Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Sustained phonation and connected-speech timing capture complementary motor and cognitive features in Parkinson's disease

Honghao Xu
Chengdu Medical College
Zhuochi Zou
Chengdu Medical College
Jie Zhou
The Third People's Hospital of Sichuan Province
Yanxia Wen
Chongqing Western Hospital
Zichen Wang
The First Affiliated Hospital of Chengdu Medical College
Haroona Bashir
Chengdu Medical College
Mengxi Wu
Chengdu Medical College
Tao Hu
The Third People's Hospital of Sichuan Province
Xianwei Zou
chinazxw002@cmc.edu.cn

The First Affiliated Hospital of Chengdu Medical College

Article

Keywords: Parkinson's disease, sustained phonation, connected speech, acoustic biomarkers, cognition, hypokinetic dysarthria

Posted Date: October 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10605653/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Sustained phonation and connected-speech timing capture complementary motor and cognitive features in Parkinson's disease

Honghao Xu[1]†, Zhuochi Zou[1]†, Jie Zhou[2]†, Yanxia Wen[3], Zichen Wang[4], Haroona Bashir[1], Mengxi Wu[1], Tao Hu[2]*, Xianwei Zou[4]*

[1] School of Clinical Medicine, Chengdu Medical College, Chengdu 610500, Sichuan, China
[2] Department of Neurology, The Third People's Hospital of Sichuan Province, Chengdu 610100, Sichuan, China
[3] Department of Neurology, Chongqing Western Hospital, Chongqing 400055, China
[4] Department of Neurology, The First Affiliated Hospital of Chengdu Medical College, Chengdu 610500, Sichuan, China

† These authors contributed equally to this work: Honghao Xu, Zhuochi Zou and Jie Zhou.

* Correspondence:
Tao Hu, Department of Neurology, The Third People's Hospital of Sichuan Province, Chengdu 610100, Sichuan, China. Email: hutao2010@qq.com
Xianwei Zou, Department of Neurology, The First Affiliated Hospital of Chengdu Medical College, No. 278, Middle Section of Baoguang Avenue, Xindu District, Chengdu, Sichuan 610500, China. Email: chinazxw002@cmc.edu.cn
Tao Hu and Xianwei Zou jointly supervised this work and are co-corresponding authors.

Abstract

Speech impairment in Parkinson's disease (PD) reflects interacting respiratory, phonatory, articulatory and cognitive processes, but the complementary value of sustained phonation and connected speech remains unclear. We analysed two independent, non-overlapping cohorts comprising 78 participants with PD and 27 healthy controls. Cohort 1 (28 PD, 27 controls) characterized sustained-phonation abnormalities, whereas Cohort 2 (50 PD) examined associations of sustained vowels and passage reading with motor severity, perceived voice handicap and cognition. PD was associated with shorter maximum phonation time, lower Dysphonia Severity Index, restricted intensity range, higher shimmer and reduced vowel space. Across six vowels, shimmer was consistently elevated, while vowel-space reduction was similar in men and women. In Cohort 2, slower speech, more and longer pauses, and shorter speech runs were associated with greater motor severity and voice handicap. Mean pause length remained independently associated with both outcomes after adjustment for maximum phonation time and was also related to poorer global cognition and immediate verbal recall. A small acoustic feature set showed promising internally cross-validated discrimination between PD and controls, although external validation is required. These findings indicate that sustained phonation and connected-speech timing provide complementary, clinically interpretable markers of motor and cognitive involvement in PD.

Keywords: Parkinson's disease; sustained phonation; connected speech; acoustic biomarkers; cognition; hypokinetic dysarthria

1. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by a heterogeneous combination of motor and non-motor manifestations [1]. Speech and voice impairment is among its most frequent and disabling features, commonly presenting as hypokinetic dysarthria with reduced loudness, restricted pitch and intensity variation, breathy or harsh voice quality, imprecise articulation, and disrupted speech timing [2–4]. Acoustic manifestations are heterogeneous, and distinct speech phenotypes have been described even in de novo PD [5]. These changes can emerge before advanced motor disability and may substantially limit intelligibility, communication, social participation, and quality of life. The clinical relevance of speech dysfunction is further supported by recent randomized evidence showing that targeted voice treatment can reduce the patient-reported impact of communication problems in PD [6].

Objective acoustic analysis offers a non-invasive and repeatable approach for quantifying speech dysfunction and has attracted increasing interest as a digital biomarker for detection, phenotyping, and longitudinal monitoring [7–9]. Sustained phonation minimizes linguistic demands and permits relatively controlled assessment of respiratory-phonatory capacity, laryngeal stability, and intensity regulation. Studies in untreated or clinically established PD have identified abnormalities in maximum phonation time (MPT), sound pressure level (SPL), perturbation measures, and composite voice-quality indices [10–13]. The Dysphonia Severity Index (DSI) integrates MPT, highest fundamental frequency, minimum intensity, and jitter into an interpretable multidimensional measure [14]. Formant-derived indices, including triangular vowel space area (tVSA) and the Vowel Articulation Index (VAI), additionally quantify articulatory working space and vowel centralization, which may be altered even when overt dysarthria is mild [15–18].

Sustained vowels, however, capture only a constrained component of speech production. Connected speech imposes additional demands on respiratory control, articulation, prosody, sequencing, lexical access, and self-monitoring. Automated analysis of reading and sentence production has shown that speech rate, pause structure, and respiratory timing can reveal parkinsonian abnormalities [19,20]. These temporal features may also reflect cognitive-linguistic

processes rather than motor impairment alone. Recent work in Chinese-speaking cohorts linked slower speech and longer pauses to cognitive impairment in PD [3,21], while discourse studies have demonstrated reduced productivity and informativeness after accounting for motor speech changes [22]. Thus, phonatory measures and connected-speech timing may provide complementary information about the motor-cognitive interface of PD.

Despite rapid growth in speech-based PD research, several limitations remain. Many studies evaluate a single task, emphasize classification performance without examining clinically interpretable associations, or combine heterogeneous recordings in ways that complicate comparison across populations and languages [7,8,23]. Recent cross-language research also indicates that some acoustic markers are relatively language-general whereas others depend on task and linguistic context [24]. Sustained-phonation and connected-speech features are also frequently analysed in separate samples, leaving uncertainty about which acoustic domains primarily reflect disease-related phonatory phenotype, motor severity, perceived voice handicap, or cognition. A design that preserves cohort-specific acquisition and analysis while comparing shared measures can help distinguish reproducible directional patterns from artefacts of protocol or instrumentation.

The present study therefore integrated two independently recruited, non-overlapping cohorts. Cohort 1 was used to define a multidomain sustained-phonation phenotype of PD relative to healthy controls and to examine vowel-specific perturbation, articulatory working space, clinical associations, and exploratory discrimination. Cohort 2 was used to evaluate sustained-phonation and connected-speech correlates of motor severity, perceived voice handicap, and global and domain-specific cognition. We hypothesized that PD would be characterized by impaired phonatory capacity, intensity regulation, stability, and articulatory working space; that connected-speech timing would show stronger associations with motor and cognitive outcomes than isolated phonatory measures; and that shared measures would demonstrate partial directional consistency across cohorts despite differences in recording protocols and clinical assessments.

2. Results

2.1 Participant characteristics

A total of 118 individuals were initially assessed across the two non-overlapping cohorts. In Cohort 1, 34 patients with Parkinson's disease (PD) and 31 healthy controls (HC) were assessed;

six patients with PD and four HC were excluded because they did not complete the speech tasks or because their recordings were unsuitable for acoustic analysis, leaving 28 patients with PD and 27 HC. In Cohort 2, 53 patients with PD were assessed and three were excluded for the same reasons, leaving 50 patients in the final analytical sample. The final study population therefore comprised 105 independent participants: 78 patients with PD and 27 HC (Fig. 1). Because recruitment, recording protocols, and clinical assessments differed between cohorts, no formal between-cohort comparisons were performed.

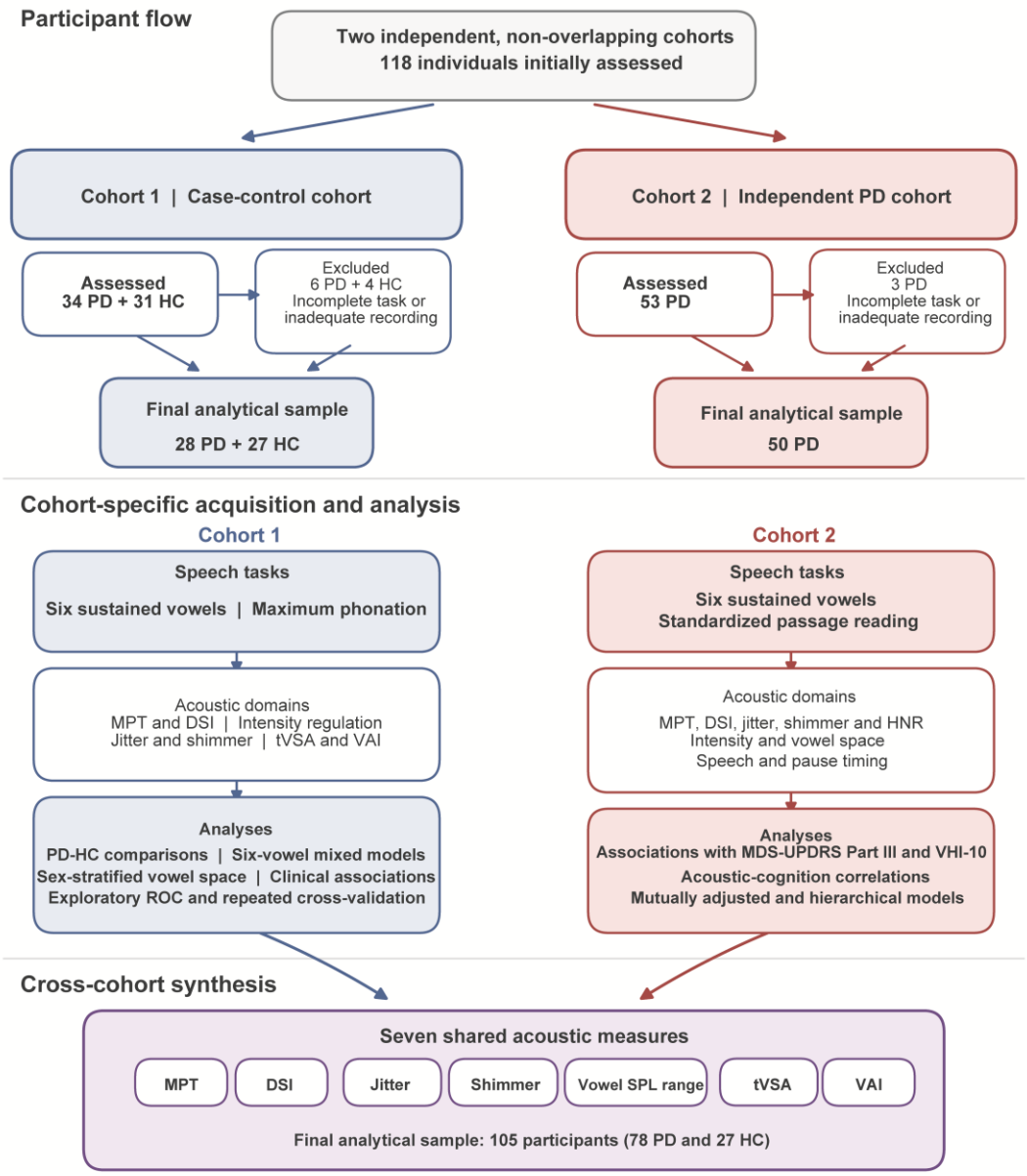

Fig. 1 Participant flow and two-cohort analytical framework. Of 118 individuals assessed, 105 entered the final analyses.

