

Speech, Language, p-tau217 and Neurofilament Light Chain in an Alzheimer’s Disease At-Risk Population

Alveena Siddiqui^{1,2}, Peter J. Snyder^{3,4*}, Jessica Alber^{3,5}, Thayabaran Kathiresan^{1,2}, Adam P. Vogel^{1,2*}

¹School of Health Sciences, The University of Melbourne, Melbourne, VIC, Australia · ²Redenlab Inc., Melbourne, VIC, Australia · ³College of Pharmacy, University of Rhode Island, Kingston, RI, USA · ⁴Department of Neurology, Alpert Medical School of Brown University, Providence, RI, USA · ⁵Memory and Aging Program, Butler Hospital, Providence, RI, USA

*Joint senior authors | Corresponding author: Adam P. Vogel, PhD - vogela@unimelb.edu.au



1 BACKGROUND

- Alzheimer’s disease (AD) progressively disrupts memory, executive function, and language.
- Speech and language analyses are non-invasive, repeatable, and sensitive to neurological dysfunction - promising scalable digital markers of cognitive decline.
- Slowed articulation, increased pausing, and reduced idea density / lexical diversity can distinguish MCI or early AD from healthy controls.
- Few studies directly link speech/language features to established plasma biomarkers, especially in at-risk (preclinical) populations.
- Plasma p-tau217 closely reflects tau tangle accumulation - high diagnostic specificity for AD.
- Plasma neurofilament light chain (NfL) is a general marker of axonal injury / neurodegeneration, elevated in AD and linked to poorer cognition.

2 AIM

To examine associations between speech/language PCA components and plasma p-tau217 and NfL across Low-Risk, High-Risk, and Mild Cognitive Impairment groups, using Principal Component Analysis to address dimensionality and multiple comparisons.

Hypothesis: speech/language features reflecting motor timing, lexical diversity, and informativeness would be significantly associated with biomarker levels, with distinct profiles across clinical risk groups.

STUDY AT A GLANCE

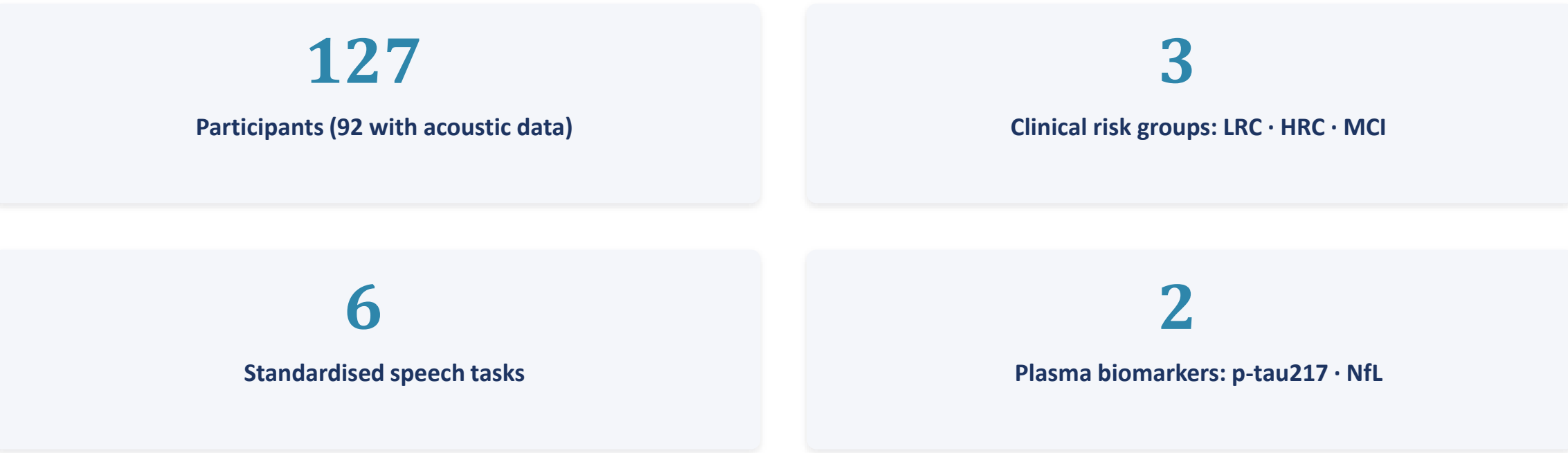


Table 1. Participant Characteristics

	LRC	HRC	MCI
N	49	60	18
Female / Male	24 / 25	43 / 17	7 / 11
Age, y (M±SD)	66.7±6.6	65.5±5.4	69.3±5.2
MoCA score	27.2±1.6	27.8±1.8	23.5±2.7
CDR score	0	0	0.5
1*-relative with AD	Negative	Pos./Suspect	-
APOE ε4 status	Non-carrier	Carrier (≥1)	-
Plasma Aβ42/40	0.046±0.008	0.043±0.007	0.045±0.010
p-tau217 (pg/mL)	0.33±0.27	0.48±0.31	0.67±0.57
NfL (pg/mL)	17.78±6.92	13.89±4.90	17.69±5.37

