013
White	56 (61.5%)	43 (37.7%)	
Black or African American	34 (37.4%)	71 (62.3%)	
Asian	1 (1.1%)	0 (0.0%)	
Education (years)	16.1 ± 2.5	15.2 ± 2.9	0.033
Marital status, no. (%)			0.016
Married	45 (49.5%)	43 (37.7%)	
Divorced/separated	17 (18.7%)	37 (32.5%)	
Widowed	7 (7.7%)	18 (15.8%)	
Single/never married	22 (24.2%)	15 (13.2%)	
BMI, kg/m^2	28.3 ± 6.4	31.0 ± 7.9	0.0093
Family history of dementia, no. (%)	54 (60.7%)	42 (37.5%)	0.0011
MoCA total score	26.7 ± 2.3	21.6 ± 3.7	<0.0001
Trail A completion time (sec)	34.2 ± 11.5	42.2 ± 17.7	0.0002
Trail B completion time (sec)	83.6 ± 47.4	137.5 ± 77.5	<0.0001
HVLT-R, delayed recall	9.6 ± 2.1	6.8 ± 3.1	<0.0001
Boston Naming Test	14.3 ± 1.2	13.3 ± 1.7	0.0005
CDR SOB	0.05 ± 0.12	0.61 ± 0.42	<0.0001
CDR-SOB, yearly change	0.05 ± 0.12	0.61 ± 0.63	<0.0001
Hippocampal volume (mm^3)	7548 ± 871	7022 ± 1030	0.0003
CSF AD biomarkers			
Aβ42, pg/ml	256 ± 65	228±74	0.02
% Aβ42 positive	43.4%	55.1%	0.03
Total tau, pg/ml	53 ± 30	65 ± 45	0.0007
p-tau, pg/ml	14 ± 7	16 ± 9	0.009

Abbreviations: AD, Alzheimer's disease; Aβ, amyloid beta; BMI, body mass index; CDR SOB, Clinical Dementia Rating - Sum of Boxes; CSF, cerebrospinal fluid; CU, cognitively unimpaired; HVLT-R, Hopkins Verbal Learning Test-Revised; MCI, mild cognitive impairment; MoCA, Montreal Cognitive Assessment; p-tau, phosphorylated tau; SD, standard deviation.
*Values are mean ± SD or n (%).

other neurological, psychiatric or health issues which precluded their inclusion, four did not have a study partner, one did not speak English, and the remainder declined due to a lack of time or loss of interest). Three additional participants did not have adequate voice recordings. The final sample included 206 participants (51% African American), of whom 92 (45%) were CU and 114 (55%) met clinical consensus criteria for MCI. No participant refused audio recordings or reported privacy concerns related to audio recordings. At baseline, 40 (43%) CU and 63 (55%) MCI participants were Aβ-positive. Participants were reassessed after 2 years and showed an average change in CDR-SOB of 0.29 (SD = 0.5, range = 0–3). Table 1 describes demographics, psychosocial, neu-

ropsychological, and biomarker characteristics between the CU and MCI groups.

3.1 | Acoustic and lexical-semantic score analysis

Table 2 shows the acoustic and lexical-semantic scores as a function of cognitive status, CSF biomarker status, and cognitive/biomarker status. The MCI group had significantly higher acoustic scores compared to the CU group (p < 0.0001). These differences remained significant after adjusting for age, race, sex, and years of education

TABLE 2 Derived acoustic and lexical-semantic scores presented as adjusted least square mean ± standard error in the cognitive, Aβ positive, and combined cognitive and Aβ positive subgroups

	Acoustic score LSM ± SE*	p-value	Lexical-semantic score LSM ± SE*	p-Value
Cognitive subgroups				
Cognitively Unimpaired	−0.47 ± 0.08	<0.0001	0.23 ± 0.07	<0.0001
MCI	0.85 ± 0.03		−0.22 ± 0.08	
Aβ subgroups				
Aβ negative	−0.03 ± 0.11	0.813	0.23 ± 0.08	0.0004
Aβ positive	0.09 ± 0.11		−0.19 ± 0.08	
Cognitive and Aβ subgroups				
CU/Aβ-	−0.48 ± 0.23	<0.0001	0.34 ± 0.12	<0.0001
CU/Aβ+	−0.53 ± 0.24		0.17 ± 0.09	
MCI/Aβ-	0.73 ± 0.26		0.12 ± 0.10	
MCI/Aβ+	0.63 ± 0.25		−0.51 ± 0.09	

Abbreviations: Aβ-, amyloid beta negative; Aβ+, amyloid beta positive; CSF, cerebrospinal fluid; CU, cognitively unimpaired; LSM, Least square mean; MCI, mild cognitive impairment; SE, standard error.

*Values are reported as adjusted least square means ± standard errors. All means and p-values are adjusted for age, sex, education, and race. p-Values are model-derived for between group differences.

