

AI-based Speech Error Detection to Differentiate Primary Progressive Aphasia Variants

*Jet M.J. Vonk¹

*Jiachen Lian²

Cheol J. Cho²

Giada Antonicelli¹

Zoe Ezzes^{3,4}

Lisa D. Wauters⁵

Willa Keegan-Rodewald¹

G. Lynn Kurteff¹

Diana Alejandra Rodriguez¹

Nina Dronkers⁶

Maya L. Henry⁵

Zachary A. Miller¹

Maria Luisa Mandelli¹

‡Gopala K. Anumanchipalli²

‡Maria Luisa Gorno-Tempini¹

*co-first authors

‡co-last authors

¹ Edward and Pearl Fein Memory and Aging Center, Department of Neurology, University of California San Francisco, San Francisco, CA

² Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, CA, USA

³ School of Speech, Language, And Hearing Sciences, San Diego State University, San Diego, CA, USA

⁴ Department of Cognitive Science, University of California San Diego, San Diego, CA, USA

⁵ Department of Speech, Language, and Hearing Sciences, The University of Texas at Austin, Austin, TX, USA

⁶ Department of Psychology, University of California Berkeley, Berkeley, CA, USA

Corresponding author:

Jet M.J. Vonk, PhD PhD

jet.vonk@ucsf.edu

Abstract

Background: Artificial Intelligence (AI) based approaches to speech analysis have the potential to assist with objective speech error analysis in aphasia but off-the shelf tools often fail to detect speech errors due to prioritizing “fluent transcription.” Speech production errors (dysfluencies) are hallmark diagnostic features of the nonfluent (nfvPPA) and logopenic (lvPPA) variants of primary progressive aphasia, yet they can be challenging to detect and characterize even by expert clinicians. This study aimed to evaluate whether the novel automated lightweight Scalable Speech Dysfluency Modeling system (SSDM-L), specifically designed to detect dysfluencies, could accurately distinguish PPA variants using voice recordings of individuals reading a brief passage.

Method: Participants included a total of 104 individuals, 40 with nfvPPA, 40 with lvPPA (matched on disease severity), and 24 healthy controls who read aloud the ‘Grandfather Passage’ as part of a widely used motor speech evaluation (MSE). We automatically extracted ten speech error (dysfluency) variables using SSDM-L, including insertions, replacements, and deletions at both phoneme- and word-levels, and phoneme-level prolongations and repetitions. Group differences were assessed via ANCOVAs controlling for age, education, and disease severity (MMSE, CDR sum-of-boxes). To test clinical relevance, we performed correlation analyses with MSE ratings provided by experienced speech-language pathologists (i.e., gold standard) within the nfvPPA group. Classification performance was assessed by training random forest and XGBoost machine-learning models including 5-fold cross-validation.

Results: All individuals read the entire passage in less than five minutes. SSDM-L detected eight of the ten predefined dysfluency features at sufficient frequency to include them in subsequent analyses. All eight features distinguished PPA from controls ($p < .006$). Individuals with nfvPPA made more errors than the lvPPA group on every feature (all $p < .023$). Each feature showed a moderate positive correlation with a global MSE apraxia/dysarthria score ($r = .31-.56$; $p < .001-.053$). Together, the eight features were able to classify nfvPPA versus lvPPA at AUC = .806 (random forest) and AUC = .776 (XGBoost).

Discussion: AI-based automated speech error analysis accurately distinguished nfvPPA and lvPPA variants using a brief reading task. This quick error-sensitive scalable AI system has the potential of providing a practical tool to aid diagnosis in aphasia and motor speech disorders.

Keywords: language, automated speech analysis, frontotemporal dementia, motor speech, apraxia of speech, speech biomarkers

Introduction

Advances in speech technologies, such as OpenAI Whisper, Google Speech-to-Text, and others, have significantly improved transcription through automated speech recognition (ASR) and subsequent speech analysis of large-scale spoken language datasets.^{1,2} By providing automated transcriptions, these tools facilitate extracting linguistic features such as lexical diversity and syntactic complexity in individuals who produce relatively fluent speech and have shown promise as cognitive markers in dementias such as Alzheimer's disease.³⁻⁵ However, applying these technologies to the dysfluent speech of individuals who present with specific language and articulation difficulties is far more challenging. Speech produced by such individuals often contains word- and phoneme-level errors (dysfluencies) that are critical for diagnostic purposes but are poorly captured by standard ASR systems, creating limitations in the use of current tools for these populations.⁶ Off-the-shelf ASR systems such as Whisper are biased toward fluent speech transcription and often "correct away" clinically meaningful speech errors like substitutions, repetitions, deletions, and pauses.⁷ These dysfluencies are essential for early detection and monitoring of conditions such as primary progressive aphasia (PPA), a neurodegenerative disease that selectively affects speech and language functions.⁸

Among the three canonical variants of PPA, the nonfluent/agrammatic variant (nfvPPA), with motor speech and grammatical deficits,⁹ and the logopenic variant (lvPPA), with phonological deficits,^{10,11} are the two that show clear dysfluencies in spontaneous speech. Distinguishing nfvPPA and lvPPA is particularly challenging,¹² as nfvPPA involves apraxia of speech and phonetic errors (i.e., focused on how a sound is produced), while lvPPA produces phonemic errors (i.e., focused on what sound is intended).^{13,14} Distinguishing between these errors requires specialized expertise from speech-language pathologists (SLPs), and, even then, inter-rater reliability among SLPs often varies.¹⁵ Advanced AI techniques tailored to detect a range of error patterns within speech samples across patients could improve diagnostic accuracy.^{16,17} Differential diagnosis between these two variants is crucial since lvPPA is most often caused by Alzheimer's disease pathology,^{11,18,19} while nfvPPA is related to frontotemporal lobar degeneration,¹⁹ each with distinct implications for prognosis, treatment, clinical trial eligibility, and disease monitoring.

To address this limitation, we developed a lightweight Scalable Speech Dysfluency Modeling system (SSDM-L), a novel AI framework designed to accurately transcribe and analyze speech errors.²⁰ This system was created to effectively characterize speech errors within a given speech sample against a known ground-truth target script. SSDM-L is explainable and specifically designed to be sensitive to speech dysfluencies. SSDM-L comprises three sub-modules: a) Dysfluent-WFST44: a high-performance phonetic decoder that transcribes the phonetic sequence of the patient's speech sample;²¹ b) Neural-LCS45: an alignment module that maps the phonetic sequence from (a) to the reference phoneme sequence (e.g., the Grandfather script);²² and c) Speech Error Analysis Module: a component that uses the alignment produced in (b) to categorize specific types of speech errors.²⁰ SSDM-L is capable of detecting ten distinct types of dysfluencies, including insertions, repetitions, deletions, and replacements at both phoneme and word levels, thus capturing a wide range of error patterns underlying diverse speech-language disorders.

This study evaluated whether the SSDM-L AI system could reliably capture speech production errors in a large cohort of well-characterized individuals with nvPPA and lvPPA. We investigated whether these dysfluencies obtained from voice recording of patients with PPA reading a short 2-minute paragraph would: (1) identify group-level differences between PPA variants; (2) correlate with expert SLP severity ratings for motor speech impairment as a proxy for speech output deficits; and (3) accurately classify individuals into nvPPA and lvPPA. We hypothesized that dysfluencies identified by SSDM-L would be more prevalent in nvPPA than in lvPPA, reflecting the fluency deficits typical of the nvPPA variant. The overall goal of this study is to demonstrate that advanced, automated AI speech modeling techniques hold the potential to aid and scale diagnostic precision of progressive speech and language disorders.

Method

Participants

This study included 104 participants: 40 diagnosed with nonfluent/agrammatic variant PPA (nvPPA), 40 with logopenic variant PPA (lvPPA), and 24 cognitively healthy controls. All participants were evaluated at the University of California, San Francisco (UCSF) Edward and Pearl Fein Memory and Aging Center (MAC) PPA programs. Individuals seen at the MAC undergo detailed clinical, neuropsychological, and imaging assessments. Healthy controls were recruited primarily through the Brain Aging Network for Cognitive Health (BrANCH) cohort and were cognitively assessed to confirm the absence of neurological or psychiatric disorders.