Cohort 1 assessed 34 patients with Parkinson's disease (PD) and 31 healthy controls (HC); six patients with PD and four HC were excluded because of incomplete speech tasks or inadequate recordings, leaving 28 patients with PD and 27 HC. Sustained vowels and maximum phonation supported case-control, vowel-specific, clinical-association, and exploratory ROC analyses. Cohort 2 assessed 53 patients with PD; three were excluded for the same reasons, leaving 50 patients. Sustained vowels and standardized passage reading supported analyses of MDS-UPDRS Part III, VHI-10, and cognition. Cross-cohort synthesis compared directionality for seven shared measures (MPT, DSI, jitter, shimmer, vowel SPL range, tVSA, and VAI) without pooling raw data or estimating a combined effect. DSI, Dysphonia Severity Index; HNR, harmonics-to-noise ratio; MDS-UPDRS, Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale; MPT, maximum phonation time; ROC, receiver operating characteristic; SPL, sound pressure level; tVSA, triangular vowel space area; VAI, Vowel Articulation Index; VHI, Voice Handicap Index.

In Cohort 1, the PD and HC groups did not differ in age, sex, education, or Mini-Mental State Examination (MMSE) score (all $P > 0.20$; Table 1). Patients with PD had higher Voice Handicap Index-30 (VHI-30) and Geriatric Depression Scale-30 scores and lower modified Frenchay Dysarthria Assessment (mFDA) scores than HC (all $P < 0.001$). Among patients with PD, median disease duration was 3.0 years, median Hoehn and Yahr stage was 2.0, and mean Unified Parkinson's Disease Rating Scale (UPDRS) total score was 37.18.

Cohort 2 included 31 men and 19 women, with a mean age of 69.06 years. Median disease duration was 3.0 years, median levodopa equivalent daily dose was 557.5 mg/day, median Hoehn and Yahr stage was 2.0, and median MDS-UPDRS Part III score was 25.0. Median VHI-10 and MMSE scores were 10.5 and 26.0, respectively, and the mean Montreal Cognitive Assessment score was 18.22. Four participants reached the prespecified 360-s ceiling for Trail Making Test part A and 15 reached the ceiling for part B. Detailed clinical, affective, and neuropsychological results are provided in Supplementary Table S1.

Table 1. Demographic and clinical characteristics of the two study cohorts

Characteristic	Cohort 1 HC (n = 27)	Cohort 1 PD (n = 28)	P value	Cohort 2 PD (n = 50)
Age, years	68.78 ± 6.61	69.96 ± 7.55	0.538	69.06 ± 7.99
Male sex, n (%)	10 (37.0)	15 (53.6)	0.282	31 (62.0)
Education level, n (%)			0.585	
Illiterate	5 (18.5)	4 (14.3)		13 (26.0)
Junior high school	13 (48.1)	11 (39.3)		24 (48.0)
High school	8 (29.6)	9 (32.1)		9 (18.0)
University	1 (3.7)	4 (14.3)		4 (8.0)
MMSE score	28.0 [26.0-29.0]	27.0 [25.0-28.0]	0.210	26.0 [24.0–27.0]
MoCA score	-	-	-	18.22 ± 4.43
Disease duration, years	-	3.0 [2.0-4.1]	-	3.0 [1.0–5.8]
LEDD, mg/day	-	-	-	557.5 [375.0–668.8]
Hoehn and Yahr stage	-	2.0 [2.0-3.0] (1-4)	-	2.0 [1.5–2.5] (1–4)
UPDRS total score	-	37.18 ± 14.15	-	-
MDS-UPDRS Part III score	-	-	-	25.0 [14.0–35.8]
VHI-30 score	2.0 [0-4.5]	24.0 [15.8-35.3]	<0.001	-

Characteristic	Cohort 1 HC (n = 27)	Cohort 1 PD (n = 28)	P value	Cohort 2 PD (n = 50)
VHI-10 score	-	-	-	10.5 [2.3–19.0]
mFDA score	25.0 [23.0-26.0]	19.0 [16.8-20.0]	<0.001	-
GDS-30 score	0 [0-1.0]	11.0 [5.8-16.0]	<0.001	-

Values are mean ± standard deviation, median [interquartile range], or n (%). Hoehn and Yahr stage is followed by the observed range in parentheses. P values compare PD and HC within Cohort 1 only; no formal between-cohort comparisons were performed.

Age was compared using Welch's t test, sex using Fisher's exact test, education using the Fisher-Freeman-Halton exact test, and MMSE, VHI-30, mFDA, and GDS-30 using the Mann-Whitney U test.

GDS-30, 30-item Geriatric Depression Scale; HC, healthy controls; LEDD, levodopa equivalent daily dose; mFDA, modified Frenchay Dysarthria Assessment; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; PD, Parkinson's disease; MDS-UPDRS, Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale; UPDRS, Unified Parkinson's Disease Rating Scale; VHI, Voice Handicap Index.

2.2 Multidomain sustained-phonation phenotype and exploratory discrimination in Cohort 1

Representative spectral and waveform displays illustrate the sustained-phonation recordings in HC and PD (Fig. 2a). In age- and sex-adjusted analyses, eight of the ten prespecified acoustic measures differed between groups after false discovery rate (FDR) correction (Fig. 2b and Table 2). Maximum phonation time (MPT) was shorter in PD (adjusted difference, -3.25 s; 95% confidence interval [CI], -5.50 to -0.99; q = 0.010), and the Dysphonia Severity Index (DSI) was lower (adjusted difference, -1.82 points; 95% CI, -2.49 to -1.14; q < 0.001; Fig. 2c).

Intensity regulation was also altered in PD (Fig. 2d). DIVAS-derived minimum sound pressure level (SPL) was higher (adjusted difference, 4.16 dB; 95% CI, 2.03 to 6.29; q < 0.001), whereas SPL range was narrower (adjusted difference, -6.45 dB; 95% CI, -11.03 to -1.86; q = 0.010). Praat-derived mean intensity was lower in PD (adjusted difference, -5.55 dB; 95% CI, -10.31 to -0.80; q = 0.032), whereas maximum SPL did not differ after FDR correction (q = 0.298). For the sustained vowel /a/, shimmer was higher in PD (adjusted difference, 0.94 percentage points; 95% CI, 0.52 to 1.37; q < 0.001), while the difference in jitter did not remain significant after correction (q = 0.075).

Formant-derived measures indicated a smaller articulatory working space in PD (Fig. 2e). The adjusted geometric mean triangular vowel space area (tVSA) in PD was 0.64 times that in HC (95% CI, 0.50 to 0.83; q = 0.001), and the Vowel Articulation Index (VAI) was lower by 0.056 (95% CI, -0.109 to -0.003; q = 0.048).