($p < 0.0001$). In contrast, lexical-semantic scores were lower in the MCI versus CU group ($p < 0.0001$) and remained significant after adjusting for covariates ($p < 0.0001$). In the whole sample, there was no significant difference in acoustic scores between Aβ-positive (Aβ+) and negative (Aβ-) participants ($p = 0.42$). However, in the MCI subgroup, acoustic scores were significantly higher in participants who were Aβ+ compared to those who were Aβ- ($p < 0.0001$). Lexical-semantic scores were significantly different between Aβ+ versus Aβ- participants ($p = 0.0003$) in the full sample and MCI subgroup ($p < 0.0001$). These results remained significant after covariate adjustments.

3.2 | Diagnostic performance for detecting MCI and amyloid status

We further compared digital voice biomarker scores to the BNT in detecting MCI or Aβ status using ROC analyses. Lexical-semantic scores (AUC = 0.80) and acoustic scores (AUC = 0.77) had comparable or higher diagnostic performance for detecting MCI relative to the BNT (AUC = 0.66). This was also true for verbal fluency (AUC = 0.68 for MCI and 0.52 for Aβ status) and episodic memory (AUC = 0.75 for MCI and 0.58 for Aβ status) (Table S5). Lexical-semantic scores also demonstrated high diagnostic performance for detecting Aβ+ status (AUC = 0.77), and for distinguishing status within the CU/Aβ and MCI/Aβ subgroups (AUC = 0.61 and 0.81, respectively). Overall, lexical-semantic scores outperformed acoustic scores, BNT and MoCA in Aβ detection (Figure 2, Table S6). In addition, we compared diagnostic performance by race as potential bias might be present in audio recordings. Our analysis did not show a difference by race in either lexical seman-

tic or acoustic analyses. These results are provided in Figure S2 in supplemental materials.

3.3 | Association with tau biomarkers, hippocampal volume, and disease progression

Acoustic scores showed a significant association with hippocampal volume ($R^2 = 0.03$, beta = -60.7, $p = 0.017$), whereas lexical-semantic scores did not ($R^2 = 0.04$, beta = -57.6, $p = 0.38$). None of the digital voice biomarkers were associated with total tau or p-tau (Table S7). Both lexical-semantic (beta = -0.18; $p = 0.0097$) and acoustic scores (beta = 0.26, $p = 0.009$) were associated with change in CDR-SOB after adjusting for covariates (Figure 3, Table S7).