We identified all cases in the UCSF Fein MAC PPA program database with a clinical diagnosis of nvPPA or lvPPA, following established diagnostic criteria,⁸ who had completed research visits since 2001, had an available recorded Grandfather Passage reading task,^{23,24} and were native English speakers. To focus on early-stage differentiation between PPA variants, we excluded individuals who did not complete at least nearly half of the passage (4/9 sentences, 45% of words), had Mini-Mental State Examination (MMSE) scores below 17, or had Clinical Dementia Rating (CDR) sum of boxes scores greater than 9. These thresholds were selected to exclude individuals with moderate or severe dementia, generally considered as CDR sum of boxes > 9 and MMSE < 18. We used a slightly lower cut-off on MMSE to account for the language-dependent nature of the MMSE, which often underestimates cognitive function in PPA.²⁵ Data from each participant's earliest available research visit that included administration of the Grandfather Passage were used for analysis. To minimize confounding factors and account for differences in PPA subgroup sizes, we performed frequency matching to match PPA groups on age, sex/gender, education, MMSE, CDR, and group size (all $p > .05$), while also ensuring the control group did not differ from either PPA group on age, sex/gender, and education (all $p > .05$). A detailed speech and language evaluation is completed with every individual participant as previously described.²⁶ All participants or caregivers provided informed consent following procedures aligned with the Declaration of Helsinki, and the study was approved by the UCSF Institutional Review Board (IRB # 10-03946).

Neuroimaging analysis of group-level atrophy patterns

A group-level neuroimaging analysis was conducted to confirm typical patterns of atrophy for each variant. MRI T1 images were acquired for 92 subjects (39 nvPPA, 38 lvPPA, 15 controls) with sequences, previously described, on either 1.5T (n=7),²⁶ 3T (n=81),²⁷ or 4T (n=4).²⁸ After a visual quality check, one participant from the nvPPA group was excluded for excessive motion. All the structural images underwent standard preprocessing for thickness analysis using the Computational Anatomy Toolbox 12 (CAT12) version 12.9 (dbm.neuro.uni-jena.de/cat) in Statistical Parametric Mapping 12 software version 7771 (fil.ion.ucl.ac.uk/spm/software/spm12) as detailed in Mandelli et al. (2023).²⁹

Whole-brain analyses of between-group differences in cortical thickness (nvPPA vs. controls; lvPPA vs. controls) were estimated using independent sample t-tests covaried for age and sex. Alpha levels were set at a peak-level threshold of $p < .05$ family-wise error (FWE) corrected. Visualization of group atrophy patterns was performed using CAT12.

Motor Speech Evaluation including the Grandfather Passage

Each participant underwent a Motor Speech Evaluation (MSE)³⁰ during their initial clinical language assessment, which was recorded on video. As part of this evaluation, participants read the Grandfather Passage aloud; the audio from this recording was extracted for use in the present study's speech analysis.

The MSE is a clinical assessment designed to detect and diagnose motor speech disorders, specifically AOS and dysarthria,³¹ and is widely used in clinical research and diagnosis, including in individuals with PPA.^{9,13,32} The test protocol consists of nine subtests that evaluate various aspects of speech production, including automatic speech production (i.e., counting from 1-20), sustained phonation, sequential motion rates (i.e., repeating /pʌpʌpʌ/; /tʌtʌtʌ/; /kʌkʌkʌ/), alternating motion rates (i.e., repeating /pʌtʌkʌ/), multisyllabic word repetition, production of words of increasing length, sentence repetition, and reading aloud of the Grandfather Passage. Each participant receives a 0-7 severity rating for both AOS and dysarthria, which is informed by characteristic features observed across the tasks (0: Within normal limits, 1: Slight, 2: Mild, 3: Mild to moderate, 4: Moderate, 5: Moderate to severe, 6: Severe, 7: Profound). Notably, the MSE is designed to detect motor speech and not phonological errors, thus SLPs are likely to only rate errors that are considered motoric in nature.

The Grandfather Passage is a brief, phonetically balanced paragraph used to elicit connected speech in a controlled context through reading aloud. Originally developed by Charles Van Riper in 1963²³ and adapted by Darley, Aronson, & Brown (1975)²⁴ for use in speech and language assessment, the passage includes all phonemes in the English language and contains complex phonotactic combinations (e.g., *frock*, *zest*), as well as syntactic and semantic variety.³³ These characteristics make it particularly useful for eliciting speech production deficits. The brevity of the passage (typically 1-3 minutes) makes it efficient to administer while still providing a rich sample for analysis. Because the spoken text is predetermined, the Grandfather Passage provides an optimal context for validating the SSDM-L technique using phoneme-level forced alignment.

AI-based automated speech transcription and feature extraction

SSDM-L comprises a phoneme transcription module, phoneme alignment module, phoneme-level dysfluency detector, as well as word-level transcription, alignment, and dysfluency detection modules. The phoneme transcription component captures the actual produced phonemes, adapted from Dysfluent-WFST,²¹ a state-of-the-art phonetic transcriber. With the Grandfather Passage serving as the reference text, SSDM-L applies the Neural Longest Common Subsequence (NLCS)²² algorithm to align the uttered speech with the reference text at the phoneme level. Phonetic dysfluencies are detected as follows: if the uttered phoneme matches the reference phoneme, no dysfluency is reported; otherwise, insertions, omissions, substitutions, or repetitions are identified. Each dysfluency is reported at most once for each reference phoneme. An analogous set of four features is extracted at the word level. Additionally, SSDM-L integrates Unconstrained Dysfluency Modeling (UDM),³⁴ a state-of-the-art dysfluent phoneme aligner, to detect irregular phoneme durations, particularly for vowels. If the duration exceeds a predefined threshold, phoneme prolongation is reported. UDM also detects irregular intra-word pauses, which are counted at the word level. Thus, SSDM-L extracts raw counts of ten features in total (Table 1): phoneme insertion, phoneme replacement, phoneme repetition, phoneme deletion, phoneme prolongation, word insertion, word replacement, word repetition, word deletion, and intra-word pauses (i.e., pauses within a word). Of note, none of these features are consistently or accurately captured by off-the-shelf ASR tools. We computed normalized values for each feature by dividing the dysfluency counts by the total number of corresponding phonemes or words, respectively.

Statistical analysis

Sample characteristics and distributions of extracted speech features were examined using descriptive statistics, general linear models, chi-square tests, and Pearson correlation coefficients. Because participants did not always complete the full standardized passage, we normalized phoneme- and word-level error counts by dividing each by the total number of phonemes or words read, respectively. Group differences in individual features were assessed using analysis of variance (ANOVA; unadjusted) and analysis of covariance (ANCOVA), adjusting for age, education, and disease severity (z-scored composite of CDR Sum of Boxes and MMSE³⁵). Post-hoc pairwise comparisons between diagnostic groups were corrected for multiple testing using the false discovery rate (FDR).

To assess whether the AI-extracted dysfluencies correlate with expert SLP ratings of motor speech output difficulties, we examined their relationship with clinical ratings of the overall MSE AOS and dysarthria ratings in the nvPPA group. Pearson correlation coefficients were computed between each feature and the clinician-rated severity scores for AOS and dysarthria, as well as a composite “global” MSE score defined as the sum of the AOS and dysarthria ratings (higher scores indicating more severe impairment). While AOS and dysarthria are different motor speech disorders with different neurological substrates (and the MSE ratings would usually not be combined by a practicing SLP), we aim at correlating the AI-based scores with a general, SLP-derived score reflecting the overall severity of motor speech output errors that patients make, regardless of their pathophysiological origin. In other words, the purpose of our global MSE score is not to make claims about the substrates

of AOS and dysarthria, but to serve as a proxy of overall severity of motor speech impairment that is easily comparable to the SSDM-L system's estimates of severity.

To evaluate the utility of the extracted features in distinguishing between diagnostic groups and to ensure robustness across modeling approaches, we trained two machine learning classifiers: random forest and extreme gradient boosting (XGBoost). Random forest was included as a widely adopted, interpretable machine learning method that reduces overfitting through ensemble learning and has demonstrated strong performance in prior classification studies using automated speech data in dementia research.³⁶ XGBoost was selected for its ability to handle small sample sizes, deliver stable performance with minimal tuning, and its similarly proven success in classification tasks involving speech-based markers of neurodegenerative disease.^{37,38} The data were split into 65% training and 35% testing sets, with model performance assessed using five-fold cross-validation within the training set. Area under the receiver operating characteristic curve (AUC) was used as an evaluation metric. All extracted speech error (dysfluency) features used in the ANCOVA analysis were included in the classification models without prior feature selection. All analyses and visualizations were conducted in R (version 4.4.2) using a range of packages as listed in the analysis code, which is available at <https://github.com/jmjonk>.