MoCA = Montreal Cognitive Assessment; CDR = Clinical Dementia Rating; Aβ = amyloid-beta. Values are Mean ± SD.

KEY TAKEAWAY

Speech and language features tracked general neurodegeneration (NfL) rather than AD-specific tau pathology (p-tau217) - supporting speech as a scalable, non-invasive marker of neurodegeneration.

3 METHODS

Study Design & Participants

- Cross-sectional study; N = 127 (acoustic feature subset, n = 92); ages 55–80, fluent English speakers.
- Low-Risk Cognitively Normal (LRC): no complaints, no family history, no APOE ε4 allele.
- High-Risk Cognitively Normal (HRC): subjective memory concerns, positive family history, ≥ 1 APOE ε4 allele.
- MCI: clinical diagnosis, RBANS Delayed Memory Index ≤ 85, positive amyloid biomarker (CSF/PET).
- Exclusions: neurological/psychiatric illness, substance misuse, uncorrected hearing/vision impairment.

Procedures & Speech Tasks

- Fasting blood draw → plasma p-tau217 (ALZpath) & NfL, quantified via Simoa (Quanterix).
- 6 standardised tasks (lavalier mic + iPad, Redenlab®): sustained vowel phonation, DDK (/p/-/t/-/k/), reading (Grandfather Passage), open monologue, days-of-week recitation, picture description.
- Acoustic features: Redenlab Analyze (v0.14.42) - HNR, MFCCs, articulation rate, pause/timing metrics.
- Linguistic features: Stanza (v1.10.1) + custom Python, POS tagging & lemmatisation on monologue and picture-description transcripts.

Statistical Analysis

- Features z-standardised; missing values median-imputed; outliers (±3 SD) winsorised.
- Separate PCA for linguistic (37 features → 3 PCs, 46% variance) and acoustic (112 features → 4 PCs, 33.2% variance) domains.
- PC scores entered as predictors of log p-tau217 and NfL in linear regression, adjusted for age, sex, clinical group.
- Group × PC interaction models tested for effect-modification.
- Benjamini-Hochberg FDR correction applied within domain (6 tests language, 8 acoustic).

PCA OVERVIEW

PCA reduced high-dimensional speech/language feature sets to a small number of interpretable components, used as predictors in the biomarker regression models shown at right.