3.4 | Mapping digital voice biomarkers to brain connectivity

Derived acoustic and lexical-semantic scores were mapped to specific brain networks involved in cognitive and language control. Specifically, significant positive correlations were observed between digital biomarkers in the cognitive control somatomotor (CC-SM) networks; for instance, connections of superior medial frontal gyrus (SMFG) and inferior frontal gyrus (IFG) in the CC domain compared to the paracentral lobule, right postcentral gyrus, superior parietal lobule regions in the SM domain. Likewise, a number of somatomotor default mode (SM-DM) connections showed positive correlations with lexical-semantic scores. In contrast, a number of subcortical default mode (SC-DM), subcortical cognitive control (SC-CC), and somatomotor
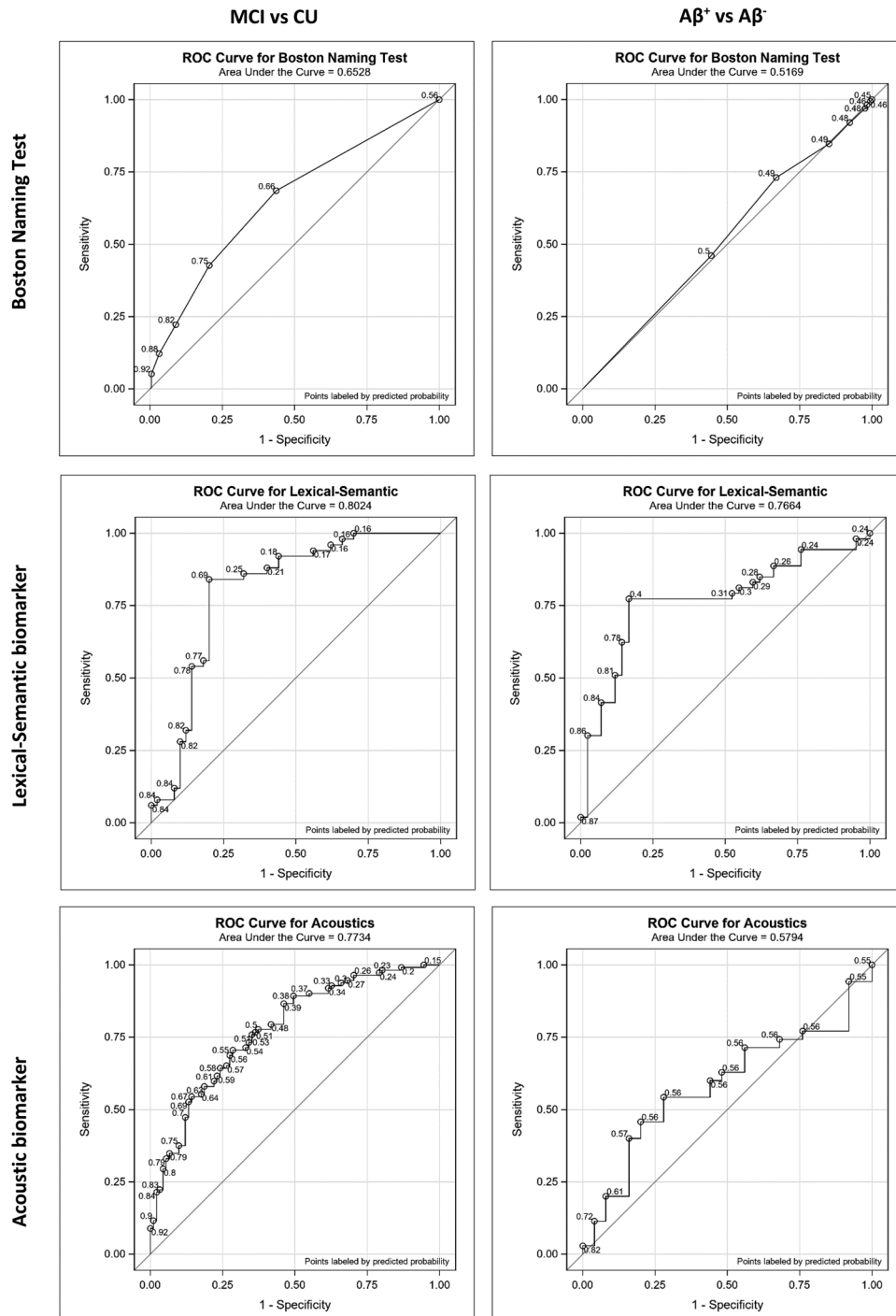
FIGURE 2 Comparisons of ROC analyses of derived digital voice biomarkers (acoustic and lexical-semantic) and BNT as measures of linguistic performance, by cognitive status (MCI vs. CU, left column) and amyloid status (Aβ positive vs. Aβ negative, right column). Aβ negative, amyloid beta negative; Aβ positive, amyloid beta positive; BNT, Boston Naming Test; CU, cognitively unimpaired; ROC, Receiver Operating Characteristic; MCI, mild cognitive impairment.

visual (SM-VIS) links showed significant negative correlations with lexical-semantic scores. Figure 4 shows the brain connectogram and domain-wise modularized correlation matrix representing brain connections and highlighting significant correlations between FNC and the lexical-semantic score. Of note, many regions in the SCN and CCN were highly activated and correlated with digital biomarker measures.

4 | DISCUSSION

This study found that, not only did digital voice biomarkers differ by cognitive status, but they were also accurate in detecting AD biomarker status. These derived voice biomarkers also tracked disease progression, measured by changes in the 2-year follow-up CDR-SOB score.
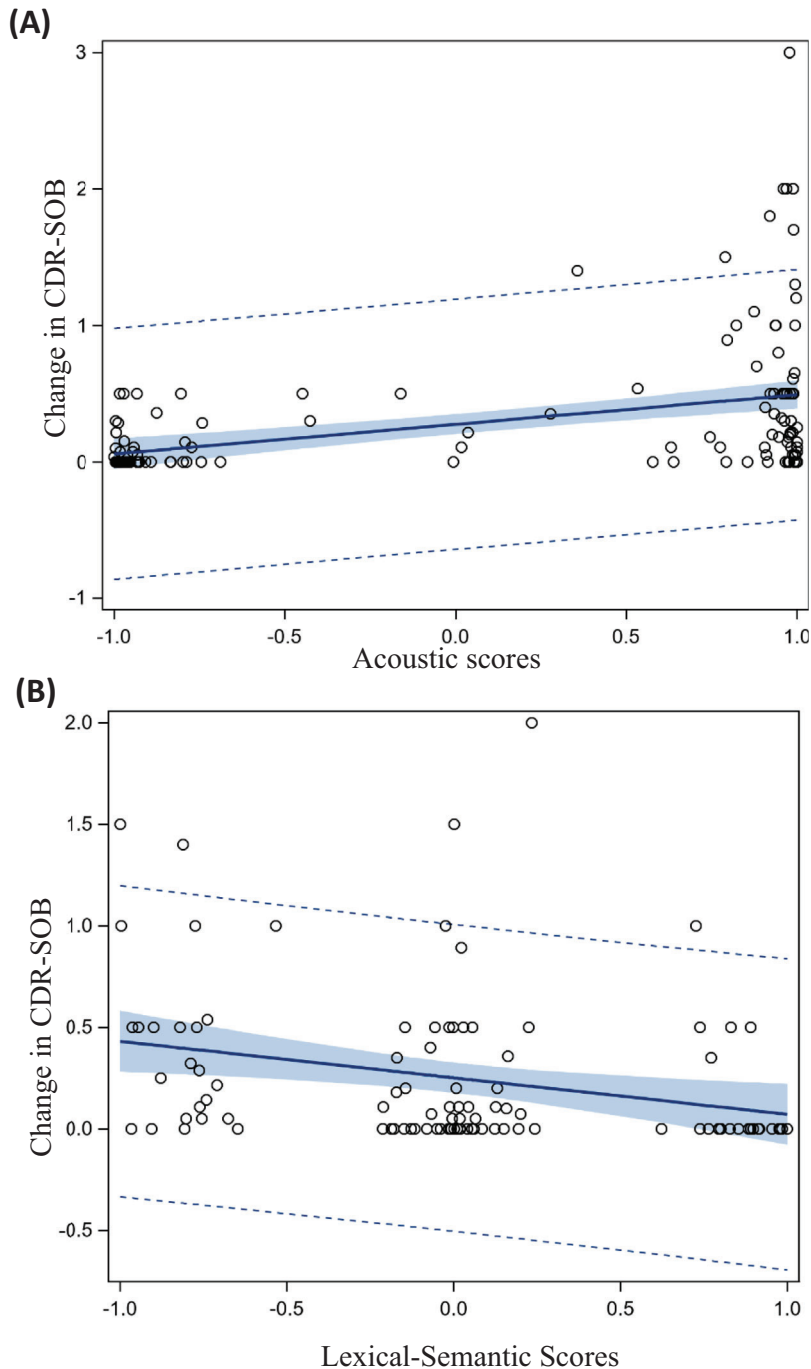
(A)