Data availability

UCSF Fein MAC ethics approval does not permit depositing anonymized study data in a public repository. Data can be obtained on reasonable request via the UCSF Fein MAC Resource Request form (<http://memory.ucsf.edu/resources/data>). Access is granted to named investigators under ethical regulations governing reuse of sensitive data, and all requests require a Material Transfer Agreement (MTA) processed through UCSF (see <https://icd.ucsf.edu/material-transfer-and-data-agreements>). Requests for commercial use will not be approved.

Results

Participant and feature characteristics

Sample characteristics for demographics, disease severity, standard cognitive and speech-language tests, and MSE scores across groups are summarized in Table 2. As the cognitively healthy control, lvPPA, and nvPPA groups were matched on age, sex/gender, and education, there were no significant differences among groups for these variables (all $p > .05$). There were also no differences between groups on race/ethnicity or handedness (all $p > .05$). The cognitively healthy control group had higher MMSE scores and lower CDR Sum of Boxes scores than both PPA groups, while the lvPPA and nvPPA groups were matched on these variables and thus did not differ from each other in disease severity scores. The two PPA variants showed speech and language characteristics compatible with their diagnosis, with lvPPA showing greater repetition and confrontation naming deficits, and the nvPPA showing greater fluency and motor speech difficulties.⁸ The lvPPA group scored 0 on MSE severity ratings for both AOS and dysarthria, except for one individual with an AOS score of 1, compatible with the MSE's focus on motor speech errors, whereby other types of speech production errors are not captured. In the nvPPA group, all individuals had at least one

score higher than 0 on either AOS or dysarthria ratings; 5/40 had dysarthria without AOS, 9/40 had AOS without dysarthria, and 26/40 had co-occurring AOS and dysarthria. For 16/40 AOS was more severe (>1 difference in score) than dysarthria, for 9/40 dysarthria was more severe (>1 difference in score) than AOS, and 15/40 had relatively equal scores (max 1 difference in score) on AOS and dysarthria. Scores on the AOS (max 4) and dysarthria (max 5) ratings in the nvPPA group were in the mild range,¹³ consistent with the early disease stage of the selected individuals.

The average time to complete reading the Grandfather Passage aloud was 48.8 seconds (SD = 5.7, range = 34.6-60.8) for cognitively healthy participants, 76.6 seconds (SD = 30.4, range = 41.5-168.6) for the lvPPA group, and 107.0 seconds (SD = 48.8, range = 51.5-285.0) for the nvPPA group. SSDM-L automatically extracted counts of ten phoneme-level and word-level speech errors during reading aloud the Grandfather Passage: phoneme insertion, phoneme replacement, phoneme repetition, phoneme deletion, phoneme prolongation, word insertion, word replacement, word repetition, word deletion, and intra-word pauses. Inspection of feature distributions across all participants showed that phoneme-level errors were common, each occurring 94-104 times. Word-level features were on average slightly less frequent but still widely present, with word insertion occurring 93 times, word replacement 97 times, and word deletion 69 times. In contrast, word repetition and intra-word pauses were rare, occurring only 18 and 24 times, respectively, across all participants. Because word repetition and intra-word pause errors were infrequently captured by the SSDM-L system, they were excluded from further group comparisons and classification analyses due to low prevalence.

Neuroimaging atrophy patterns

Participants in the nvPPA group exhibited left-sided atrophy in the inferior frontal gyrus, supplementary motor area, and precentral gyrus (Figure 1A). Participants in the lvPPA group exhibited left-sided atrophy in the temporal-parietal junction and temporal lobe, particularly the superior, middle, and inferior temporal gyri (Figure 1B).

Group differences

Differences in error counts on the eight remaining phoneme- and word-level features between diagnostic groups is visualized in the boxplots in Figure 2. Each feature shows a similar visual pattern, in which the nvPPA group made the most errors, followed by the lvPPA group, followed by almost no errors in the control group. AN(C)OVA models to statistically assess group differences in features showed similar results unadjusted or adjusted for covariates (i.e., age, education, disease severity); this report focuses primarily on the adjusted models. All features differed across the three diagnostic groups (all $p < .006$). Pairwise comparisons (FDR-corrected) showed the nvPPA group made more errors than the lvPPA group on every feature (all $p < .023$; Table 3). While unadjusted, the lvPPA group made significantly more errors than the control group on nearly every feature except word deletion (Fig 2); no group differences between lvPPA and controls remained significant after adjustment for covariates and FDR correction in the pairwise comparisons (all $p > .084$). In contrast, the nvPPA group robustly made more errors than the control group on all features

in unadjusted and adjusted models, with all $p < .024$ after adjustment for covariates and FDR correction, except for word deletion at $p = .055$.

Validation with expert-rated Motor Speech Evaluation

The relationships between automated speech features and SLP MSE ratings within the nvPPA group are visualized in Figure 3, which includes a correlation matrix, scatterplots, and histograms of all variables. Across the eight speech error (dysfluency) features, moderate positive correlations were observed with the global MSE score (sum of AOS and dysarthria ratings), ranging from $r = .30$ to $.58$ ($p < .001$ -.064).

Machine learning classification

Figure 4 displays receiver operating characteristic (ROC) curves for classifying individuals with nvPPA and lvPPA using automated speech error (dysfluency) features. A random forest machine learning model trained on the eight extracted features achieved an AUC of .806 [95% CI: .636–.976], indicating good classification performance based on speech errors alone. When age, education, and disease severity were added, model performance remained similar with an AUC of .805 [.631–.978]. An XGBoost classifier yielded similar results, with an AUC of .776 [.590–.961] using the speech features alone, and an improved AUC of .831 [.673–.989] when demographic and clinical variables were included. These converging results across two machine learning classifiers support the robustness of the automated features in distinguishing between PPA variants.

Discussion

This study demonstrated that AI-based automated analysis of speech errors (dysfluencies) extracted using the novel SSDM-L system, can detect PPA from controls and differentiate between individuals with nvPPA and lvPPA. Using voice recordings from a brief reading task, the system captured dysfluencies that reflected known differences in speech production error profiles between variants. All features showed group differences, with the nvPPA group making the most errors, followed by lvPPA, with minimal errors in cognitively healthy controls. These automated features also showed correlations with expert SLP-rated scores of motor speech impairment, confirming that an AI-based classification of speech dysfluencies aligns with clinical judgment as a preliminary proof-of-concept analysis. Finally, classification performance was good based on features alone, and further improved when demographic and disease severity variables were added to the model. Taken together, these findings suggest that the SSDM-L AI system provides a promising solution to automatically detect speech errors in connected speech.

Our findings provide a crucial advance to a growing body of work using automated speech analysis to study neurodegenerative disease, including PPA.^{32,39-44} A handful of studies have successfully applied natural language processing and audio signal processing techniques to characterize broad linguistic and/or acoustic patterns in PPA, such as reduced lexical diversity, simplified syntax, and semantic impairments.^{32,39,41} Our findings address a key limitation of previous off-the-shelf ASR tools, which prioritize fluent, intelligible output and systematically normalize dysfluencies, making them ill-suited for detecting speech errors that are crucial to the diagnosis and monitoring of PPA, and other

neurological diseases. Dysfluency detection remains one of the most persistent challenges in ASR, particularly in clinical populations where speech is degraded or non-standard.⁶ Our study provides evidence that the SSDM-L system, combining phoneme-level alignment with rule-based detection, enables the capture of clinically meaningful speech disruptions automatically and efficiently. As automated speech tools gain traction in both clinical and research settings, developing systems that preserve diagnostically informative signals (rather than correcting them away) will be critical for advancing their clinical utility, particularly in disorders such as PPA where speech production deficits are a defining feature.

The machine learning classification results further support the potential of automated speech error (dysfluency) analysis for assisting in differential diagnosis of PPA. Using SSDM-L features, the models achieved good discrimination between nfvPPA and lvPPA, demonstrating that fine-grained phoneme- and word-level errors carry meaningful diagnostic signal. Models remained similar or only slightly improved after adding demographic and clinical variables, likely since groups were matched on these variables. These findings are consistent with prior studies applying machine learning to speech and language features in PPA and other neurodegenerative syndromes, which have similarly shown promising accuracy in distinguishing subtypes.^{43,45}

The prominence of motor speech impairments in individuals with nfvPPA has been well-documented.⁹ AOS is one of the two core diagnostic criteria for nfvPPA and is frequently the earliest and most prominent symptom;⁸ dysarthria is also often observed in individuals with nfvPPA.^{46,47} Consistently, our participants with nfvPPA showed significantly higher scores in AOS and dysarthria severity than those with lvPPA on the motor speech evaluation carried out by SLP experts. Determining the presence of motor speech impairments typically requires expert perceptual judgment by an SLP—a process that is time-consuming, resource-intensive, and often limited by the scarcity of clinicians with specialized expertise in dementia clinics.^{48,49} The automated speech features identified in this study were consistently correlated with expert SLP-rated global MSE scores in the nfvPPA group, supporting the interpretation that higher feature counts are at least partially related to underlying motor speech impairments. These findings reinforce the construct validity of the SSDM-L-derived metrics and suggest that automated analysis may offer a scalable complement to clinician-based assessments of speech errors in PPA. Further studies are necessary to determine whether the AI system can match the precision of clinicians in identifying specific error types (e.g., differentiating dysarthria and AOS) in individual cases.