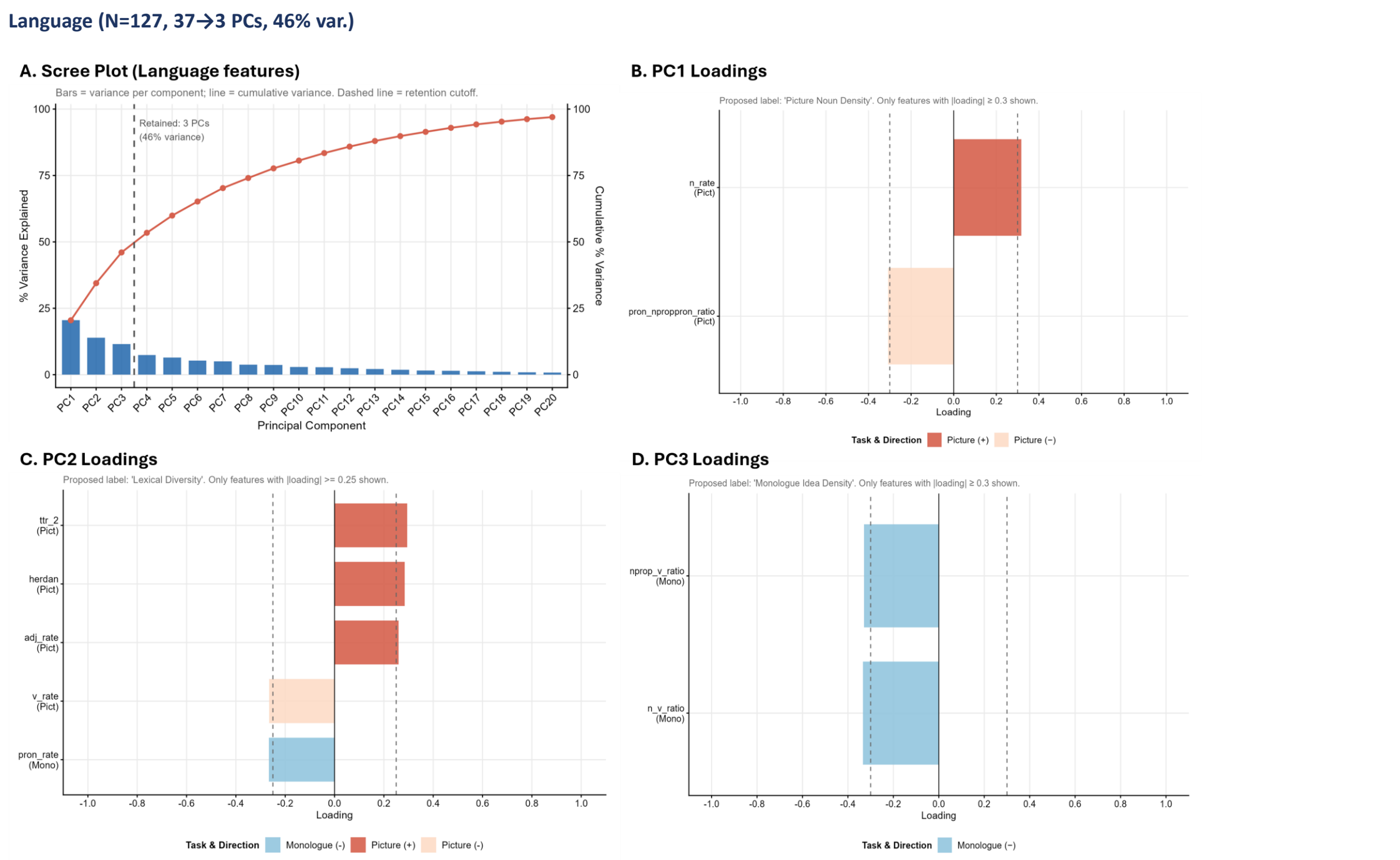


Figure 1. Language PCA loadings. PC1: Picture Noun Density (20.0%), PC2: Monologue Idea Density (14.0%), PC3: Lexical Diversity (12.0%)