(B)

FIGURE 3 Association of baseline acoustic (panel A) and lexical-semantic (panel B) scores with the 2-year change of CDR-SOB (higher value indicates greater disease progression). X-axis is the derived digital biomarker score. Y-axis is the change in CDR-SOB over 2 years. Higher baseline acoustic score and lower lexical-semantic scores were associated with greater increases in CDR-SOB reflecting greater disease progression with greater cognitive or functional impairment at 2-year follow-up. CDR SOB, Clinical Dementia Rating - Sum of Boxes.

A recent report by Verfaillie et al. showed that in 63 non-demented individuals (19 with Aβ+ status), there were no significant associations between amyloid load and performance on conventional neuropsychological language tests, whereas fewer content words, abstract nouns, and syntactic complexity were associated ($p < 0.01$) with amyloid load.[43] All their participants were enrolled in the Subjective Cognitive Impairment Cohort (SciencE),[44] a prospective study of individuals with cognitive complaints but normal cognitive performance. In contrast, our CU participants did not have significant cognitive complaints and as such, the findings indicate that associations between language features and amyloid status can be detected even before changes in cognition are subjectively experienced. These findings are especially important for early detection of at-risk groups less likely to be referred for evaluations, but may benefit from interventions including clinical trial enrollment.

We used advances in NLP and artificial intelligence to detect lexical-semantic patterns that would otherwise be undetectable or labor intensive to identify. Semantic degradation occurs early in AD resulting in a reduction in the amount of specific content information, while changes in features such as syntax and grammar occur later in the disease.[45] Our lexical-semantic indices captured semantic, pragmatic, and discourse aspects of language that transcended traditional NLP