A core strength of our approach is that it is grounded in expert clinical knowledge, incorporating input from trained SLPs and aligning with real-world observations of speech behavior in PPA. Using the Grandfather Passage allowed us to control lexical content and isolate dysfluency patterns, avoiding confounds related to language formulation and vocabulary access, while the use of forced alignment enabled precise detection of phonemic deviations that reflect underlying motor or phonological impairments. Nonetheless, several limitations should be noted. Although our sample sizes are comparable to those in other studies of PPA,^{39,43,50,51} they remain relatively small for achieving optimal statistical power and broad generalizability. Nonetheless, the consistent group differences observed across all features suggest that the automated measures are

robust and sensitive to meaningful speech impairments, even in a modestly-sized sample. In addition, all participants were native English speakers, which may restrict the applicability of the SSDM-L system to individuals with different linguistic backgrounds. Additionally, the current analysis focused on a single speech elicitation task. Although the Grandfather Passage offers important advantages for controlled error analysis, future research should examine the generalizability of these findings to spontaneous speech tasks such as picture description⁵² or personal narratives.⁵³ Finally, a limitation of the AI based system is that similar coding conflicts that exist within human-expert annotators can manifest within AI-based automated error analysis. For example, a single phoneme replacement error can equivalently be represented as one deletion followed by one insertion. While such distinctions can often be resolved by trained SLPs, current AI-based error analyzers (including SSDM-L) are not sufficiently precise to consistently disambiguate these mechanistically equivalent error modes. Future refinements of the AI system that integrate more SLP-guided error classifications are likely to further improve both the accuracy and consistency of automated error analysis beyond what we have demonstrated in the present study.

Our study provides strong evidence of the potential of AI methods to transform how speech errors are identified and monitored in neurodegenerative diseases, particularly in under-resourced settings where expertise in speech-language pathology is limited or unavailable. The integration of engineered speech analytics into routine care could bring the field closer to more equitable, efficient, and precise models of diagnosis and disease tracking in neurodegenerative conditions.

Acknowledgments

This work was supported by the National Institutes of Health (NIH) under the following grants: R01NS050915, K24DC015544, R01AG075775, P01AG019724 (M.L. Gorno-Tempini) and R00AG066934, RF1AG091509 (J.M.J. Vonk). We are grateful to our participants and their families for participating in this research.

Table 1. Features extracted by SSDM-L with examples

Level	Feature Type	Example	Notes
Word	Deletion	Please [-call] Stella	Omission of an intended lexical item. A target word is missing from the original reference text, reflecting breakdowns in reading, lexical planning, and/or execution.
Word	Insertion	Please call [what's the name again] Stella	Insertion of non-target lexical material. Extra words appear in the original reference text that are not aligned to the reference, often associated with fillers or self-repairs.
Word	Replacement	Please [dial] Stella	Substitution of an intended lexical item with a different word. May reflect reading, lexical retrieval, and/or semantic substitution errors.
Word	Repetition	[Please please please] call Stella	Repetition of lexical units, where multiple produced words align to a single word in the original reference text.
Word/phoneme	Pause	please call s[pause]tella	Silent interval inserted within or between words, where no speech segment aligns to the original reference text for an extended duration, often reflecting hesitations or breakdowns in speech planning or execution.
Phoneme	Deletion	pl'iz [-z] k'ɔl st'ɛlə	Omission of an intended speech segment. A reference phoneme in the original text has no aligned produced phoneme, often reflecting phonological or motor planning difficulty.
Phoneme	Insertion	pl'iz k'ɔl [ɛ]st'ɛlə	Epenthesis of an additional speech segment. Extra produced phoneme(s) appear without a reference counterpart in the original text, potentially due to compensatory articulation.
Phoneme	Replacement	pl'iz [t]'ɔl st'ɛlə	Substitution of one phoneme for another. This category also captures speech sound distortions that manifest as sub-phonological changes in motor speech output (e.g., devoicing, slurring).
Phoneme	Repetition	pl'iz [k..k..k]'ɔl st'ɛlə	Repeated production of a speech segment, where multiple produced phonemes align to a single phoneme from the original reference text.
Phoneme	Prolongation	pl'iz[extend duration]z k'ɔl st'ɛlə	Abnormally extended duration of a speech segment, capturing temporal distortion without segment substitution.

Table 2. Participant characteristics

	N	Cognitively healthy <i>n</i> = 24 ¹	lvPPA <i>n</i> = 40 ¹	nfvPPA <i>n</i> = 40 ¹	<i>p</i> (all groups)²
Age	104	69.1 (6.0)	66.0 (7.0)	68.5 (7.1)	0.081
Sex/gender	104				0.4
Female		15 (63%)	20 (50%)	26 (65%)	
Male		9 (38%)	20 (50%)	14 (35%)	
Race/Ethnicity	104				0.6
White		21 (88%)	37 (93%)	34 (85%)	
Black/African American		1 (4.2%)	0 (0%)	1 (2.5%)	
Hispanic/Latinx		1 (4.2%)	1 (2.5%)	1 (2.5%)	
Asian		1 (4.2%)	0 (0%)	2 (5.0%)	
Unknown		0 (0%)	2 (5.0%)	2 (5.0%)	
Handedness	100				0.071
Left		6 (25%)	7 (18%)	2 (5.3%)	
Right		18 (75%)	30 (79%)	36 (95%)	
Ambidextrous		0 (0%)	1 (2.6%)	0 (0%)	
Mini Mental State Examination (MMSE)	89	28.94 (1.64)	25.57 (2.75)	26.70 (2.26)	<0.001
Clinical Dementia Rating (CDR)	96	0.00 (0.00)	2.25 (1.31)	1.76 (1.68)	<0.001
Education	100	17.50 (1.56)	16.45 (2.91)	16.34 (3.29)	0.14
Forward digit span (<i>n</i>)	72	7.38 (1.15)	4.73 (1.04)	5.40 (1.16)	<0.001
Backward digit span (<i>n</i>)	82	5.50 (1.32)	3.42 (0.81)	3.34 (1.11)	<0.001
Phonemic fluency (<i>n</i>)	82	17.6 (5.2)	9.0 (4.9)	6.0 (4.5)	<0.001
Animal fluency (<i>n</i>)	91	24 (4)	12 (5)	11 (7)	<0.001
Brief Boston naming test (<i>n</i>)	82	14.88 (0.50)	10.23 (3.17)	13.26 (1.92)	<0.001
Peabody Picture Vocabulary Test (<i>n</i>)	65	16.00 (NA)	14.39 (2.12)	14.79 (1.39)	0.4
Single word reading (<i>n</i>)	46	49 (2)	39 (8)	41 (7)	0.2
WAB total repetition (%)	72	NA (NA)	79 (10)	90 (9)	<0.001
WAB spontaneous speech	69	NA (NA)	17.17 (1.99)	15.94 (2.49)	0.049
MSE Apraxia rating	100	0.00 (0.00)	0.03 (0.17)	2.18 (1.30)	<0.001
MSE Dysarthria rating	100	0.00 (0.00)	0.00 (0.00)	1.90 (1.58)	<0.001

¹ Mean (SD); *n* (%)

² Kruskal-Wallis rank sum test; Pearson's Chi-squared test; Fisher's exact test

Note. Continuous variables: mean (SD, range), categorical variables: count (percentage); lvPPA = logopenic Primary Progressive Aphasia (PPA); nfvPPA = non-fluent PPA; MSE = Motor Speech Evaluation; MMSE was primarily missing for Controls (*n*=7); Apraxia of Speech, Dysarthria, and MSE Global scores were missing for *n*=4 lvPPA, no missingness in nfvPPA

Table 3. Group differences in phoneme- and word-level errors including pairwise comparisons