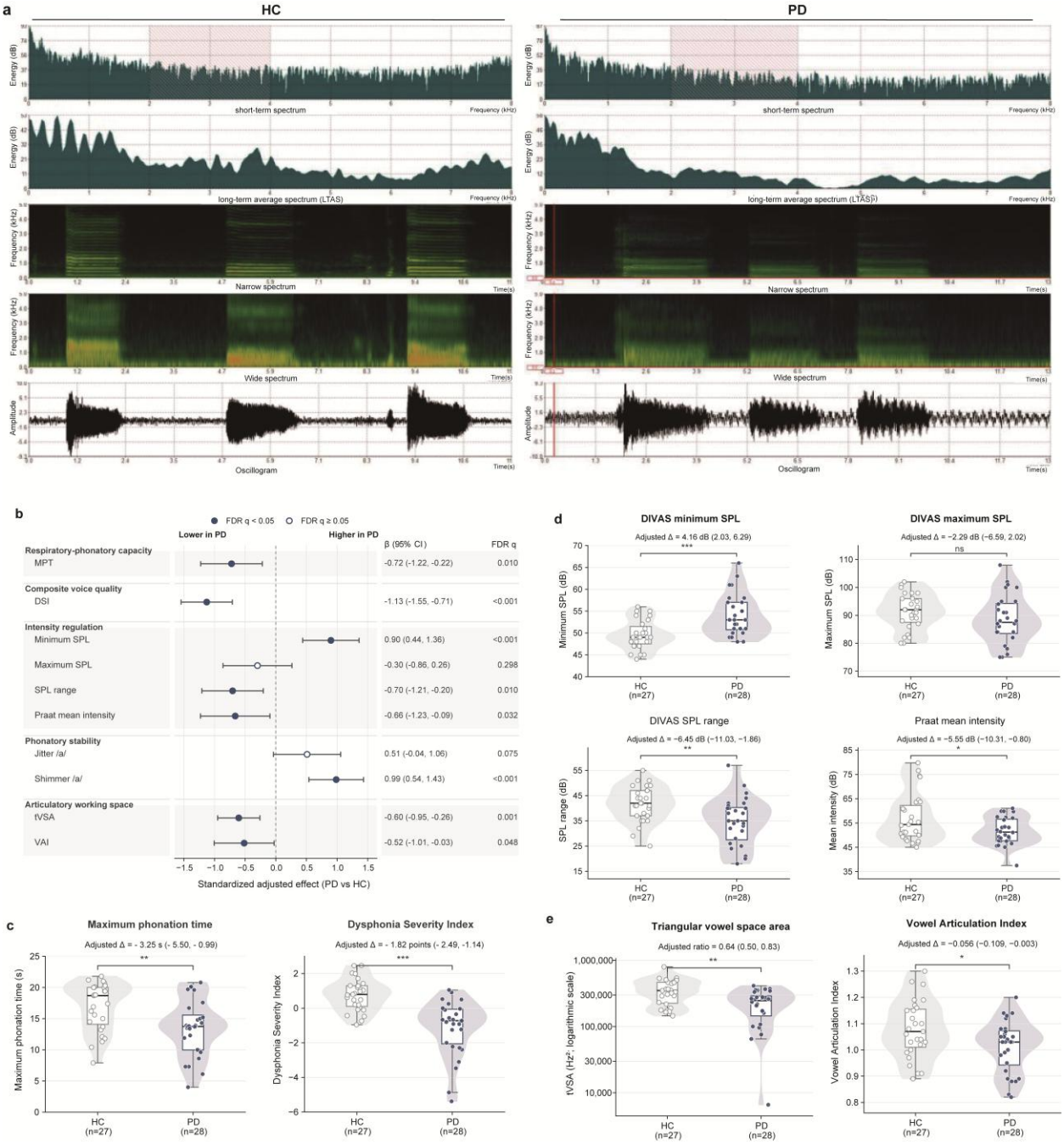

Fig. 2 Representative acoustic displays and multidomain sustained-phonation phenotype in Cohort 1.

(a) Representative frequency-domain, time-frequency, and waveform displays from one healthy control (HC) and one participant with Parkinson's disease (PD). From top to bottom, the rows show a short-term spectrum, long-term average spectrum, narrow-band spectrogram, wide-band spectrogram, and waveform. The displays are illustrative and do not constitute independent inferential evidence. (b) Age- and sex-adjusted standardized group effects for the ten prespecified sustained-phonation measures, estimated using linear regression with HC3 heteroscedasticity-robust standard errors. Filled symbols indicate false discovery rate (FDR)-adjusted $q < 0.05$. (c-e) Individual distributions of maximum phonation time (MPT) and Dysphonia Severity Index (DSI) (c), intensity measures (d), and triangular vowel space area (tVSA) and Vowel Articulation Index (VAI) (e). Violin envelopes show kernel-density estimates, boxes indicate medians and interquartile ranges, whiskers extend to 1.5 times the interquartile range, and points represent individual participants. tVSA is displayed on a logarithmic scale, and its adjusted effect is expressed as a geometric mean ratio; all other adjusted effects are mean differences. Significance symbols are based on FDR-adjusted values across the ten prespecified measures: *$q < 0.05$, **$q < 0.01$, ***$q < 0.001$; ns, not significant. CI, confidence interval; SPL, sound pressure level.

Table 2. Adjusted differences in prespecified sustained-phonation measures in Cohort 1

Measure	HC (n = 27)	PD (n = 28)	Adjusted effect, PD vs HC (95% CI)	FDR q
MPT, s	16.90 ± 3.86	13.46 ± 4.48	-3.25 (-5.50 to -0.99)	0.010
DSI, points	0.76 ± 0.92	-1.13 ± 1.60	-1.82 (-2.49 to -1.14)	<0.001
DIVAS minimum SPL, dB	49.59 ± 3.30	53.93 ± 4.78	4.16 (2.03 to 6.29)	<0.001
DIVAS maximum SPL, dB	91.30 ± 6.37	88.57 ± 8.61	-2.29 (-6.59 to 2.02)	0.298
DIVAS SPL range, dB	41.70 ± 7.16	34.64 ± 9.63	-6.45 (-11.03 to -1.86)	0.010
Praat mean intensity, dB	57.08 ± 9.91	51.75 ± 5.63	-5.55 (-10.31 to -0.80)	0.032
Jitter /a/, %	1.58 ± 0.53	1.83 ± 0.59	0.29 (-0.02 to 0.60)	0.075
Shimmer /a/, %	2.35 ± 0.51	3.41 ± 1.01	0.94 (0.52 to 1.37)	<0.001
tVSA, Hz^2	358,502.05 ± 154,988.25	226,826.54 ± 109,733.39	0.64 (0.50 to 0.83)*	0.001
VAI	1.079 ± 0.110	1.013 ± 0.099	-0.056 (-0.109 to -0.003)	0.048

HC and PD values are unadjusted arithmetic means ± standard deviations. Adjusted effects were estimated using linear regression models including study group, age, and sex with HC3 heteroscedasticity-robust standard errors. Except for tVSA, effects are adjusted mean differences calculated as PD minus HC.

* tVSA was natural-log transformed before analysis. The reported effect is the exponentiated group coefficient and represents the adjusted geometric mean ratio for PD relative to HC.

FDR correction was applied across the ten prespecified measures. CI, confidence interval; DSI, Dysphonia Severity Index; FDR, false discovery rate; HC, healthy controls; MPT, maximum phonation time; PD, Parkinson's disease; SPL, sound pressure level; tVSA, triangular vowel space area; VAI, Vowel Articulation Index.

In sensitivity analyses additionally adjusted for GDS-30, group differences in DSI, minimum SPL, SPL range, Praat-derived mean intensity, and shimmer remained significant after FDR correction, whereas MPT, tVSA, and VAI were attenuated and no longer met the FDR-adjusted significance threshold (Supplementary Table S2).

In exploratory receiver operating characteristic analyses, DSI showed the highest apparent discrimination among the four individual measures examined (area under the curve [AUC], 0.864; bootstrap 95% CI, 0.759 to 0.944), followed by shimmer /a/ (AUC, 0.796; 95% CI, 0.672 to 0.902), log-transformed tVSA (AUC, 0.742; 95% CI, 0.603 to 0.865), and MPT (AUC, 0.710; 95% CI, 0.563 to 0.840). Across 200 repetitions of stratified fivefold cross-validation, the mean AUC was 0.909 (95% empirical range, 0.890 to 0.925) for the acoustic model, 0.895 (0.862 to 0.923) for the combined demographic-acoustic model, and 0.523 (0.422 to 0.593) for the demographic model (Supplementary Fig. S1). These estimates represent internal exploratory validation rather than external diagnostic validation.

2.3 Vowel-specific perturbation and sex-stratified vowel-space patterns in Cohort 1

Jitter and shimmer varied across the six sustained vowels (both $P < 0.001$), but the group-by-vowel interactions were not significant (jitter, $P = 0.432$; shimmer, $P = 0.810$; Fig. 3a,b). In additive mixed-effects models adjusted for vowel, age, and sex, geometric mean jitter was 1.27 times higher in PD than in HC (95% CI, 1.07 to 1.51; $P = 0.006$), and geometric mean shimmer was 1.36 times higher (95% CI, 1.20 to 1.54; $P < 0.001$). Exploratory within-vowel contrasts

showed higher jitter in PD for /o/, /e/, /i/, /u/, and /ü/ after FDR correction, but not for /a/. Shimmer was higher in PD for all six vowels (all $q \leq 0.001$). Complete mixed-effects and vowel-specific estimates are reported in Supplementary Table S3.

Sex-stratified F1-F2 plots showed a smaller vowel-space triangle in PD in both men and women (Fig. 3c,d). The age-adjusted PD-to-HC tVSA ratio was 0.54 in men (bootstrap 95% CI, 0.32 to 0.82) and 0.74 in women (95% CI, 0.61 to 0.88; Fig. 3e). The corresponding adjusted VAI differences were -0.045 in men (95% CI, -0.120 to 0.029) and -0.065 in women (95% CI, -0.129 to -0.001). There was no evidence that sex modified the group effect for tVSA (P for interaction = 0.261) or VAI (P for interaction = 0.702).

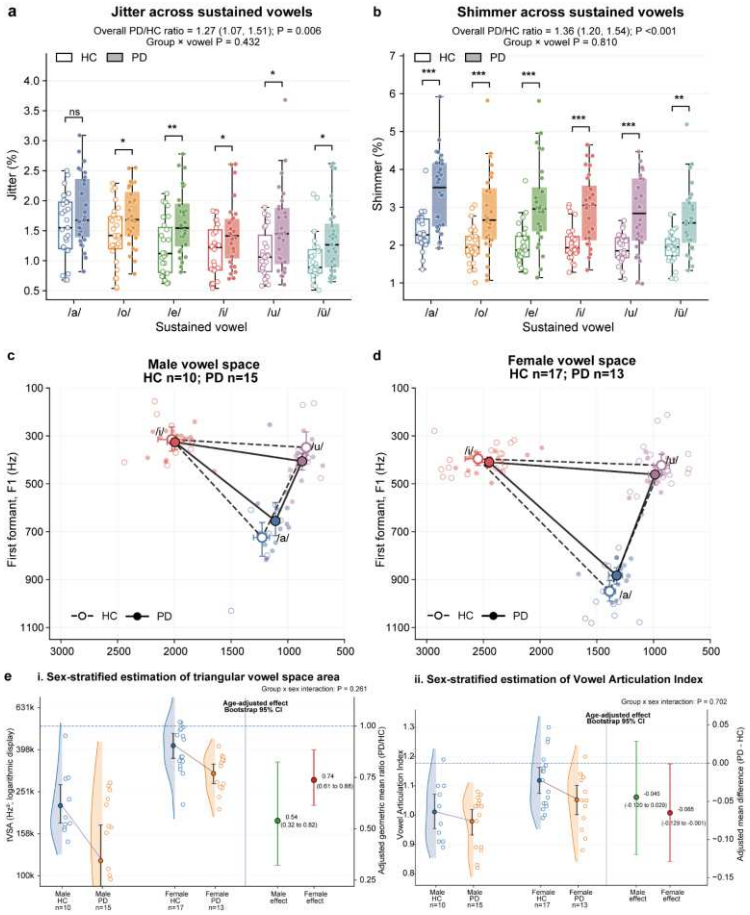

Fig. 3 Vowel-specific phonatory perturbation and sex-stratified vowel-space patterns in Cohort 1.