FIGURE 4 Significant associations were observed between the derived digital biomarkers (acoustic and lexical-semantic scores) and functional network connectivity of the brain. The brain connections highlighted in panel A demonstrate positive (red) or negative (blue) covariation with the derived digital biomarkers, whereas panel B displays the equivalent connectogram illustration between seven brain domains and their underlying brain regions. Interestingly, the significant links were predominantly focused on higher cognitive brain networks. Significant positive correlations ($p < 0.05$) of the digital biomarkers were found between several CC, CC-SM; strongest effect and CC-DM brain connections, whereas significant negative correlations ($p < 0.05$) were found between several SC-DM, SC-CC, and SC-SM brain connections. AUD, auditory; CB, cerebellar; CC, cognitive control; DM, default mode; SC, subcortical; SM, somatomotor; VIS, visual.
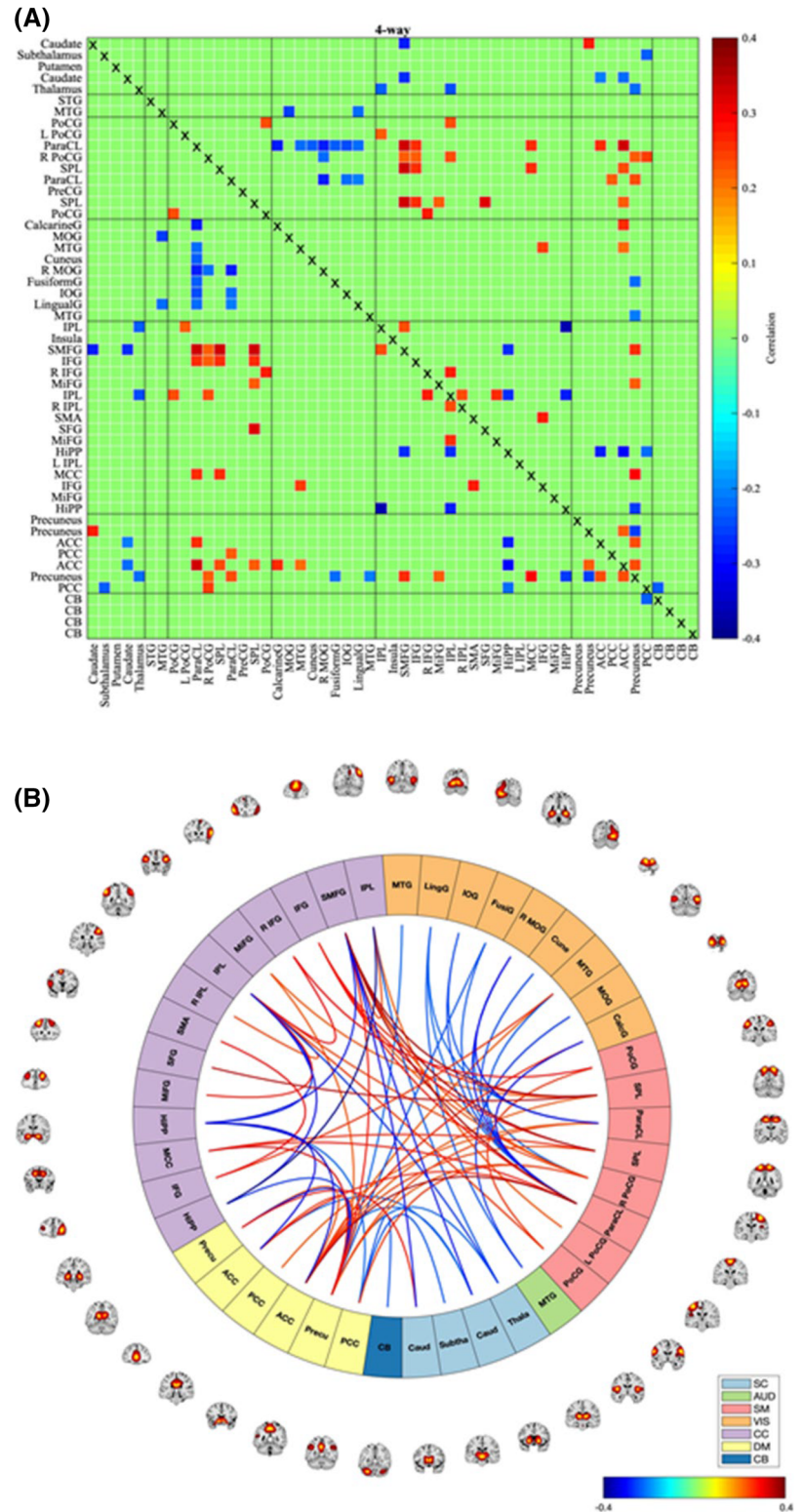
approaches which have used metrics such as text-level counts (tokens and sentences), grammatical categories (nouns and verbs), or syntactic structures (coordination, clausal modifiers, or complements). These indices were sensitive to amyloid positivity in both normal controls and MCI participants, had higher diagnostic performance for detecting amyloid status than measures of confrontation naming, and were

associated with disease progression. It is not surprising that AUC for individual tasks is low as a highly specific and sensitive categorization requires lengthy and detailed cognitive interviews and testing.

Neuropsychological measures of semantic processing such as timed category fluency and naming are frequently administered to assist in clinically identifying patients who have possible or probable AD.

However, negligible results have been reported on their utility in detecting amyloid burden.[46] A comprehensive and innovative feature of our study was the analysis of acoustic speech. Acoustics capture the essence of the expression and manner of connected speech which can provide critical contextual information and underlying neurological pathologies. While analysis of connected speech has been the focus of cognitive impairment detection, acoustic analysis has been utilized less often even though it is a viable and important research area.[47] Our findings indicated that acoustic scores were sensitive to cognitive status (CU or MCI), had higher diagnostic accuracy compared to a naming test, and were associated with changes in CDR-SOB. However, acoustic scores differed between MCI participants but not CU controls who were Aβ+ versus Aβ-, suggesting that acoustics may be more sensitive to amyloid status in prodromal versus preclinical AD.