Feature	ANCOVA <i>F</i> (df)	<i>p</i>	Pairwise Comparison	Mean Difference	<i>p</i> (FDR-corrected)
<i>Phoneme insertion</i> (per phoneme)	5.37 (2, 94)	.006	Control - lvPPA	-.007	.511
			Control - nfvPPA	-.026	.023
			lvPPA - nfvPPA	-.019	.023
<i>Phoneme replacement</i> (per phoneme)	14.83 (2, 94)	<.001	Control - lvPPA	-.008	.084
			Control - nfvPPA	-.021	<.001
			lvPPA - nfvPPA	-.012	.001
<i>Phoneme repetition</i> (per phoneme)	6.57 (2, 94)	.002	Control - lvPPA	-.002	.524
			Control - nfvPPA	-.009	.010
			lvPPA - nfvPPA	-.007	.010
<i>Phoneme deletion</i> (per phoneme)	19.35 (2, 94)	<.001	Control - lvPPA	-.006	.781
			Control - nfvPPA	-.085	<.001
			lvPPA - nfvPPA	-.079	<.001
<i>Phoneme prolongation</i> (per phoneme)	5.86 (2, 94)	.004	Control - lvPPA	-.001	.914
			Control - nfvPPA	-.027	.024
			lvPPA - nfvPPA	-.026	.013
<i>Word insertion</i> (per word)	6.32 (2, 94)	.003	Control - lvPPA	-.015	.313
			Control - nfvPPA	-.042	.008
			lvPPA - nfvPPA	-.027	.023
<i>Word replacement</i> (per word)	10.74 (2, 94)	<.001	Control - lvPPA	.003	.937
			Control - nfvPPA	-.096	.003
			lvPPA - nfvPPA	-.099	<.001
<i>Word deletion</i> (per word)	8.13 (2, 94)	.001	Control - lvPPA	.016	.392
			Control - nfvPPA	-.035	.055
			lvPPA - nfvPPA	-.050	.001

Note. lvPPA = logopenic Primary Progressive Aphasia (PPA), nfvPPA = non-fluent PPA; ANCOVA = Analysis of Covariance; df = degrees of freedom; FDR = False Discovery Rate

Figure 1. Brain atrophy patterns

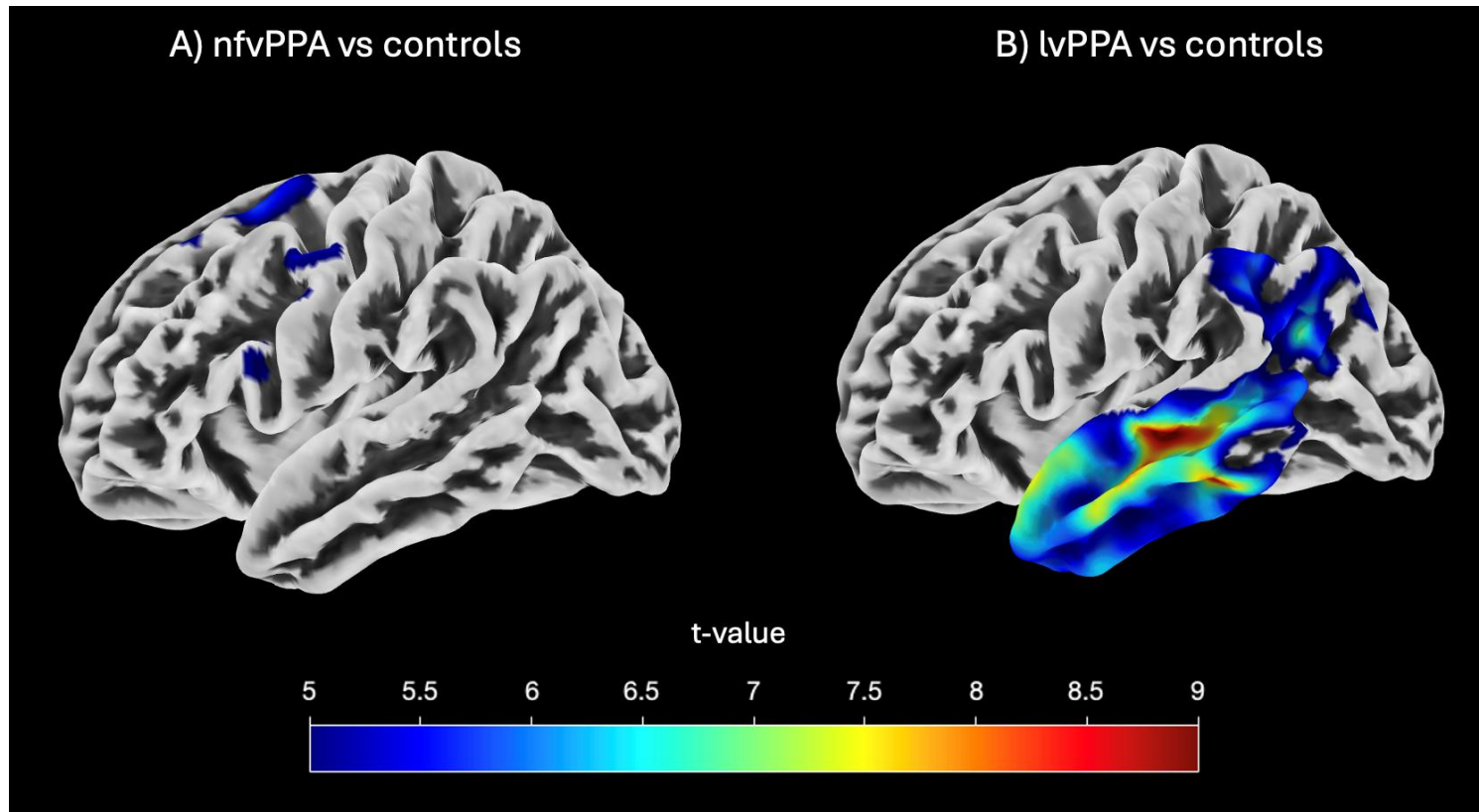
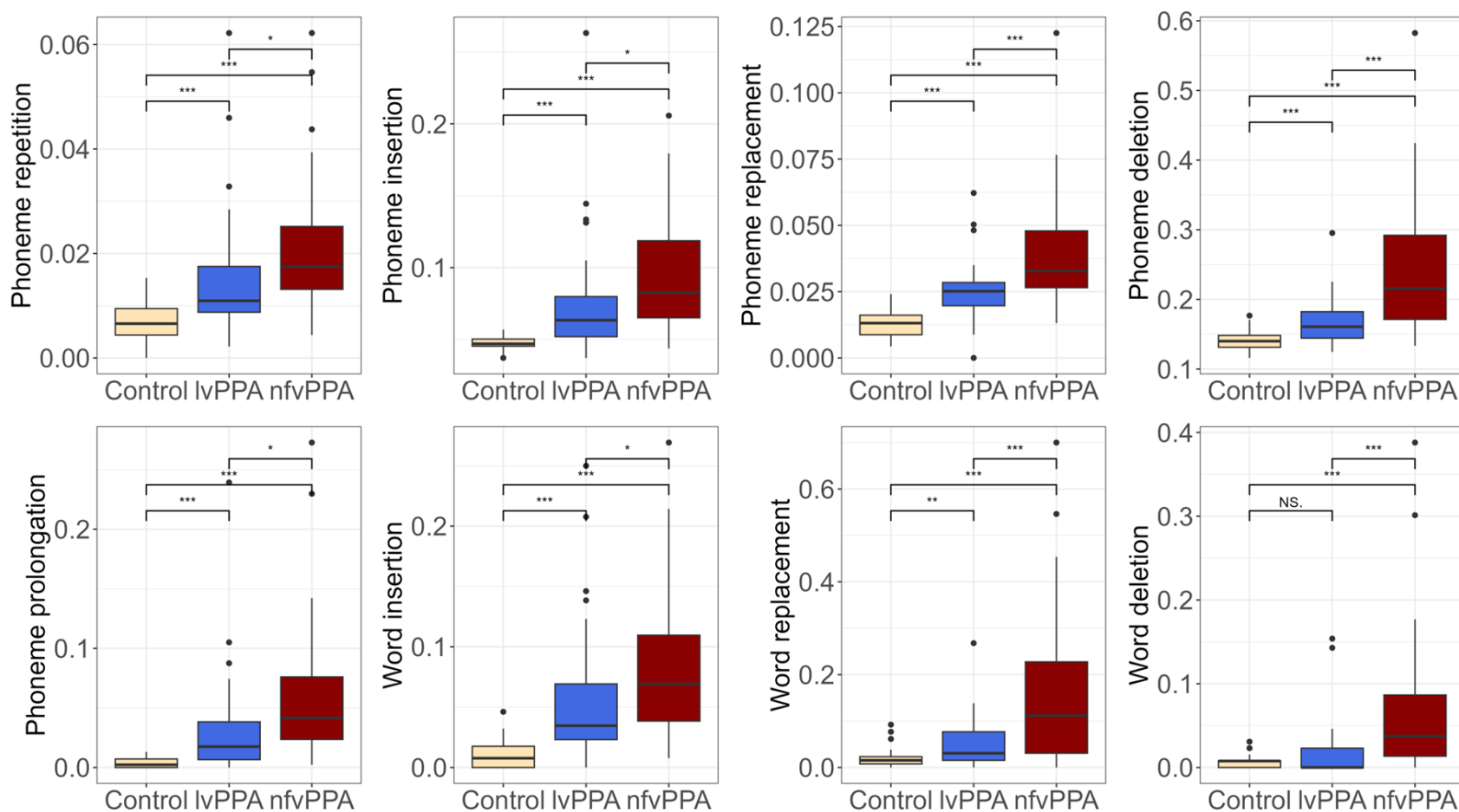