(a,b) Jitter (a) and shimmer (b) across six sustained vowels in healthy controls (HC) and patients with Parkinson's disease (PD). Boxes indicate medians and interquartile ranges, whiskers extend to 1.5 times the interquartile range, and points represent individual participants. Overall PD-to-HC ratios were estimated using mixed-effects models with participant-specific random intercepts. Within-vowel symbols indicate FDR-adjusted contrasts: *$q < 0.05$, **$q < 0.01$, ***$q < 0.001$; ns, not significant. (c,d) Sex-stratified F1-F2 vowel-space plots for men (c) and women (d). Small points represent individual corner-vowel observations; large points and error bars represent group centroids and bootstrap 95% confidence intervals. F1 and F2 axes are reversed according to conventional acoustic vowel-space presentation. (e) Sex-stratified age-adjusted effects for tVSA and VAI with bootstrap 95% confidence intervals. CI, confidence interval; F1, first formant frequency; F2, second formant frequency; FDR, false discovery rate; tVSA, triangular vowel space area; VAI, Vowel Articulation Index.

2.4 Acoustic interrelationships and correlates of motor severity and perceived voice handicap

Within-cohort correlation matrices showed expected clustering within acoustic domains (Supplementary Fig. S2). In Cohort 1, the strongest FDR-significant correlations included DSI with minimum SPL (Spearman's rho = -0.83), maximum SPL with SPL range (rho = 0.85), and tVSA with VAI (rho = 0.74). In Cohort 2, MPT correlated with DSI (rho = 0.68), jitter with shimmer (rho = 0.70), shimmer inversely with harmonics-to-noise ratio (rho = -0.85), and tVSA with VAI (rho = 0.81). Connected-speech timing measures were strongly interrelated, including speech rate with silent-pause count (rho = -0.88) and silent-pause count with mean run length (rho = -0.95). Cross-domain correlations were generally smaller, although MPT correlated with speech rate and mean run length (both rho = 0.57).

Within Cohort 1, larger log-transformed tVSA was associated with lower UPDRS total score (standardized beta, -0.50; 95% CI, -0.83 to -0.18; q = 0.024; Fig. 4a). Higher minimum SPL was associated with lower VHI-30 score (standardized beta, -0.54; 95% CI, -0.88 to -0.19; q = 0.022). No acoustic measure remained associated with mFDA score after FDR correction. Full estimates are provided in Supplementary Table S4.

In Cohort 2, higher MDS-UPDRS Part III scores were associated with shorter MPT (standardized beta, -0.31; 95% CI, -0.54 to -0.09; q = 0.023), slower speech rate (beta, -0.40; 95% CI, -0.61 to -0.19; q = 0.001), more silent pauses (beta, 0.32; 95% CI, 0.09 to 0.55; q = 0.023), longer mean pause length (beta, 0.55; 95% CI, 0.30 to 0.81; q < 0.001), and shorter mean run length (beta, -0.30; 95% CI, -0.53 to -0.08; q = 0.025; Fig. 4b). Higher VHI-10 scores were associated with shorter MPT, lower DSI, narrower sustained-vowel SPL range, and lower reading SPL mean (q = 0.003 to 0.039). Slower speech rate, more silent pauses, longer mean pause length, and shorter mean run length were also associated with higher VHI-10 scores (all q < 0.001).

In mutually adjusted models, mean pause length remained associated with MDS-UPDRS Part III (standardized beta, 0.52; 95% CI, 0.25 to 0.79; q < 0.001) and VHI-10 (beta, 0.44; 95% CI, 0.19 to 0.70; q = 0.001), whereas MPT was not independently associated with either outcome after adjustment for mean pause length (Fig. 4d and Supplementary Table S4). Among the seven acoustic measures available in both cohorts, coefficient directions were concordant for six measures in the motor-severity analyses and four measures in the voice-handicap analyses (Fig.

4c). These cross-cohort comparisons were descriptive; no pooled effect or formal test of heterogeneity was calculated.

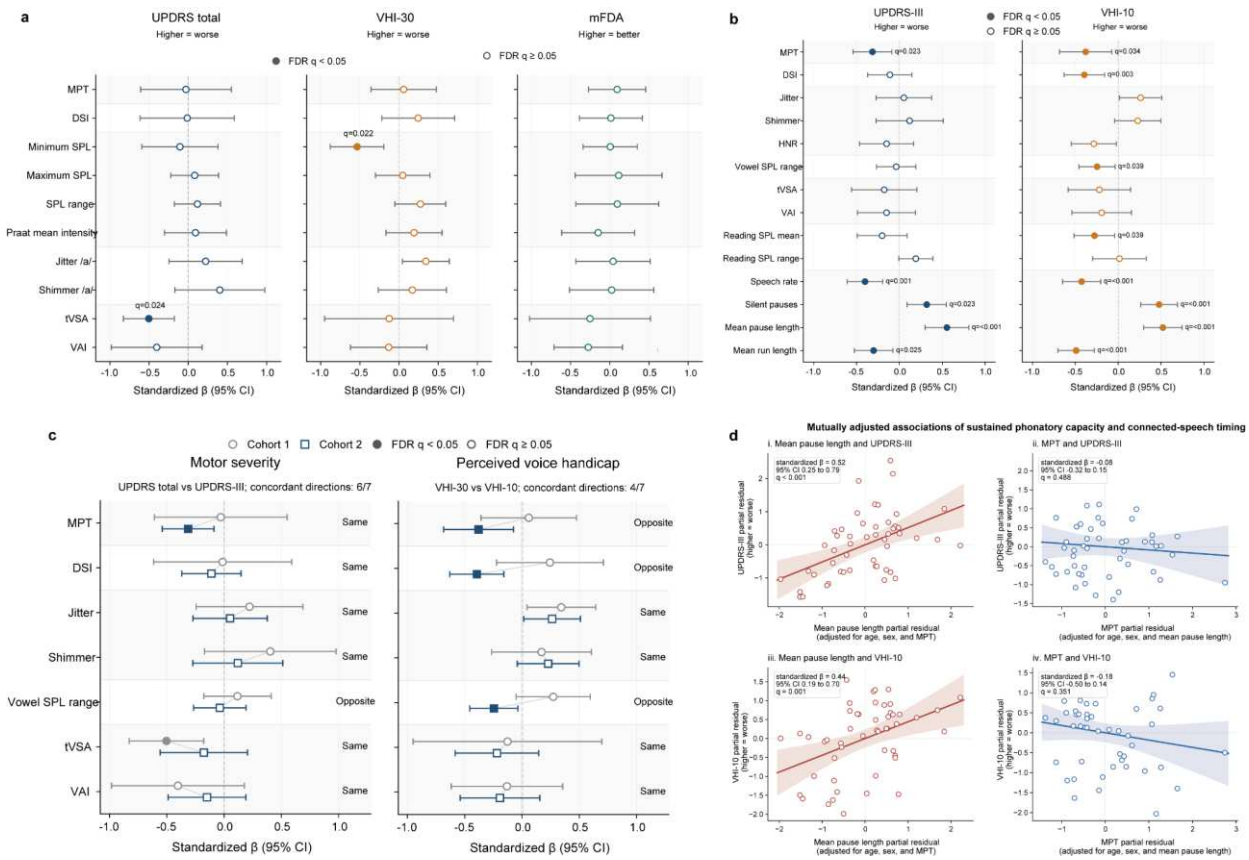

Fig. 4 Acoustic associations with motor severity and perceived voice handicap.

(a) Age- and sex-adjusted standardized associations of ten sustained-phonation measures with UPDRS total score, VHI-30, and mFDA in Cohort 1. (b) Corresponding associations of fourteen sustained-phonation and connected-speech measures with MDS-UPDRS Part III and VHI-10 in Cohort 2. Points indicate standardized regression coefficients and horizontal lines indicate 95% confidence intervals; filled symbols denote FDR-adjusted q < 0.05. (c) Descriptive cross-cohort comparison of seven shared acoustic measures. Circles denote Cohort 1 and squares denote Cohort 2. Lines connect cohort-specific estimates for visual comparison only; no pooled effect was estimated. (d) Partial-regression plots from mutually adjusted Cohort 2 models comparing mean pause length and MPT as correlates of MDS-UPDRS Part III and VHI-10. Points represent partial residuals, solid lines indicate fitted trends, and shaded areas indicate 95% confidence bands. CI, confidence interval; DSI, Dysphonia Severity Index; FDR, false discovery rate; HNR, harmonics-to-noise ratio; mFDA, modified Frenchay Dysarthria Assessment; MPT, maximum phonation time; SPL, sound pressure level; tVSA, triangular vowel space area; MDS-UPDRS, Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale; UPDRS, Unified Parkinson's Disease Rating Scale; VAI, Vowel Articulation Index; VHI, Voice Handicap Index.

2.5 Acoustic correlates of cognition and incremental contribution of pause timing in Cohort 2

Four age-, sex-, and education-adjusted acoustic-cognition associations remained significant after FDR correction (Fig. 5a and Supplementary Table S5). Longer mean pause length was associated with lower MMSE score (partial Spearman's rho, -0.50; 95% CI, -0.69 to -0.24; q = 0.007) and poorer Rey Auditory Verbal Learning Test (RAVLT) immediate recall (rho, -0.44; 95% CI, -0.65 to -0.17; q = 0.031). Longer MPT was associated with better RAVLT immediate

recall (rho, 0.42; 95% CI, 0.15 to 0.64; q = 0.031) and shorter Trail Making Test part A completion time (rho, -0.43; 95% CI, -0.64 to -0.16; q = 0.045; Fig. 5b). No other acoustic-cognition association remained significant after FDR correction.

After exclusion of the four participants who reached the TMT-A ceiling, the association between MPT and TMT-A remained directionally consistent but no longer survived FDR correction (partial Spearman's rho = -0.36, q = 0.280). After exclusion of the 15 TMT-B ceiling cases, MPT, DSI, and mean pause length showed nominal associations with TMT-B performance, but none remained significant after FDR correction (all q = 0.110; Supplementary Table S6).