The medial temporal lobe and temporo-parietal region, implicated in semantic processing, are reportedly susceptible to pathological change in early AD.[46] Many theories have been proposed for lexical-semantic representation in AD. The most relevant is the theory of parallel-distributed representation where a homogeneous network of equivalent neuronal units process every aspect of semantics.[48] Connectivity between anterior and posterior left superior temporal gyri (STG) correlated with lexical-semantic processing. In contrast, connectivity between STG and the middle temporal and inferior frontal gyri correlated with phonological processing. These connections identified for both types of processing are vulnerable to AD neuropathology.[49,50]

The neural representation of acoustic performance has not yet been clarified, but we find it beneficial to view it as being related to speech motor control. Here, speech production is based on adequate connectivity in the language-dominant hemisphere in the left supplementary motor area with dorsal frontal cortex/anterior insula and superior cerebellum (preparatory loop), while corticobulbar systems (motor cortex, thalamus, and putamen) participate in motor execution (executive loop).[51,52] These culminate in acoustic signals where the vocal organs disturb air molecules leading to the sounds we hear during articulations. Derangement in preparatory or executive loops may be detected in acoustic analysis and decomposing the components of the two loops can reveal neural disruption. In this study, we address this gap by including glottic waveform, derived from acoustic waves in detecting depression. Limited evidence suggests that glottal waveform features are altered in late stage AD.[53] Our study findings support the idea that derived lexical-semantic features using this modeling approach map to areas heavily involved in semantic abilities, and use of these tools in refining and advancing voice biomarker research will likely be successful.

4.1 | Strengths, limitations, and future directions

This study's strengths lie in the inclusion of key components such as AD CSF biomarker status, neuroimaging assessment, consideration of confounders, and more comprehensive linguistic acoustic approaches successful in detecting AD biomarker and cognitive status. There are some limitations which underscore the need for further research.

There was a potential for misclassification of MCI verus CU, especially with a 50% African American sample. We used the modified ADNI criteria that includes the MoCA for clinically diagnosing persons as CU or MCI. Previous reports have suggested that cutoff scores of ≤26 on MoCA may be unreliable and a lower cutoff may be indicated.[54,55] The MoCA is an imperfect tool, and we did not rely on a single test to classify participants. We believe that additional neuropsychological criteria beyond the MoCA strengthened the likelihood of accuracy.

We used the short (15-item) BNT which while sensitive to cognitive impairment, may not have been as sensitive as the 60-item version. Some features of the picture we used to capture connected speech may be dated and culturally dependent. Our goal was to select a detailed scene available in the public domain that had actions, colors, and objects which captured the richness of connected speech. This scene provided an advantage over the widely used Cookie Theft Picture[56] which is not public-domain, shows few activities and objects, and portrays stereotyped roles. Future pictures will likely need to span multiple racial, ethnic, and socioeconomic backgrounds while future research examines the sensitivity of digital biomarkers to cognitive functioning and differences based on sociocultural and demographic mediators. Finally, the inherent ML approaches may lead to models that are overfitting the data. We attempted at mitigating these by incorporating analytical approaches, but overfitting is still possible.

The time spent for voice acquisition was 6–8 min. Comparatively, a traditional neuropsychological battery for MCI categorization and longitudinal monitoring would require several hours. We believe this approach to be time efficient, given its diagnostic performance for detecting cognitive and amyloid status in under 10 min, and warrants further development and validation. However, the ease of collecting voice recording even without the acknowledgement of the individuals should also be considered and safeguards for privacy and prevention of bias and discrimination should be implemented as part of this area development. This is particularly true in underrepresented minorities.

4.2 | Conclusion

Preliminary findings suggest that digital voice biomarkers are not only able to detect cognitive impairment using brief audio recordings, but distinguish AD biomarker status and disease progression. These findings reveal the strong potential of advancing technologies combined with connected speech as digital biomarkers for AD clinical care and research participation, especially in preclinical and prodromal stages.

tension and Aging Research Program) team and the Emory Goizueta Alzheimer's Disease Research Center staff.

CONFLICT OF INTEREST

Ihab Hajjar, Maureen Okafor, Jinho D. Choi, Elliot Moore II, Anees Abrol, Vince Calhoun, and Felicia C. Goldstein report no conflicts. Author disclosures are available in the supporting information.