Figure 2. Normalized[†] counts of phoneme- and word-level errors across diagnostic groups (Control n=24; lvPPA n=40; nvfPPA n=40)



Note. [†]Scores were normalized by dividing phoneme- and word-level error counts by the total number of phonemes or words read, respectively, to account for incomplete passage completion; Visualized significance used t-tests (i.e., unadjusted for covariates), *p<.05, **p<.01, ***p<.001, NS=non-significant; lvPPA = logopenic variant primary progressive aphasia (PPA); nvfPPA = nonfluent variant PPA;

Figure 3. Scatterplot (lower triangle) and correlation matrix (upper triangle) of the three Motor Speech Evaluation (MSE) scores and the eight speech error (dysfluency) metrics, including histograms (diagonal); .p < .1, *p < .05, **p < .01, ***p < .001

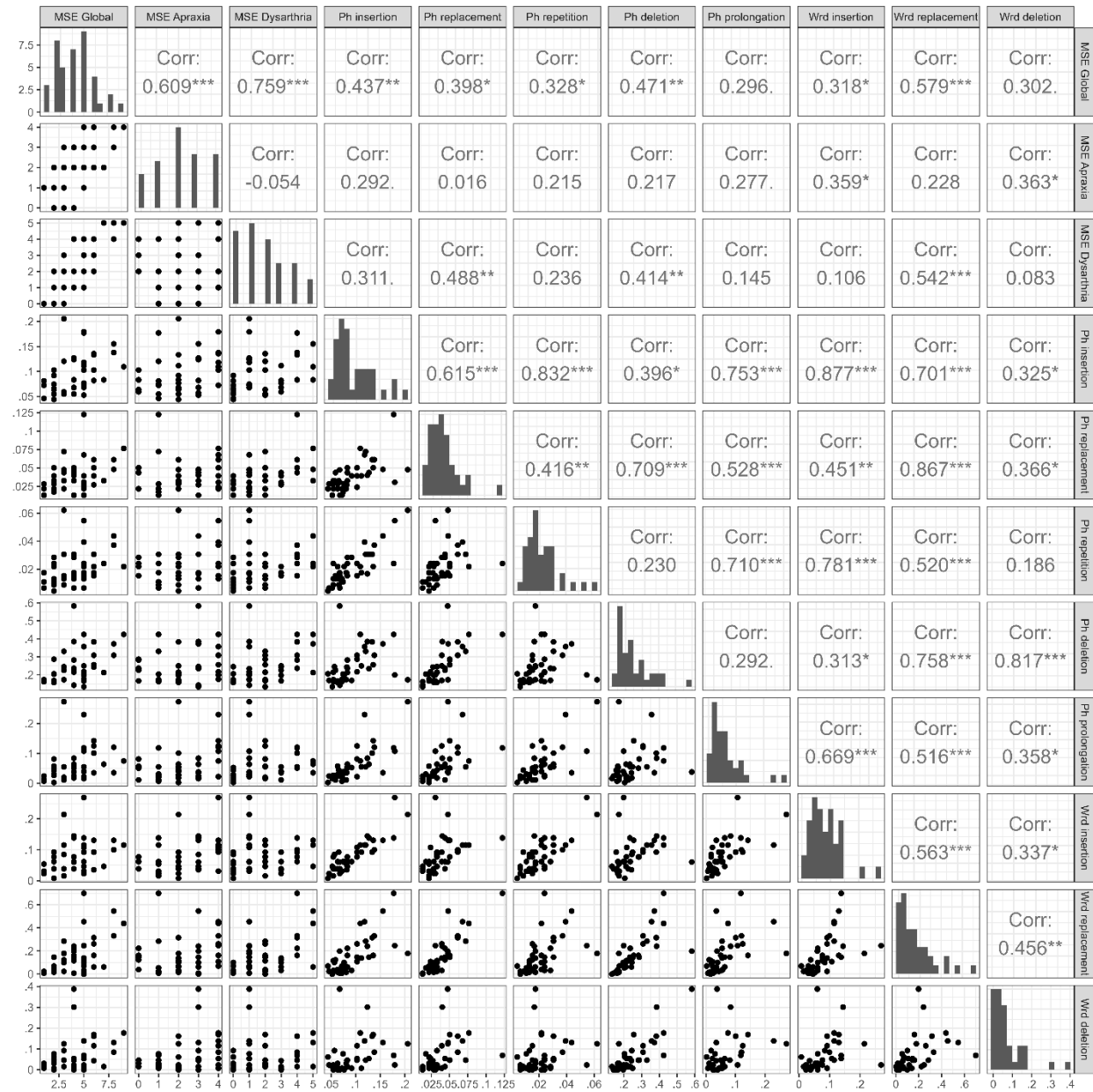
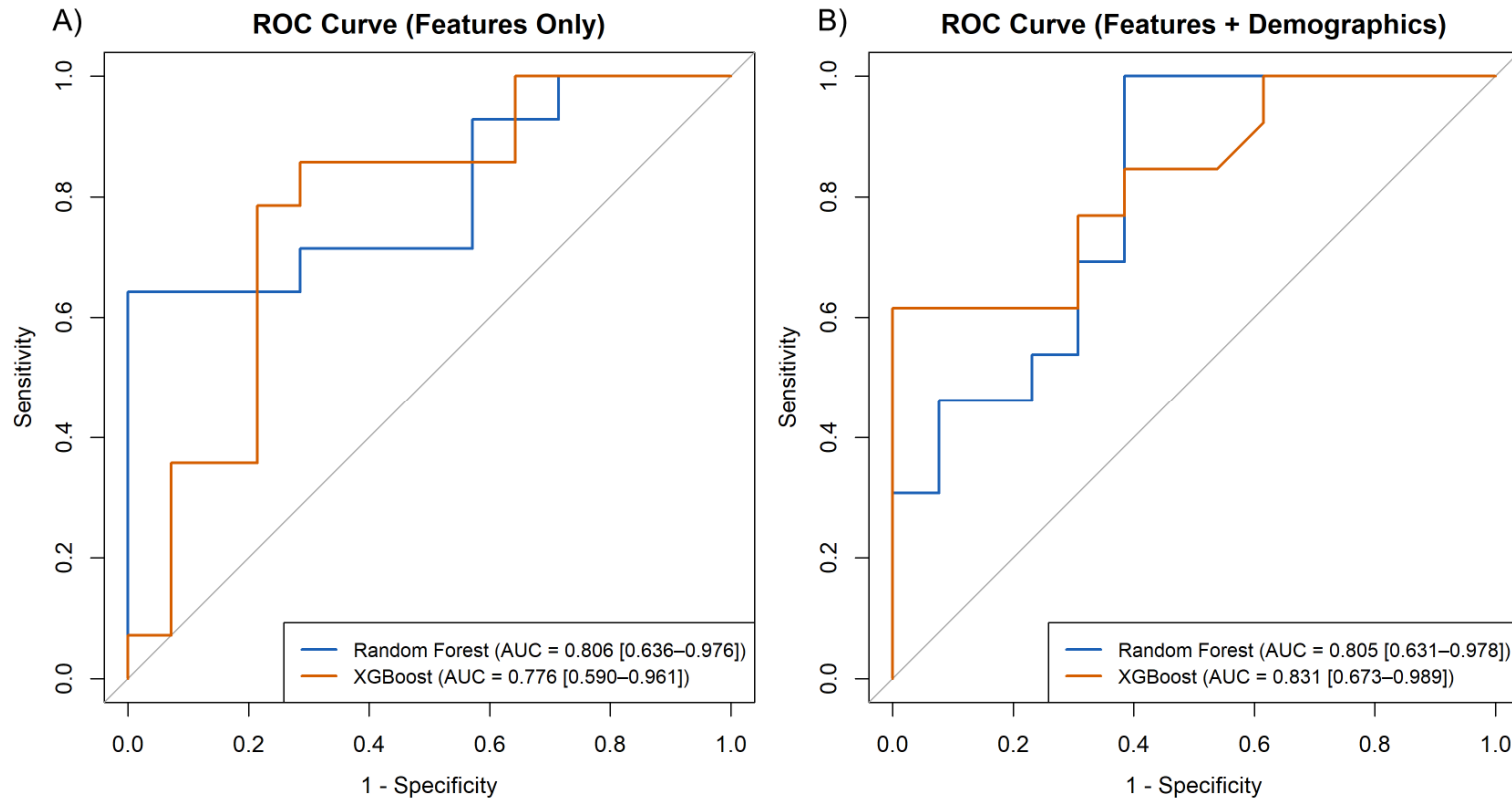