Hierarchical rank-based models assessed whether mean pause length explained cognitive variation beyond demographic and clinical variables (Fig. 5c and Supplementary Table S7). For MMSE, adjusted R-squared increased from 0.117 in the demographic model to 0.166 after adding disease duration and MDS-UPDRS Part III and to 0.300 after adding mean pause length. The clinical block increased unadjusted R-squared by 0.078 (P = 0.115), whereas mean pause length contributed a further 0.130 (P = 0.004). For RAVLT immediate recall, adjusted R-squared increased from 0.146 to 0.219 and 0.270 across the three models. The corresponding increases in unadjusted R-squared were 0.098 for the clinical block (P = 0.057) and 0.058 for mean pause length (P = 0.055).

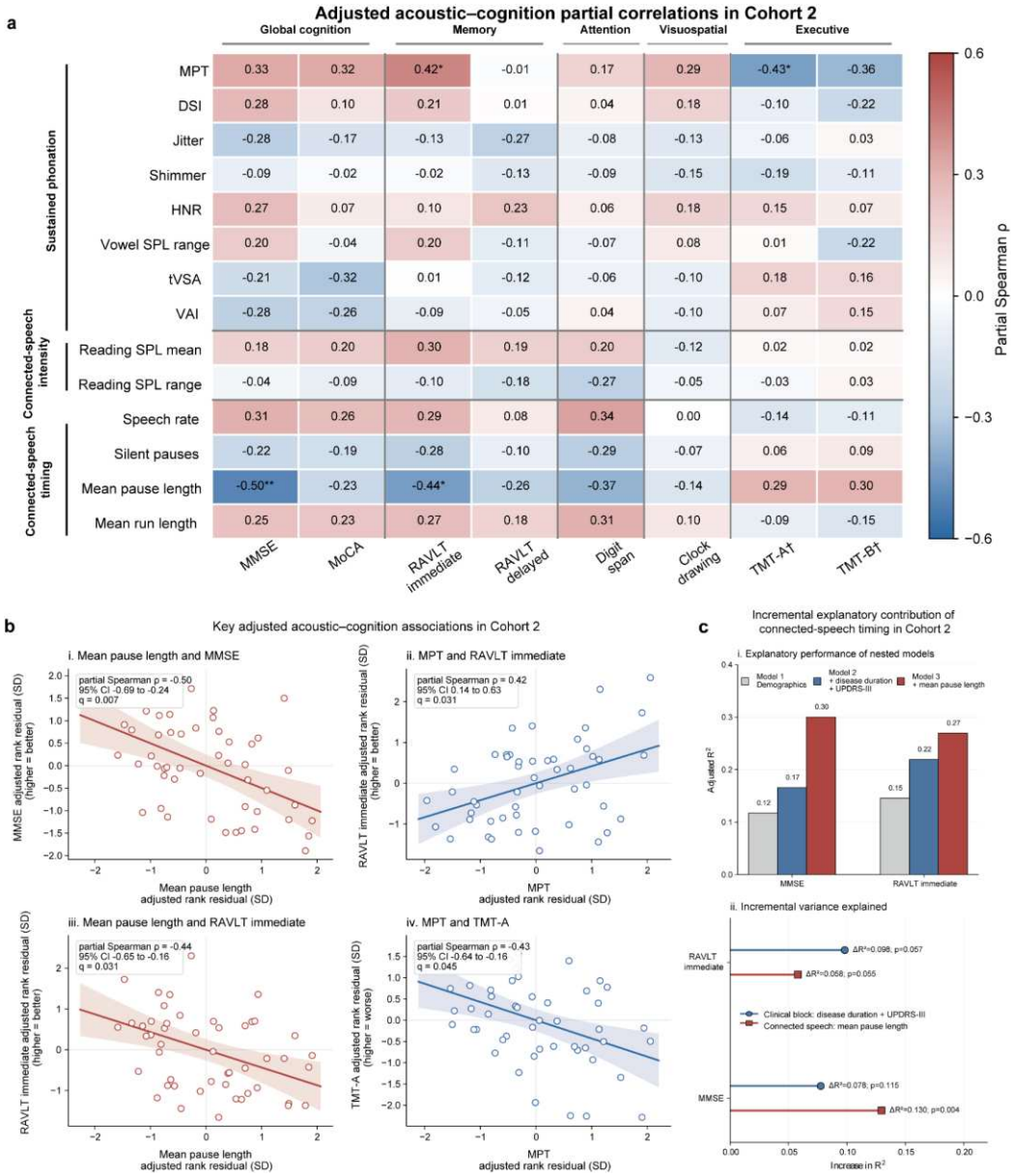

Fig. 5 Acoustic correlates of cognition and incremental contribution of pause timing in Cohort 2.

(a) Heatmap of age-, sex-, and education-adjusted partial Spearman correlations between fourteen acoustic measures and eight cognitive outcomes. Cell values are partial correlation coefficients. Asterisks indicate FDR-adjusted significance: *q < 0.05, **q < 0.01, ***q < 0.001. Higher Trail Making Test values indicate poorer performance; higher scores on the remaining tests indicate better performance. (b) Partial-regression plots for the four associations that remained significant after FDR correction. Points are standardized residuals of rank-transformed variables, solid lines indicate fitted trends, and shaded areas indicate 95% confidence bands. (c) Hierarchical rank-based models for MMSE and RAVLT immediate recall. Model 1 included age, sex, and education; Model 2 additionally included disease duration and MDS-UPDRS Part III; Model 3 additionally included mean pause length. The upper panel shows adjusted R-squared for the three nested models, and the lower panel shows incremental unadjusted R-squared with nested-model P values. CI, confidence interval; DSI, Dysphonia Severity Index; FDR, false discovery rate; HNR, harmonics-to-noise ratio; MMSE, Mini-Mental State Examination; MPT, maximum phonation time; RAVLT, Rey Auditory Verbal Learning Test; SPL, sound pressure level; TMT, Trail Making Test; tVSA, triangular vowel space area; VAI, Vowel Articulation Index.

3. Discussion

This two-cohort study delineated complementary acoustic dimensions of Parkinson's disease (PD). In the case-control cohort, PD was associated with shorter maximum phonation time (MPT), lower Dysphonia Severity Index (DSI), restricted intensity regulation, greater amplitude perturbation, and a smaller articulatory working space. Across the six sustained vowels, the group effect was more consistent for shimmer than for jitter, and the reduction in vowel space was evident in both men and women without statistical evidence of sex-specific effect modification. In the independent clinical cohort, connected-speech timing measures, particularly mean pause length, showed the clearest associations with motor severity, perceived voice handicap, and cognition. Cross-cohort comparisons further indicated partial directional consistency for shared acoustic measures, while the exploratory classification analysis suggested that a small set of interpretable acoustic features could distinguish PD from healthy controls within Cohort 1. Together, these findings support a multidomain view of parkinsonian speech in which sustained phonation primarily characterizes respiratory-phonatory, laryngeal, and articulatory impairment, whereas connected-speech timing captures additional motor-cognitive burden.

The sustained-phonation findings are broadly consistent with recent acoustic studies showing that PD affects several speech-production subsystems rather than a single voice feature [4,10–13]. Shorter MPT points to reduced respiratory-phonatory efficiency, whereas lower DSI reflects the combined influence of reduced phonation time, altered frequency capacity, elevated minimum intensity, and perturbation. The higher shimmer observed in PD indicates less stable cycle-to-cycle amplitude control and agrees with recent acoustic evidence of phonatory instability in parkinsonian dysarthria [12,13]. Jitter showed a weaker main comparison for /a/ but was elevated across the six-vowel mixed-effects model, illustrating that a conclusion based on one sustained vowel may depend on vowel selection. This task dependence is compatible with studies showing that phonetic context and speaking rate affect the discriminative information carried by parkinsonian speech [20,25,26].

The intensity results warrant a nuanced interpretation. PD participants had a higher minimum SPL, a narrower SPL range, and a lower Praat-derived mean intensity, while maximum SPL did not differ after correction. This pattern does not represent uniform hypophonia across all intensity metrics. Instead, it is compatible with compression of the available intensity range and reduced modulation around a constrained operating level.

Differences between summary measures may also arise because minimum, maximum, range, and mean intensity capture distinct aspects of vocal output and can be sensitive to task instructions and signal-processing procedures. Accordingly, the present findings favour multidimensional interpretation rather than treating any single SPL metric as a direct proxy for perceived loudness.

Formant-derived measures provided complementary evidence of impaired articulation. Both tVSA and VAI indicated a smaller articulatory working space in PD, in agreement with previous work on vowel centralization and recent studies of task-dependent and Mandarin-language vowel production [11,13,15–18]. The absence of a significant group-by-sex interaction suggests that the direction of the articulatory effect was similar in men and women, although the sex-stratified confidence intervals were wide. Sex remains important in speech-biomarker research because fundamental and formant frequencies, spectral energy, and some derived features can differ systematically between men and women [24,27]. Larger samples are therefore needed before concluding that the magnitude of vowel-space reduction is sex invariant.

The strongest clinical associations in Cohort 2 involved connected-speech timing. Slower speech rate, more frequent pauses, longer pauses, and shorter runs were associated with greater motor severity and perceived voice handicap, and mean pause length remained independently related to both outcomes after simultaneous adjustment for MPT. These results align with recent studies showing that connected-speech timing and pause behaviour reveal parkinsonian abnormalities that are not fully captured by isolated phonation [19,20]. Longitudinal evidence has likewise identified pause duration, speech rate, and syllable timing as features that change across the clinical course of PD [9,28]. Together, these observations suggest that timing measures integrate multiple processes, including respiratory phrasing, initiation and sequencing of speech movements, articulatory transitions, and online monitoring.

The partial dissociation between MPT and mean pause length is particularly informative. MPT reflects the capacity to sustain a relatively simple phonatory act, whereas reading requires repeated initiation, coordination of respiration with linguistic units, and continuous adjustment of speech output. Their moderate cross-domain correlations indicate shared physiological constraints, but the mutually adjusted models show that pause timing carried additional clinical information in this cohort. This does not imply that connected speech is uniformly superior to sustained phonation. Rather, the two tasks appear to probe different levels of the speech-production hierarchy: sustained vowels offer controlled, interpretable measures of phonatory and articulatory function, while connected speech increases ecological and cognitive demands.