REFERENCES

1. Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on aging Alzheimer disease centers, 2005-2010. *J Neuropathol Exp Neurol*. 2012;71:266-273.
2. Fyffe DC, Mukherjee S, Barnes LL, Manly JJ, Bennett DA, Crane PK. Explaining differences in episodic memory performance among older african americans and whites: the roles of factors related to cognitive reserve and test bias. *J Int Neuropsychol Soc*. 2011;17:625-638.
3. Rabinovici GD, Gatsonis C, Apgar C, et al. Association of Amyloid positron emission tomography with subsequent change in clinical management among Medicare beneficiaries with mild cognitive impairment or dementia. *JAMA*. 2019;321:1286-1294.
4. Jack CR, Jr., Bennett DA, Blennow K, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement*. 2018;14:535-562.
5. Riley K, Snowdon DA, Desrosiers MF, Markesbery WR. Early life linguistic ability, late life cognitive function, and neuropathology: findings from the Nun Study. *Neurobiol Aging*. 2005;26(3):341-347.
6. Medina LD, Rodriguez-Agudelo Y, Geschwind DH, et al. Propositional density and apolipoprotein E genotype among persons at risk for familial Alzheimer's disease. *Dement Geriatr Cogn Disord*. 2011;32:188-192.
7. Stewart R. Mild cognitive impairment–the continuing challenge of its "real-world" detection and diagnosis. *Arch Med Res*. 2012;43:609-614.
8. 2022 Alzheimer's disease facts and figures. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2022;18(4):700789. https://doi.org/10.1002/alz.12638
9. Mitchell AJ, Meader N, Pentzek M. Clinical recognition of dementia and cognitive impairment in primary care: a meta-analysis of physician accuracy. *Acta Psychiatrica Scandinavica*. 2011;124:165-183.
10. Sabbagh MN, Boada M, Borson S, et al. Early detection of mild cognitive impairment (MCI) in primary care. *J Prev Alzheimers Dis*. 2020;7:165-170.
11. Khodabakhsh A, Kuscuoglu S, Demiroglu C. Detection of Alzheimer's disease using prosodic cues in conversational speech. 22nd Signal Processing and Communications Applications Conference (SIU). IEEE; 2014:1003-1006.
12. Meilan JJ, Martinez-Sanchez F, Carro J, Sanchez JA, Perez E. Acoustic markers associated with impairment in language processing in Alzheimer's Disease. *Span J Psychol*. 2012;15:487-494.
13. Jarrold W, Peintner B, Wilkins D, et al. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. Association for Computational Linguistics; 2014:27-37.
14. Beltrami D, Gagliardi G, Favretti RR, Ghidoni E, Tamburini F, Calza L. Speech analysis by natural language processing techniques: a possible tool for very early detection of cognitive decline? *Front Aging Neurosci*. 2018;10:369.
15. Toth L, Hoffmann I, Gosztolya G, et al. A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Curr Alzheimer Res*. 2018;15:130-138.
16. Grossman M, Payer, F, Onishi, K, et al. Constraints on the cerebral basis for semantic processing from neuroimaging studies of Alzheimer's disease. *J Neurol Neurosurg Psychiatry*. 1997;63(2):152-158.
17. deToledo-Morrell L, et al. MRI-derived entorhinal volume is a good predictor of conversion from MCI to AD. *Neurobiol Aging*. 2004;25:1197-1203.
18. Kearney E, Guenther FH. Articulating: the neural mechanisms of speech production. *Lang Cogn Neurosci*. 2019;34:1-16.
19. Montembeault M, Chapleau M, Jarret J, et al. Differential language network functional connectivity alterations in Alzheimer's disease and the semantic variant of primary progressive aphasia. *Cortex*. 2019;117:284-298.
20. Mascali D, DiNuzzo M, Serra L, et al. Disruption of semantic network in mild Alzheimer's disease revealed by resting-state fMRI. *Neuroscience*. 2018;371:38-48.
21. Whitney C, Kirk M, O'Sullivan J, Lambon Ralph MA, Jefferies E. The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cereb Cortex*. 2010;21:1066-1075.
22. Fu Z, Caprihan A, Chen J, et al. Altered static and dynamic functional network connectivity in Alzheimer's disease and subcortical ischemic vascular disease: shared and specific brain connectivity abnormalities. *Human Brain Mapping*. 2019;40:3203-3221.
23. Pistono A, Senoussi M, Guerrier L, et al. Language network connectivity increases in early Alzheimer's disease. *J Alzheimers Dis*. 2021;82:447-460.
24. Teipel SJ, Metzger CD, Brosseron F, et al. Multicenter resting state functional connectivity in prodromal and dementia stages of Alzheimer's disease. *J Alzheimers Dis*. 2018;64:801-813.
25. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Kokmen E, Tangelos EG. Aging, memory, and mild cognitive impairment. *Int Psychogeriatr*. 1997;9(Suppl 1):65-69.
26. Lezak MD, Howieson DB, Bigler ED, Tranel D. *Neuropsychological Assessment*. 5th ed. Oxford University Press; 2012.
27. Mcloughlin Bros, Juvenile Collection.. *The Circus procession*. McLoughlin Bro's; 1888.
28. Boschi V, Catricala E, Consonni M, Chesi C, Moro A, Cappa SF. Connected speech in neurodegenerative language disorders: a review. *Front Psychol*. 2017;8:269.
29. Geranmayeh F, Leech R, Wise RJ. Semantic retrieval during overt picture description: left anterior temporal or the parietal lobe? *Neuropsychologia*. 2015;76:125-135.
30. Ahmed S, Haigh AM, de Jager CA, Garrard P. Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*. 2013;136:3727-3737.
31. Garrard P, Forsyth R. Abnormal discourse in semantic dementia: a data-driven approach. *Neurocase*. 2010;16:520-528.
32. Eyben F, Scherer KR, Schuller BW, et al. The geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Trans Affect Comput*. 2016;7:190-202.
33. Eyben F, Wöllmer M, Schuller B, et al. Opensmile: the munich versatile and fast open-source audio feature extractor. Proceedings of the 18th ACM international conference on Multimedia. ACM Press; 2010:1459-1462.
34. Eyben F. *Real-time Speech and Music Classification by Large Audio Feature Space Extraction*. Springer Thesis, 1st ed. Springer; 2016. p. 1 online resource (XXXVIII, 298 pages 41 illustrations, 39 illustrations in color).
35. Choi JD, Li M, Goldstein F, Hajjar I. Meta-semantic representation for early detection of Alzheimer's disease. Proceedings of the First International Workshop on Designing Meaning Representations. Association for Computational Linguistics; 2019:82-91.
36. Hall MA. Correlation-Based Feature Selection for Machine Learning. *Department of Computer Science*. 2000;19.
37. Yu L, Liu H. Feature selection for high-dimensional data: a fast correlation-based filter solution. Proceedings of the Twentieth International Conference on International Conference on Machine Learning. AAAI Press; 2003:856-863.