Figure 4. Receiver operating characteristic (ROC) curves for classifying nvPPA versus lvPPA



References

1. Schultz, B.G., Tarigoppula, V.S.A., Noffs, G., Rojas, S., van der Walt, A., Grayden, D.B., & Vogel, A.P. (2021). Automatic speech recognition in neurodegenerative disease. *International Journal of Speech Technology*, 24(3), 771-779.
2. Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Whisper: Robust Speech Recognition via Large-Scale Weak Supervision*. Retrieved from <https://github.com/openai/whisper>
3. Tang, L., Zhang, Z., Feng, F., Yang, L.-Z., & Li, H. (2023). Explainable alzheimer's disease detection using linguistic features from automatic speech recognition. *Dementia and geriatric cognitive disorders*, 52(4), 240-248.
4. Vivekanand, C.V., Nayagam, S., & Rajeev, S. (2025). *Integrating Acoustic and Linguistic Feature Extraction for Enhanced Dementia Diagnosis*. Paper presented at the 2025 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS).
5. Mueller, K.D., Hermann, B., Mecollari, J., & Turkstra, L.S. (2018). Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *J Clin Exp Neuropsychol*, 40(9), 917-939. doi:10.1080/13803395.2018.1446513
6. Mujtaba, D., Mahapatra, N.R., Arney, M., Yaruss, J.S., Gerlach-Houck, H., Herring, C., & Bin, J. (2024). Lost in transcription: Identifying and quantifying the accuracy biases of automatic speech recognition systems against disfluent speech. *arXiv preprint arXiv:2405.06150*.
7. Farzana, S., Deshpande, A., & Parde, N. (2022). *How you say it matters: Measuring the impact of verbal disfluency tags on automated dementia detection*. Paper presented at the Proceedings of the 21st workshop on biomedical language processing.
8. Gorno-Tempini, M.L., Hillis, A.E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S.F., Ogar, J.M., Rohrer, J.D., Black, S., Boeve, B.F., Manes, F., Dronkers, N.F., Vandenberghe, R., Rascovsky, K., Patterson, K., Miller, B.L., Knopman, D.S., Hodges, J.R., Mesulam, M.M., & Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, 76(11), 1006-1014.
9. Ogar, J.M., Dronkers, N.F., Brambati, S.M., Miller, B.L., & Gorno-Tempini, M.L. (2007). Progressive nonfluent aphasia and its characteristic motor speech deficits. *Alzheimer Disease & Associated Disorders*, 21(4), S23-S30.
10. Gorno-Tempini, M.L., Brambati, S.M., Ginex, V., Ogar, J., Dronkers, N.F., Marcone, A., Perani, D., Garibotto, V., Cappa, S.F., & Miller, B.L. (2008). The logopenic/phonological variant of primary progressive aphasia. *Neurology*, 71(16), 1227-1234.
11. Conca, F., Esposito, V., Giusto, G., Cappa, S.F., & Catricalà, E. (2022). Characterization of the logopenic variant of Primary Progressive Aphasia: A systematic review and meta-analysis. *Ageing Res Rev*, 82, 101760. doi:10.1016/j.arr.2022.101760

12. Antonicelli, G., Morin, B.T., Bogley, R., Newbury, R., Ezzes, Z., Ramkrishnan, S., Miller, Z., Mandelli, M. L., Tee, B. L., Martín García, A., Gorno-Tempini, M.L., Vonk, J.M.J. Differentiating non-fluent and logopenic primary progressive aphasia via automated acoustic speech analysis. *under review*.
13. Haley, K.L., Jacks, A., Jarrett, J., Ray, T., Cunningham, K.T., Gorno-Tempini, M.L., & Henry, M.L. (2021). Speech metrics and samples that differentiate between nonfluent/agrammatic and logopenic variants of primary progressive aphasia. *Journal of Speech, Language, and Hearing Research*, 64(3), 754-775.
14. Johnson, J.C.S., Jiang, J., Bond, R.L., Benhamou, E., Requena - Komuro, M.C., Russell, L.L., Greaves, C., Nelson, A., Sivasathiseelan, H., Marshall, C.R., Volkmer, A.P., Rohrer, J.D., Warren, J.D., & Hardy, C.J.D. (2020). Impaired phonemic discrimination in logopenic variant primary progressive aphasia. *Annals of Clinical and Translational Neurology*, 7(7), 1252-1257.
15. Duncan, E.S., Donovan, N.J., & Sajjadi, S.A. (2020). Clinical assessment of characteristics of apraxia of speech in primary progressive aphasia. *American Journal of Speech-Language Pathology*, 29(1S), 485-497.
16. Tetzloff, K.A., Wiepert, D., Botha, H., Duffy, J.R., Clark, H.M., Whitwell, J.L., Josephs, K.A., & Utianski, R.L. (2024). Automatic speech recognition in primary progressive apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 67(9), 2964-2976.
17. Gallée, J., & Volkmer, A. (2021). A window into functional communication: Leveraging naturalistic speech samples in primary progressive aphasia. *Perspectives of the ASHA Special Interest Groups*, 6(4), 704-713.
18. Giannini, L.A., Irwin, D.J., McMillan, C.T., Ash, S., Rascovsky, K., Wolk, D.A., Van Deerlin, V.M., Lee, E.B., Trojanowski, J.Q., & Grossman, M. (2017). Clinical marker for Alzheimer disease pathology in logopenic primary progressive aphasia. *Neurology*, 88(24), 2276-2284.
19. Spinelli, E.G., Mandelli, M.L., Miller, Z.A., Santos - Santos, M.A., Wilson, S.M., Agosta, F., Grinberg, L.T., Huang, E.J., Trojanowski, J.Q., & Meyer, M. (2017). Typical and atypical pathology in primary progressive aphasia variants. *Annals of Neurology*, 81(3), 430-443.
20. Lian, J., Zhou, X., Guo, C., Ye, Z., Ezzes, Z., Vonk, J.M.J., Morin, B., Baquirin, D., Miller, Z., & Tempini, M.L.G. (2025). Automatic Detection of Articulatory-Based Disfluencies in Primary Progressive Aphasia. *IEEE Journal of Selected Topics in Signal Processing*, 19(5), 810-826. doi:10.1109/JSTSP.2025.3579972
21. Guo, C., Lian, J., Zhou, X., Zhang, J., Li, S., Ye, Z., Park, H.J., Das, A., Ezzes, Z., Vonk, J.M.J., Morin, B.T., Bogley, R., Wauters, L., Miller, Z., Gorno-Tempini, M.L., & Anumanchipalli, G. (2025). Dysfluent WFST: A framework for zero-shot speech dysfluency transcription and detection. *Proc. Interspeech 2025*, 2205-2209. doi:10.21437/Interspeech.2025-2446
22. Ye, Z., Lian, J., Zhou, X., Zhang, J., Li, H., Li, S., Guo, C., Das, A., Park, P., Ezzes, Z., Vonk, J.M.J., Morin, B., Bogley, R., Wauters, L., Miller, Z., Gorno-Tempini, M., & Anumanchipalli, G. (2025). Seamless Dysfluent Speech Text Alignment for