Mean pause length was associated with global cognition and immediate verbal learning after adjustment for age, sex, and education, and it explained additional variance in MMSE beyond demographic variables, disease duration, and MDS-UPDRS Part III. MPT was also related to immediate verbal learning and Trail Making Test part A performance. These findings are consistent with recent evidence that speech rate and pause-related prosody are associated with cognitive impairment in PD [3,21]. They also align with discourse studies showing that communication abnormalities in PD cannot be reduced entirely to motor speech impairment [22]. From a network perspective, speech production depends on coordinated interactions among basal ganglia, frontal executive systems, language networks, auditory feedback pathways, and respiratory-laryngeal control. Recent neuroimaging and sensory-feedback research further supports this distributed motor-language framework [29,30].

The cognitive findings should nevertheless be interpreted conservatively. Reading reduces lexical retrieval demands compared with spontaneous discourse, but it still requires visual scanning, sequencing, working memory, sustained attention, and timing control. Longer pauses may therefore reflect a mixture of bradykinesia, respiratory planning, motor initiation, slowed information processing, and cognitive-linguistic monitoring. The hierarchical model provided clear incremental evidence for MMSE, whereas the additional contribution to RAVLT immediate recall was borderline. Moreover, only four of 112 partial correlations remained significant after FDR correction. The pattern therefore supports selective rather than generalized acoustic-cognitive coupling and does not establish pause duration as a standalone cognitive biomarker.

The present results have two translational implications. First, clinically interpretable acoustic domains may be more useful when considered together than when used in isolation. DSI, shimmer, and tVSA represent complementary aspects of composite voice quality, phonatory stability, and articulation. Their internally cross-validated acoustic model showed promising discrimination in Cohort 1, whereas adding age and sex did not improve performance. Similar studies in previous studies have shown that voice-derived models can identify PD with promising accuracy [23], while work on interpretable speech biomarkers has emphasized the need to relate data-driven features to recognizable physiological or acoustic constructs [27]. The current model follows that principle by using a small number of conventional measures rather than a large opaque feature set.

Second, the results support a task-complementary assessment strategy. A brief sustained-vowel protocol can quantify respiratory-phonatory capacity, perturbation, intensity regulation, and articulatory space, while a short reading task can add temporal measures that are more closely related to clinical burden and cognition. Such an approach may be suitable for repeated digital assessment because speech is inexpensive and non-invasive, and recent work has demonstrated the feasibility of automated speech-based PD evaluation and motor-state monitoring [7,23,31]. However, the ROC analysis in this study was exploratory, internally validated, and confined to Cohort 1. It should not be interpreted as evidence of diagnostic accuracy in an independent population. External validation across centres, languages, disease stages, devices, and recording environments is required before clinical implementation.

Strengths of this study include the use of two independently recruited and non-overlapping cohorts, preservation of cohort-specific acquisition and analysis, coverage of multiple acoustic domains, evaluation of six vowels rather than reliance on /a/ alone, and integration of motor, patient-reported, and neuropsychological outcomes. The analytical strategy used age- and sex-adjusted robust regression, mixed-effects modelling for repeated vowels, correction for multiple testing, mutually adjusted models, bootstrap confidence intervals, and repeated cross-validation. The separation of the cohorts also avoided inappropriate pooling of measurements obtained using different instruments and protocols.

Several limitations should be considered. First, the cross-sectional, single-centre design and modest sample sizes limit causal inference, subgroup analysis, and generalizability. Second, the cohorts differed in recruitment period, recording equipment, microphone distance, clinical scales, and available acoustic measures. Cross-cohort synthesis was therefore limited to descriptive directionality and cannot be regarded as replication of identical effect estimates. Third, all Cohort 2 assessments were performed in the medication ON state, and medication-related fluctuations were not examined. Fourth, GDS-30 scores differed substantially between PD and healthy controls in Cohort 1. Additional GDS-30-adjusted analyses preserved the DSI, intensity, and shimmer effects but attenuated MPT and vowel-space effects, indicating that depressive symptoms may explain part of the latter differences and that the modest, imbalanced sample limits precise separation of disease-group and mood effects. Fifth, the reading task does not capture the lexical, grammatical, and discourse demands of spontaneous speech, and the pause segmentation procedure did not distinguish respiratory, linguistic, and hesitation pauses. Sixth, TMT completion times were capped at 360 s, producing ceiling values, particularly for part B.

Sensitivity analyses excluding ceiling cases yielded directionally similar but FDR-nonsignificant estimates, with reduced precision because only 35 participants remained in the TMT-B analysis. Seventh, no concurrent laryngoscopic, aerodynamic, respiratory, or neurophysiological measures were available to identify the biological source of individual acoustic abnormalities. Finally, the exploratory discrimination model lacked an external test cohort, and its performance may be optimistic despite repeated cross-validation. Longitudinal multicentre studies with harmonized recording, blinded clinical ratings, explicit pause annotation, and independent validation are needed to determine sensitivity to progression and treatment response.

In conclusion, PD was associated with a multidomain sustained-phonation phenotype comprising reduced respiratory-phonatory capacity, altered intensity regulation, greater phonatory instability, and a smaller articulatory working space. Connected-speech timing provided complementary information: slower and more fragmented reading was more consistently associated with motor severity, perceived voice handicap, and selected cognitive outcomes, and mean pause length retained independent clinical and cognitive associations after adjustment. The two tasks should therefore be viewed as complementary rather than interchangeable. These findings support the use of interpretable acoustic measures for motor-cognitive phenotyping, while the exploratory classification results require external validation before diagnostic application. Harmonized longitudinal studies are now needed to establish reproducibility, treatment sensitivity, and clinical utility across languages and recording settings.

4. Methods

4.1 Study design and ethical approval

This study was an integrated cross-sectional analysis of two independently recruited, non-overlapping cohorts. Cohort 1 used a case-control design to characterize the sustained-phonation phenotype of Parkinson's disease (PD) relative to healthy controls (HC). Cohort 2 comprised an independent clinical cohort of patients with PD and was used to examine sustained-phonation and connected-speech correlates of motor severity, perceived voice handicap, and cognition. Because the cohorts were recruited under different protocols and used different recording systems and clinical assessments, analyses were conducted separately within each cohort. Cross-cohort synthesis was restricted to descriptive comparison of coefficient directions for shared acoustic measures; raw measurements were not pooled and no combined effect estimate was

466 calculated. Participant flow, cohort-specific acquisition and analysis, and the cross-cohort
467 synthesis framework are summarized in Fig. 1.

468 Cohort 1 was approved by the Medical Ethics Committee of the First Affiliated Hospital of
469 Chengdu Medical College (No. 2021CYFYrb-SQ-04). Cohort 2 was approved by the
470 Institutional Review Board of the same hospital (No. 2024CYFYIRB-SQ-28). All procedures
471 were conducted in accordance with the Declaration of Helsinki and relevant institutional
472 requirements [32]. Written informed consent was obtained from all participants before study
473 procedures.

474 **4.2 Participants**

475 Cohort 1 initially comprised 34 patients with PD and 31 HC recruited at the First Affiliated
476 Hospital of Chengdu Medical College between January and December 2021. Six patients with
477 PD and four HC were excluded because they did not complete the speech tasks or because their
478 recordings were of insufficient quality for acoustic analysis, leaving 28 patients with PD and 27
479 HC in the final analytical sample. Patients were recruited from the Department of Neurology,
480 and HC were recruited from the hospital physical examination centre. Idiopathic PD was
481 diagnosed according to the 2015 Movement Disorder Society clinical diagnostic criteria [33].
482 Cognitive eligibility was screened using education-adjusted Mini-Mental State Examination
483 (MMSE) thresholds: at least 17 points for participants with no formal education, at least 20
484 points for those with primary-school education, and at least 24 points for those with junior-high-
485 school education or above [34]. Participants were required to understand the instructions and
486 complete the clinical and speech assessments. Additional eligibility criteria included no self-
487 reported hearing impairment that interfered with testing, no professional voice training or
488 previous speech therapy, no respiratory infection within the preceding month, no clinically
489 relevant pharyngeal disease on fibre laryngoscopy, and no previous trauma or surgery involving
490 the phonatory system. Exclusion criteria included secondary parkinsonism, other
491 neurodegenerative or major neurological disease, severe systemic or psychiatric illness, and
492 inability to complete the recording protocol.

493 Cohort 2 initially comprised 53 patients with PD recruited from the Department of
494 Neurology of the First Affiliated Hospital of Chengdu Medical College between January and
495 December 2024. Three participants were excluded because they did not complete the speech
496 tasks or because their recordings were of insufficient quality for acoustic analysis, leaving 50
497 patients with PD in the final analytical sample. Eligibility criteria were a diagnosis of idiopathic

498 PD according to the 2015 Movement Disorder Society criteria [33], age 40–90 years, a stable
499 clinical condition, and sufficient reading ability to complete the standardized passage. Exclusion
500 criteria included other neurological or communication disorders, clinically relevant laryngeal
501 disease, severe psychiatric illness, major systemic disease, or inability to complete the clinical,
502 cognitive, or speech assessments. Clinical assessments and speech recordings were performed 1–
503 4 h after anti-parkinsonian medication intake during the medication ON state. Levodopa
504 equivalent daily dose was calculated using updated standard conversion procedures [35].

505 **4.3 Clinical and neuropsychological assessment**

506 Demographic variables collected in both cohorts included age, sex, and education. In Cohort
507 1, clinical characterization included disease duration, Hoehn and Yahr stage [36], Unified
508 Parkinson's Disease Rating Scale (UPDRS) total score [37], the 30-item Voice Handicap Index
509 (VHI-30) [38], the modified Frenchay Dysarthria Assessment (mFDA) [39], the 30-item
510 Geriatric Depression Scale (GDS-30) [40], and MMSE [34]. Higher UPDRS, VHI-30, and GDS-
511 30 scores indicated greater impairment, whereas higher mFDA scores indicated better speech-
512 motor or articulatory performance.

513 In Cohort 2, motor and non-motor status was assessed using the Movement Disorder
514 Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS)
515 Parts I-IV [41], Hoehn and Yahr stage [36], the Hamilton Anxiety Rating Scale [42], the
516 Hamilton Depression Rating Scale [43], and the 10-item Voice Handicap Index (VHI-10) [44].
517 Global cognition was assessed using the MMSE [34] and Montreal Cognitive Assessment [45].
518 Domain-specific testing comprised Rey Auditory Verbal Learning Test immediate and delayed
519 recall [46], Digit Span Test, Clock Drawing Test [47], and Trail Making Test parts A and B
520 (TMT-A and TMT-B) [48]. All clinical and neuropsychological assessments were administered
521 by trained neurologists. TMT completion time was capped at 360 s; participants unable to
522 complete a task within 360 s were assigned 360 s and flagged as having reached the ceiling.
523 Higher TMT values represented poorer performance, whereas higher scores on the remaining
524 cognitive measures represented better performance.