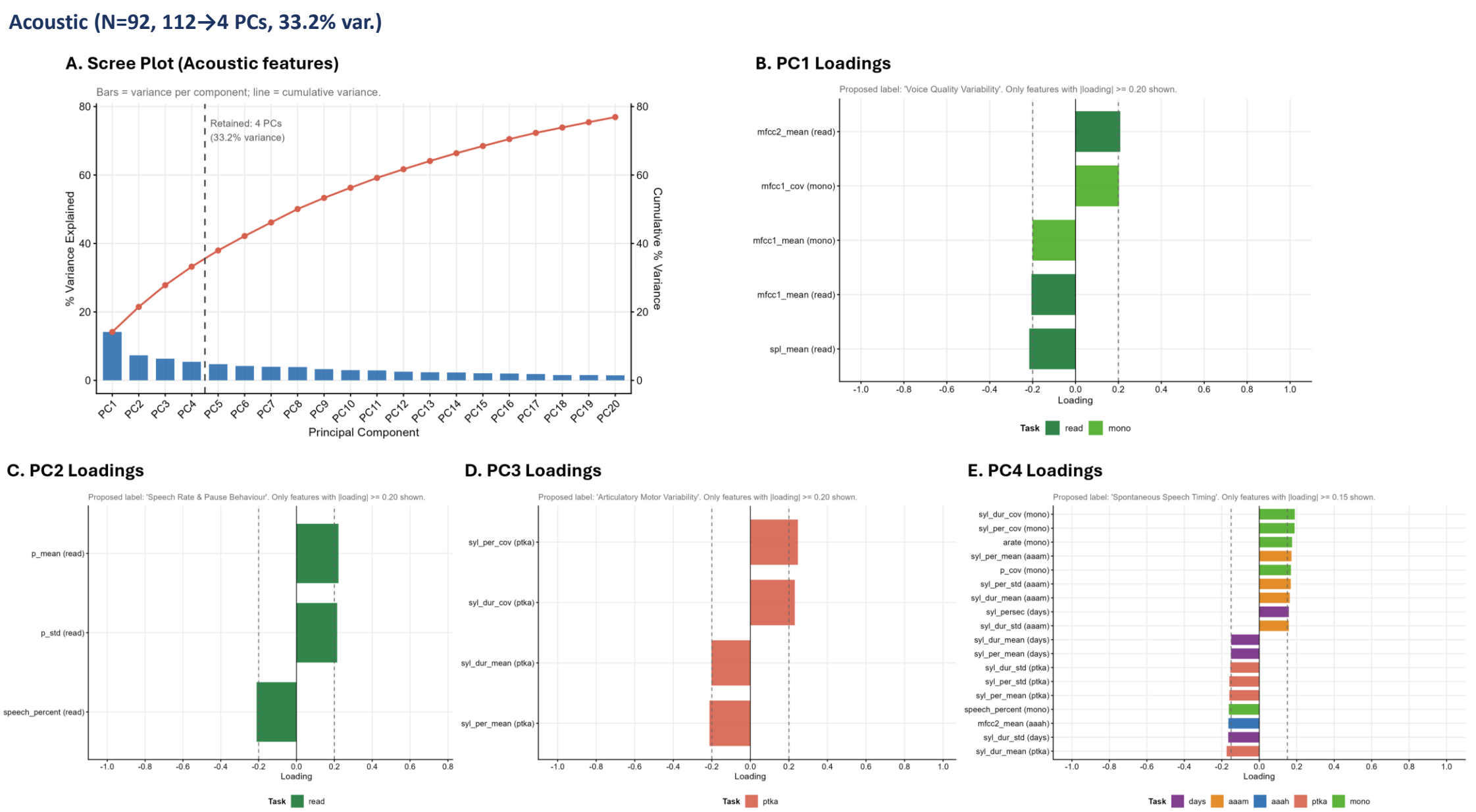


Figure 2. Acoustic PCA loadings. PC1: Voice Quality Variability (14.1%), PC2: Speech Rate & Pause (7.3%), PC3: Articulatory Motor Var. (6.3%), PC4: Spontaneous Speech Timing (5.4%)

→ These 7 components (3 language + 4 acoustic) were entered as predictors of plasma p-tau217 and NfL in the regression models shown at right (Tables 2 & 3).