- Disordered Speech Analysis. *Proc. Interspeech 2025*, 1848-1852.
doi:10.21437/Interspeech.2025-2496
23. Van Riper, C. (1963). *Speech correction* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
24. Darley, F.L., Aronson, A.E., & Brown, J.R. (1975). *Motor speech disorders* (3 ed.). Philadelphia, PA: W.B. Saunders Company.
25. Osher, J.E., Wicklund, A.H., Rademaker, A., Johnson, N., & Weintraub, S. (2008). The mini-mental state examination in behavioral variant frontotemporal dementia and primary progressive aphasia. *American Journal of Alzheimer's Disease & Other Dementias*®, 22(6), 468-473.
26. Gorno-Tempini, M.L., Dronkers, N.F., Rankin, K.P., Ogar, J.M., Phengrasamy, L., Rosen, H.J., Johnson, J.K., Weiner, M.W., & Miller, B.L. (2004). Cognition and anatomy in three variants of primary progressive aphasia. *Annals of Neurology*, 55(3), 335-346.
27. Mandelli, M.L., Caverzasi, E., Binney, R.J., Henry, M.L., Lobach, I., Block, N., Amirbekian, B., Dronkers, N., Miller, B.L., & Henry, R.G. (2014). Frontal white matter tracts sustaining speech production in primary progressive aphasia. *Journal of Neuroscience*, 34(29), 9754-9767.
28. Zhang, Y., Schuff, N., Ching, C., Tosun, D., Zhan, W., Nezamzadeh, M., Rosen, H.J., Kramer, J.H., Gorno-Tempini, M.L., & Miller, B.L. (2011). Joint Assessment of Structural, Perfusion, and Diffusion MRI in Alzheimer's Disease and Frontotemporal Dementia. *International journal of Alzheimer's disease*, 2011(1), 546871.
29. Mandelli, M.L., Lorca - Puls, D.L., Lukic, S., Montembeault, M., Gajardo - Vidal, A., Licata, A., Scheffler, A., Battistella, G., Grasso, S.M., & Bogley, R. (2023). Network anatomy in logopenic variant of primary progressive aphasia. *Human brain mapping*, 44(11), 4390-4406.
30. Wertz, R.T., LaPointe, L.L., & Rosenbek, J.C. (1984). *Apraxia of speech in adults: The disorder and its management*. Orlando, FL: Grune & Stratton.
31. Duffy, J.R. (2019). *Motor speech disorders: Substrates, differential diagnosis, and management* (4th ed.). Maryland Heights, MO: Mosby.
32. Ash, S., Nevler, N., Irwin, D.J., Shellikeri, S., Rascovsky, K., Shaw, L., Lee, E.B., Trojanowski, J.Q., & Grossman, M. (2023). Apraxia of Speech in the Spontaneous Speech of Nonfluent/Agrammatic Primary Progressive Aphasia. *Journal of Alzheimer's Disease Reports*, 7(1), 589-604.
33. Reilly, J., & Fisher, J.L. (2012). Sherlock Holmes and the strange case of the missing attribution: A historical note on "The Grandfather Passage".
34. Lian, J., Feng, C., Farooqi, N., Li, S., Kashyap, A., Cho, C.J., Wu, P., Netzorg, R., Li, T., & Anumanchipalli, G.K. (2023). *Unconstrained dysfluency modeling for dysfluent speech transcription and detection*. Paper presented at the 2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU).
35. Vonk, J.M.J., Jonkers, R., Hubbard, H.I., Gorno-Tempini, M.L., Brickman, A.M., & Obler, L.K. (2019). Semantic and lexical features of words dissimilarly affected by

- non-fluent, logopenic, and semantic primary progressive aphasia. *Journal of the International Neuropsychological Society*, 25(10), 1011-1022.
36. Hason, L., & Krishnan, S. (2022). Spontaneous speech feature analysis for alzheimer's disease screening using a random forest classifier. *Frontiers in Digital Health*, 4, 901419.
37. Tripathi, T., & Kumar, R. (2024). Speech-based detection of multi-class Alzheimer's disease classification using machine learning. *International Journal of Data Science and Analytics*, 18(1), 83-96.
38. Qian, C., Di, J., & Li, J. (2022). *Detecting mild cognitive impairment in Alzheimer's disease using speech acoustics only: a two-stage deep metric learning approach*. Paper presented at the 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).
39. Nevler, N., Ash, S., Irwin, D.J., Liberman, M., & Grossman, M. (2019). Validated automatic speech biomarkers in primary progressive aphasia. *Annals of Clinical and Translational Neurology*, 6(1), 4-14.
40. Nevler, N., Ash, S., Cho, S., Shellikeri, S., Parjane, N., Irwin, D.J., Liberman, M.Y., & Grossman, M. (2020). A longitudinal study of automated analysis of acoustic speech markers in FTD and PPA: Biomarkers (non - neuroimaging)/Longitudinal change over time. *Alzheimer's & Dementia*, 16, e045315.
41. García, A.M., Welch, A.E., Mandelli, M.L., Henry, M.L., Lukic, S., Prioris, M.J.T., Deleon, J., Ratnasiri, B.M., Puls, D.L.L., & Miller, B.L. (2022). Automated detection of speech timing alterations in autopsy-confirmed non-fluent/agrammatic variant primary progressive aphasia. *Neurology*.
42. Josephy-Hernandez, S., Rezaii, N., Jones, A., Loyer, E., Hochberg, D., Quimby, M., Wong, B., & Dickerson, B.C. (2023). Automated analysis of written language in the three variants of primary progressive aphasia. *Brain communications*, 5(4), fcad202.
43. Rezaii, N., Hochberg, D., Quimby, M., Wong, B., Brickhouse, M., Touroutoglou, A., Dickerson, B.C., & Wolff, P. (2024). Artificial intelligence classifies primary progressive aphasia from connected speech. *Brain*, 147(9), 3070-3082.
44. Ash, S., Evans, E., O'Shea, J., Powers, J., Boller, A., Weinberg, D., Haley, J., McMillan, C., Irwin, D.J., & Rascovsky, K. (2013). Differentiating primary progressive aphasias in a brief sample of connected speech. *Neurology*, 81(4), 329-336.
45. Themistocleous, C., Ficek, B., Webster, K., den Ouden, D.-B., Hillis, A.E., & Tsapkini, K. (2021). Automatic subtyping of individuals with primary progressive aphasia. *Journal of Alzheimer's Disease*, 79(3), 1185-1194.
46. Duffy, J.R., Strand, E.A., & Josephs, K.A. (2014). Motor speech disorders associated with primary progressive aphasia. *Aphasiology*, 28(8-9), 1004-1017.
47. Lorca-Puls, D.L., Gajardo-Vidal, A., Mandelli, M.L., Illán-Gala, I., Ezzes, Z., Wauters, L.D., Battistella, G., Bogley, R., Ratnasiri, B., & Licata, A.E. (2024). Neural basis of speech and grammar symptoms in non-fluent variant primary progressive aphasia spectrum. *Brain*, 147(2), 607-626.
48. Taylor-Rubin, C., Croot, K., & Nickels, L. (2021). Speech and language therapy in primary progressive aphasia: a critical review of current practice. *Expert review of neurotherapeutics*, 21(4), 419-430.

49. Volkmer, A., Spector, A., Warren, J.D., & Beeke, S. (2020). Speech and language therapy for primary progressive aphasia: Referral patterns and barriers to service provision across the UK. *Dementia*, 19(5), 1349-1363.
50. Walenski, M., Sostarics, T., Mesulam, M.M., & Thompson, C.K. (2024). The production of adjectives in narratives by individuals with primary progressive aphasia. *Journal of Neurolinguistics*, 69, 101179.
51. Agmon, G., Cho, S., Ash, S., Cousins, K.A., Blennow, K., Zetterberg, H., Shaw, L.M., Pradhan, S., Kim, Y.D., & Liberman, M.Y. (2025). Automatic quantification of syntactic complexity in natural spontaneous speech of people with primary progressive aphasia. *Aphasiology*, 1-22.
52. Vonk, J.M.J., Morin, B.T., Pillai, J., Rosado Rolon, D., Bogley, R., Baquirin, D.P., Ezzes, Z., Tee, B.L., de Leon, J., & Wauters, L. (2025). Automated Speech Analysis to Differentiate Frontal and Right Anterior Temporal Lobe Atrophy in Frontotemporal Dementia. *Neurology*, 104(9), e213556.
53. Farias, S.T., Chand, V., Bonnici, L., Baynes, K., Harvey, D., Mungas, D., Simon, C., & Reed, B. (2012). Idea density measured in late life predicts subsequent cognitive trajectories: implications for the measurement of cognitive reserve. *J Gerontol B Psychol Sci Soc Sci*, 67(6), 677-686. doi:10.1093/geronb/gbr162