525 **4.4 Speech tasks and recording procedures**

526 In Cohort 1, recordings were obtained in a sound-treated speech room with background
527 noise below 45 dB. Participants were seated upright and maintained an approximately 10–15 cm
528 mouth-to-microphone distance. Under standardized instruction, they sustained the six Mandarin

monophthongs /a/, /o/, /e/, /i/, /u/, and /ü/ at comfortable loudness and habitual pitch, with the aim of maintaining a steady phonation. Each vowel was produced twice, with an interval of approximately 2 s between repetitions; each production lasted approximately 3–5 s. The higher-quality production was retained, and the most stable 1–2-s segment from the middle of that recording was used for vowel-specific jitter and shimmer extraction. Maximum phonation time was measured twice, and the longer duration was retained. Signals were acquired using a unidirectional microphone and an LA01 pre-low-pass filter amplifier at 44.1 kHz and 16-bit resolution. The microphone gain was 10 dB and the filter cut-off was 20 kHz. Recording procedures followed established recommendations for instrumental voice assessment and microphone-based voice research [49,50].

In Cohort 2, recordings were obtained in a quiet room with ambient noise below 40 dB using a Sony ICD-PX470 digital recorder. The recorder was positioned approximately 30 cm from the participant's mouth at a 45-degree downward-forward angle. Participants completed sustained production of the same six Mandarin monophthongs and read a standardized Mandarin version of 'The North Wind and the Sun' containing approximately 250 syllables. For perturbation analysis, sustained /a/ phonation was maintained for at least 3 s and a stable middle portion was selected for extraction of jitter, shimmer, and harmonics-to-noise ratio. For formant analysis, each vowel was sustained at a comfortable habitual pitch and loudness, with at least 2 s of stable phonation. Maximum phonation time was measured twice, and the longer duration was retained. The same recording arrangement and instructions were used for all participants within the cohort, consistent with general recommendations for standardized acoustic voice acquisition [49,50].

4.5 Acoustic processing and feature definitions

Cohort 1 recordings were processed using DIVAS 2.5 and Praat version 6.0.25. DIVAS supplied maximum phonation time (MPT), minimum and maximum sound pressure level (SPL), maximum fundamental frequency (F0max), jitter, and shimmer, whereas Praat was used for formant extraction and the additional mean-intensity measure. Cohort 2 recordings were processed using Praat version 6.2.09 [51]. All acoustic measures were derived using a uniform procedure within each cohort. The primary feature sets represented respiratory-phonatory capacity, composite voice quality, intensity regulation, phonatory stability, articulatory working space, and connected-speech timing. Because the cohorts used different recording devices, microphone geometries, and processing systems, intensity measures were analysed only within cohort and were not directly pooled or compared between cohorts.

Maximum phonation time (MPT) was defined as the maximum sustained phonation duration in seconds. Minimum and maximum sound pressure level (SPL) were extracted for sustained phonation, and vowel SPL range was calculated as maximum minus minimum SPL. In Cohort 1, Praat-derived mean intensity was additionally retained as a separate intensity measure. Jitter and shimmer quantified cycle-to-cycle frequency and amplitude perturbation, respectively; vowel-specific values were available for all six sustained vowels in Cohort 1, whereas the /a/ measures were used in the main Cohort 2 analyses. Harmonics-to-noise ratio (HNR) was available in Cohort 2.

To ensure identical computation in both cohorts, the Dysphonia Severity Index (DSI) was recalculated using the original multiparametric formulation [14]: DSI = 0.13 × MPT + 0.0053 × F0max - 0.26 × SPLmin - 1.18 × jitter(/a/) + 12.4, where F0max is the maximum fundamental frequency and jitter(/a/) is expressed as a percentage. Lower DSI values indicate greater dysphonia severity.

First and second formant frequencies (F1 and F2) were extracted in Praat for the corner vowels /a/, /i/, and /u/. Triangular vowel space area (tVSA) was calculated as 0.5 x $|F_{1i}(F_{2a} - F_{2u}) + F_{1a}(F_{2u} - F_{2i}) + F_{1u}(F_{2i} - F_{2a})|$. The Vowel Articulation Index (VAI) was calculated as $(F_{2i} + F_{1a})/(F_{1i} + F_{1u} + F_{2u} + F_{2a})$, following established formant-based approaches to vowel centralization [15,16]. Larger tVSA and VAI values generally indicate a more expanded articulatory working space. Sex-stratified F1-F2 plots were generated because absolute formant frequencies differ systematically between men and women.

For the Cohort 2 reading task, mean and range of SPL were obtained from the full reading sample, with reading SPL range calculated as maximum minus minimum SPL. Temporal measures included speech rate (syllables per second), number of silent pauses, mean pause length, and mean run length. Silent pauses were defined as silent intervals lasting at least 0.20 s. Speech segmentation and annotation were performed with reference to both the waveform and spectrogram. During pitch-quality control, automatically detected erroneous periods were manually inspected and corrected before smoothing and interpolation. Silent intervals and continuous speech runs were otherwise identified using a consistent Praat-based segmentation procedure across all Cohort 2 recordings.

590 **4.6 Analytical framework and prespecified measure sets**

591 The Cohort 1 case-control analysis included ten sustained-phonation measures: MPT, DSI,
592 minimum SPL, maximum SPL, SPL range, Praat mean intensity, jitter /a/, shimmer /a/, tVSA,
593 and VAI. Vowel-specific analyses used jitter and shimmer from all six vowels. Cohort 2 analyses
594 included fourteen measures: MPT, DSI, jitter, shimmer, HNR, vowel SPL range, tVSA, VAI,
595 reading SPL mean, reading SPL range, speech rate, number of silent pauses, mean pause length,
596 and mean run length. Seven measures were available in both cohorts for descriptive cross-cohort
597 synthesis: MPT, DSI, jitter, shimmer, vowel SPL range, tVSA, and VAI.

598 **4.7 Statistical analysis**

599 No formal a priori sample-size calculation was performed because this study was a
600 secondary integrated analysis of two existing cohorts. All eligible participants with the required
601 clinical and acoustic data were included.

602 Distributional characteristics were assessed using histograms, Q-Q plots, and Shapiro-Wilk
603 tests. For linear regression models, residual-versus-fitted plots and Q-Q plots of residuals were
604 inspected to assess linearity and residual behaviour. Multicollinearity was evaluated using
605 variance inflation factors, and influential observations were assessed using Cook's distance.
606 Mixed-effects models were checked for convergence and residual patterns.

607 Continuous variables are summarized as mean and standard deviation or median and
608 interquartile range, according to their distribution, and categorical variables as number and
609 percentage. All tests were two-sided. No values were imputed; analyses used available complete
610 observations. Missing clinical values in Cohort 1 HC were structural because PD-specific
611 measures were not applicable. Statistical significance was assessed at $P < 0.05$ or false discovery
612 rate (FDR)-adjusted $q < 0.05$, as appropriate. Benjamini-Hochberg correction was used for all
613 stated FDR procedures [52].

614 For participant characteristics, Cohort 1 age was compared using Welch's t test, sex using
615 Fisher's exact test, education using the Fisher-Freeman-Halton exact test, and MMSE, VHI-30,
616 mFDA, and GDS-30 using Mann-Whitney U tests. No formal demographic or clinical
617 comparisons were made between Cohort 1 and Cohort 2.

618 For the primary Cohort 1 case-control analysis, each acoustic outcome was entered into a
619 separate linear regression model with group, age, and sex as predictors. HC3 heteroscedasticity-
620 robust standard errors were used [53]. tVSA was natural-log transformed; its group coefficient

621 was exponentiated and reported as the adjusted geometric mean ratio for PD relative to HC.
622 Other effects were reported as adjusted mean differences. Standardized group coefficients were
623 additionally calculated for the forest plot. FDR correction was applied across the ten prespecified
624 acoustic measures.

625 As a sensitivity analysis, the primary Cohort 1 case-control models were repeated with
626 additional adjustment for GDS-30 score to assess the potential influence of depressive
627 symptoms. FDR correction was again applied across the ten acoustic outcomes.

628 Vowel-specific jitter and shimmer were natural-log transformed and analysed using linear
629 mixed-effects models with participant-specific random intercepts. Full models included study
630 group, vowel, age, sex, and the group-by-vowel interaction and were fitted by maximum
631 likelihood. Because neither interaction was significant, overall PD-to-HC ratios were obtained
632 from additive models without the interaction. Exploratory within-vowel contrasts were estimated
633 from the full interaction models, with FDR correction applied separately across the six contrasts
634 for jitter and shimmer.

635 Sex-stratified tVSA and VAI effects in Cohort 1 were estimated using age-adjusted linear
636 models within men and women. tVSA was natural-log transformed and reported as a PD-to-HC
637 geometric mean ratio; VAI was reported as an adjusted mean difference. Confidence intervals
638 were obtained by nonparametric bootstrap [54]. Effect modification was evaluated in models
639 containing group, sex, age, and a group-by-sex interaction.

640 Acoustic-clinical associations were examined separately in each cohort using age- and sex-
641 adjusted linear regression with HC3 robust standard errors. Acoustic predictors and clinical
642 outcomes were standardized before modelling. In Cohort 1, tVSA and VHI-30 were transformed
643 as ln(value) and ln(score + 1), respectively. In Cohort 2, jitter, shimmer, tVSA, mean pause
644 length, and mean run length were natural-log transformed; silent-pause count and VHI-10 were
645 transformed as ln(value + 1). FDR correction was applied separately across the prespecified
646 acoustic measures for each clinical outcome. In Cohort 2, MPT and mean pause length were also
647 entered simultaneously to assess their independent associations with MDS-UPDRS Part III and
648 VHI-10; FDR correction was applied across the four focal acoustic coefficients. Cross-cohort
649 comparisons of the seven shared measures were descriptive and based on direction of the
650 standardized coefficients only.

Within-patient acoustic intercorrelations were evaluated separately in the PD subgroup of Cohort 1 and in Cohort 2 using Spearman rank correlation. FDR correction was applied across all unique pairwise correlations within each cohort.