38. Kohavi R, John GH. Wrappers for feature subset selection. *Artificial Intelligence.* 1997;97:273-324.

39. Schmidt MF, Storrs JM, Freeman KB, et al. A comparison of manual tracing and FreeSurfer for estimating hippocampal volume over the adult lifespan. *Hum Brain Mapp.* 2018;39:2500-2513.

40. Du Y, Fu Z, Sui J, et al. NeuroMark: an automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders. *Neuroimage Clin.* 2020;28:102375.

41. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med.* 2013;4:627-635.

42. Witten IH, Frank E, Hall MA. Chapter 10 - Introduction to Weka. In: Witten IH, Frank E, Hall MA, eds. *Data Mining: Practical Machine Learning Tools and Techniques.* 3rd ed. Morgan Kaufmann; 2011:403-406.

43. Verfaillie SCJ, Witteman J, Slot RER, et al. High amyloid burden is associated with fewer specific words during spontaneous speech in individuals with subjective cognitive decline. *Neuropsychologia.* 2019;131:184-192.

44. Slot RER, Verfaillie SCJ, Overbeek JM, et al. Subjective Cognitive Impairment Cohort (SCIENCe): study design and first results. *Alzheimers Res Ther.* 2018;10:1-13.

45. Ostrand R, Gunstad J. Using automatic assessment of speech production to predict current and future cognitive function in older adults. *J Geriatr Psychiatry Neurol.* 2021;34:357-369.

46. Vonk JM, Twait EL, Scholten RJ, Geerlings MI. Cross-sectional associations of amyloid burden with semantic cognition in older adults without dementia: a systematic review and meta-analysis. *Mech Ageing Dev.* 2020;192:111386.

47. Ammar RB, Ayed YB. Speech processing for early alzheimer disease diagnosis: machine learning based approach. IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA). 2018; IEEE:1-8.

48. McClelland JL, Rumelhart DE. Distributed memory and the representation of general and specific information. *J Exp Psychol Gen.* 1985;114:159-197.

49. Yu B, Williamson JR, Mundt JC, Quatieri TF. Speech-based automated cognitive impairment detection from remotely-collected cognitive test audio. *IEEE Access.* 2018;6:40494-40505.

50. Pakhomov SV, Jones DT, Knopman DS. Language networks associated with computerized semantic indices. *Neuroimage.* 2015;104: 125-137.

51. Maassen B, Van Lieshout P. Speech Motor Control : New Developments in Basic and Applied Research. Oxford University Press; 2010.

52. Guenther FH. *Neural Control of Speech.* The MIT Press; 2015.

53. Gómez-Vilda P, Rodellar-Biarge V, Nieto-Lluis V, et al. Phonation biomechanic analysis of Alzheimer's disease cases. *Neurocomputing.* 2015;167:83-93.

54. Goldstein FC, Ashley AV, Miller E, Alexeeva O, Zanders L, King V. Validity of the Montreal cognitive assessment as a screen for mild cognitive impairment and dementia in African Americans. *J Geriatr Psychiatry Neurol.* 2014;27:199-203.

55. Rossetti HC, Lacritz LH, Cullum CM, Weiner MF. Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. *Neurology.* 2011;77:1272-1275.

56. Berube S, Nonnemacher J, Demsky C, et al. Stealing cookies in the twenty-first century: measures of spoken narrative in healthy versus speakers with aphasia. *Am J Speech Lang Pathol.* 2019;28:321-329.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.