4 BIOMARKER ASSOCIATIONS

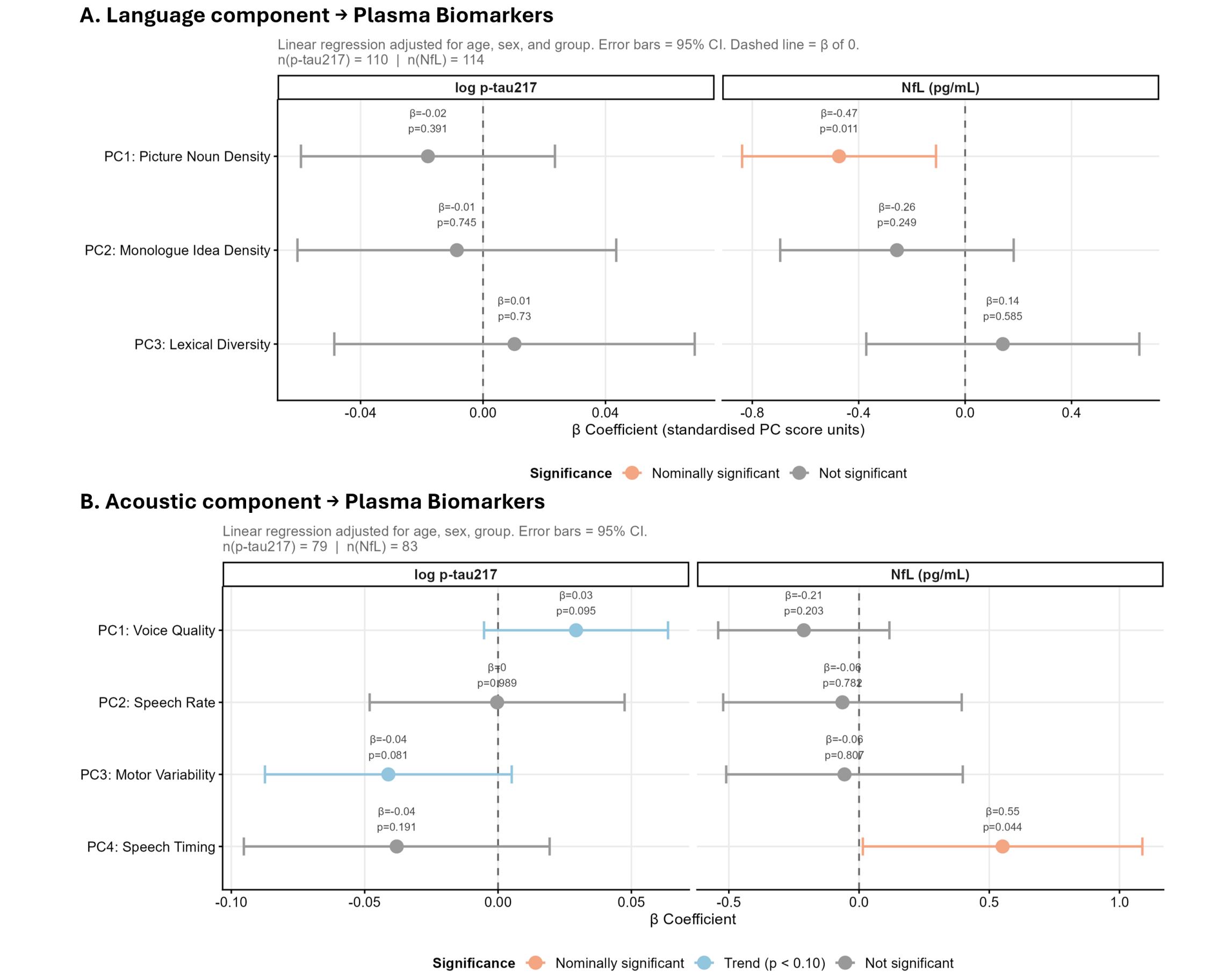


Figure 3. Forest plots of standardised β coefficients (95% CI): (A) language and (B) acoustic PC scores → log p-tau217 and NfL, adjusted for age, sex, clinical group. Orange = nominally significant; blue = trend; grey = not significant.

Table 2. Language PC Scores → Plasma Biomarkers

Component	β [95% CI]	p	p(FDR)
log p-tau217 (R² = 0.197, N = 110)			
PC1: Picture Noun Density	−0.018 [−0.059, 0.023]	.391	.745
PC2: Monologue Idea Density	−0.009 [−0.061, 0.044]	.745	.745
PC3: Lexical Diversity	0.010 [−0.049, 0.069]	.730	.745
NfL, pg/mL (R² = 0.277, N = 114)			
PC1: Picture Noun Density	−0.474 [−0.838, −0.109]	.011	.068
PC2: Monologue Idea Density	−0.256 [−0.695, 0.182]	.249	.745
PC3: Lexical Diversity	0.142 [−0.372, 0.655]	.585	.745

Orange = nominally significant (p<.05). All models adjusted for age, sex, clinical group (HRC = reference); FDR within domain.

Table 3. Acoustic PC Scores → Plasma Biomarkers

Component	β [95% CI]	p	p(FDR)
log p-tau217 (R² = 0.252, N = 79)			
PC1: Voice Quality Variability	0.029 [−0.005, 0.064]	.095	.254
PC2: Speech Rate & Pause	0.000 [−0.048, 0.047]	.989	.989
PC3: Articulatory Motor Var.	−0.041 [−0.087, 0.005]	.081	.254
PC4: Spontaneous Speech Timing	−0.038 [−0.095, 0.019]	.191	.254
NfL, pg/mL (R² = 0.284, N = 83)			
PC1: Voice Quality Variability	−0.212 [−0.541, 0.117]	.203	.324
PC2: Speech Rate & Pause	−0.064 [−0.522, 0.394]	.782	.922
PC3: Articulatory Motor Var.	−0.056 [−0.510, 0.398]	.807	.922
PC4: Spontaneous Speech Timing	0.551 [0.015, 1.088]	.044	.254

Orange = nominally significant; blue = trend (p<.10). No component survived FDR correction within either domain.

- No significant Group × Component interactions - associations were transdiagnostic (consistent across LRC, HRC, MCI).

5 DISCUSSION

- Speech/language features associated primarily with NfL, not p-tau217 - likely indexing general neurodegeneration rather than AD-specific tau pathology.
- ↑ Noun density in picture description may reflect preserved referential specificity, maintained longer through the disease trajectory.
- ↑ Spontaneous speech timing irregularity may reflect cognitive-motor integration demands vulnerable to early neurodegeneration.

Limitations

- Cross-sectional design - no temporal / progression inference.
- Small MCI acoustic subsample (n = 13) limits stability of group estimates.
- Associations did not survive FDR correction - exploratory, hypothesis-generating findings.

6 CONCLUSION

Speech-derived digital markers show promise as scalable, non-invasive indices of neurodegeneration. Longitudinal replication and multimodal validation are warranted.