In Cohort 2, associations between the fourteen acoustic measures and eight cognitive outcomes were evaluated using partial Spearman correlations after rank transformation, with adjustment for age, sex, and education level; education was represented categorically. Confidence intervals were calculated using Fisher z transformation and two-sided P values used 43 residual degrees of freedom. FDR correction was applied separately across the fourteen acoustic measures for each cognitive outcome. The capped TMT-A and TMT-B analysis values were used.

To evaluate the influence of the prespecified 360-s ceiling, analyses of TMT-A and TMT-B were repeated after excluding participants who reached the ceiling for the corresponding task. Partial Spearman correlations were recalculated with adjustment for age, sex, and education, using the same outcome-specific FDR procedure.

Exploratory hierarchical rank-based linear regression was then used to assess the incremental contribution of mean pause length to MMSE and RAVLT immediate recall. All continuous variables were rank transformed. Model 1 included age, sex, and education; Model 2 additionally included disease duration and MDS-UPDRS Part III; and Model 3 additionally included mean pause length. Education was entered categorically. Model fit was summarized using R-squared, adjusted R-squared, overall F tests, and Akaike information criterion. Incremental variance was evaluated using the change in unadjusted R-squared and nested-model F tests.

Exploratory discrimination analyses were performed in Cohort 1. Apparent receiver operating characteristic curves were generated for DSI, MPT, shimmer /a/, and natural-log-transformed tVSA, with scores oriented so that larger values indicated a greater probability of PD. Areas under the curve (AUCs) and 95% confidence intervals were estimated using 5000 stratified nonparametric bootstrap resamples [54]; operating points were selected using the Youden index. Three logistic regression models were internally evaluated: a demographic model containing age and sex, an acoustic model containing DSI, shimmer /a/, and log-transformed tVSA, and a combined model containing all five predictors. Fixed L2-regularized logistic regression (C = 1.0) was used. Standardization was fitted within each training fold, and out-of-fold predictions were obtained from 200 repetitions of stratified fivefold cross-validation using scikit-learn [55]. Performance was summarized by the mean cross-validated AUC and its 2.5th-

97.5th percentile empirical range. These analyses were considered exploratory internal validation and not external diagnostic validation.

Analyses were performed using IBM SPSS Statistics version 26.0 and Python version 3.12. Python analyses used standard scientific-computing, statistical-modelling, and machine-learning libraries. Figures were generated from the final analysis datasets using reproducible Python scripts.

4.8 Code availability

The Python scripts used for data preprocessing, statistical modelling, false-discovery-rate correction, bootstrap estimation, repeated cross-validation, and generation of the main and supplementary figures are available from the corresponding author (chinazxw002@cmc.edu.cn) on reasonable request. The code package includes variable definitions and instructions required to reproduce the reported analyses. No proprietary custom algorithm was used.

4.9 Use of generative artificial intelligence

OpenAI ChatGPT was used to assist with manuscript organization, language editing, and the drafting and debugging of selected Python analysis and visualization scripts. The tool was not used to determine participant eligibility, collect data, make autonomous research decisions, or verify study results. All statistical procedures, code, numerical outputs, references, figures, and final wording were independently reviewed and approved by the authors, who take full responsibility for the content of the manuscript.

Data availability

De-identified participant-level clinical and derived acoustic data, the data dictionary, and source data underlying the tables and figures are available from the corresponding author, Xianwei Zou (chinazxw002@cmc.edu.cn), upon reasonable request. Access is available to qualified researchers for non-commercial academic verification or secondary analysis, subject to approval by the institutional ethics committee and completion of a data-use agreement. Requests should include the research question, analysis plan, and proposed data-security measures and will normally be reviewed within 30 days. Raw voice recordings are not publicly deposited because voice recordings may remain personally identifiable; access to such recordings will be considered only where permitted by participant consent and applicable ethical and legal requirements.

Acknowledgements

The authors thank all participants for their time and contribution and the clinical staff who assisted with participant assessment and speech recording.

Author contributions

H.X. contributed to conceptualization, methodology, formal analysis, visualization, and writing the original draft. Z.Z. contributed to investigation, data curation, formal analysis, validation, and manuscript review and editing. J.Z. contributed to investigation, project administration, formal analysis, and manuscript review and editing. Y.W. contributed to conceptualization, investigation, data collection, and data validation. Z.W. contributed to participant recruitment, clinical assessment, and data curation. H.B. and M.W. contributed to formal analysis, interpretation of acoustic findings, and critical revision of the manuscript. T.H. contributed to supervision, interpretation of the findings, and critical revision of the manuscript. X.Z. contributed to conceptualization, supervision, project administration, funding acquisition, and manuscript review and editing. T.H. and X.Z. jointly supervised this work. H.X., Z.Z., and J.Z. contributed equally to this work and share first authorship. All authors reviewed and approved the submitted version and accept responsibility for their contributions.

Funding

This work was supported by the Sichuan Provincial Health Commission Medical Science and Technology Project (Grant No. 25LCYJ41), the Chengdu Municipal Health Commission Medical Research Project (Grant No. 2025125), and the Open Fund of the Key Laboratory of Smart Healthcare and Elderly Health Management (Grant No. ZHYYJCYB2522). The funders had no role in study design, participant recruitment, data collection, data analysis, interpretation of the results, preparation of the manuscript, or the decision to submit the work for publication.

Competing interests

The authors declare no competing interests.

References

[1] Bloem, B. R., Okun, M. S. & Klein, C. Parkinson's disease. *Lancet* **397,** 2284–2303 (2021).
[2] Atalar, M. S., Oguz, O. & Genc, G. Hypokinetic dysarthria in Parkinson's disease: a narrative review. *Sisli Etfal Hastan Tip Bul.* **57,** 163–170 (2023).
[3] Rohl, A. et al. Speech dysfunction, cognition, and Parkinson's disease. *Prog. Brain Res.* **269,** 153–173 (2022).

[4] Mu, L. et al. Mechanisms of swallowing, speech and voice disorders in Parkinson's disease: literature review with our first evidence for Periperal nervous system involvement. *Dysphagia* **39,** 1001–1012 (2024).
[5] Rusz, J. et al. Defining speech subtypes in de novo Parkinson disease: response to long-term levodopa therapy. *Neurology* **97,** e2124–e2135 (2021).
[6] Sackley, C. M. et al. Lee Silverman voice treatment versus NHS speech and language therapy versus control for dysarthria in people with Parkinson's disease (PD COMM): pragmatic, UK based, multicentre, three arm, parallel group, unblinded, randomised controlled trial. *BMJ* **386,** e078341 (2024).
[7] Cao, F., Vogel, A. P., Gharahkhani, P. & Renteria, M. E. Speech and language biomarkers for Parkinson's disease prediction, early diagnosis and progression. *npj Parkinsons Dis.* **11,** 57 (2025).
[8] Ngo, Q. C. et al. Computerized analysis of speech and voice for Parkinson's disease: a systematic review. *Comput. Methods Programs Biomed.* **226,** 107133 (2022).
[9] Rusz, J., Krack, P. & Tripoliti, E. From prodromal stages to clinical trials: the promise of digital speech biomarkers in Parkinson's disease. *Neurosci. Biobehav. Rev.* **167,** 105922 (2024).
[10] Rusz, J., Cmejla, R., Ruzickova, H. & Ruzicka, E. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *J. Acoust. Soc. Am.* **129,** 350–367 (2011).
[11] Diao, Y. et al. Acoustic assessment in Mandarin-speaking Parkinson's disease patients and disease progression monitoring and brain impairment within the speech subsystem. *npj Parkinsons Dis.* **10,** 115 (2024).
[12] Duranovic, M., Avdic, Z. & Klebic, J. Voice and speech characteristics in patients with Parkinson's disease. *J. Voice* https://doi.org/10.1016/j.jvoice.2026.02.007 (2026).
[13] Mou, Z. et al. Acoustic properties of vowel production in Mandarin-speaking patients with Parkinson disease-related hypokinetic dysarthria. *J. Craniofac. Surg.* **36,** e670–e677 (2025).
[14] Wuyts, F. L. et al. The Dysphonia Severity Index: an objective measure of vocal quality based on a multiparameter approach. *J. Speech Lang. Hear. Res.* **43,** 796–809 (2000).
[15] Skodda, S., Visser, W. & Schlegel, U. Vowel articulation in Parkinson's disease. *J. Voice* **25,** 467–472 (2011).
[16] Sapir, S., Ramig, L. O., Spielman, J. L. & Fox, C. Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech. *J. Speech Lang. Hear. Res.* **53,** 114–125 (2010).
[17] DeSilva, G. S. et al. Variability in vowel space in Parkinson's disease: associations with cognitive and motor impairment. *J. Speech Lang. Hear. Res.* **67,** 3566–3578 (2024).
[18] Abur, D., Cushman, M., Dunsmuir, C. J. & Stepp, C. E. Comparing token-based and trajectory-based vowel space metrics for assessing speech in Parkinson's disease. *PLoS One* **21,** e0353680 (2026).
[19] Mollaei, F., Evans, H. & Pool, A. Pause characteristics of sentence production in Parkinson's disease: insights from sentence complexity and length. *PLoS One* **21,** e0330808 (2026).
[20] Knowles, T., Adams, S. G. & Jog, M. Speech rate mediated vowel and stop voicing distinctiveness in Parkinson's disease. *J. Speech Lang. Hear. Res.* **64,** 4096–4123 (2021).
[21] Liu, Y. et al. Acoustic prosodic parameters associated with Parkinson's disease cognitive impairment. *Parkinsonism Relat. Disord.* **132,** 107306 (2025).
[22] D'Ascanio, S. et al. Narrative discourse production in Parkinson's disease: decoupling the role of cognitive-linguistic and motor speech changes. *Heliyon* **9,** e18633 (2023).
[23] Iyer, A. et al. A machine learning method to process voice samples for identification of Parkinson's disease. *Sci. Rep.* **13,** 20615 (2023).
[24] Pinto, S. et al. Do acoustic characteristics of dysarthria in people with Parkinson's disease differ across languages? *J. Speech Lang. Hear. Res.* **67,** 2822–2841 (2024).
[25] Piña Méndez, Á. et al. Speech's syllabic rhythm and articulatory features produced under different auditory feedback conditions identify Parkinsonism. *Sci. Rep.* **14,** 15787 (2024).
[26] Johansson, I.-L., Samuelsson, C. & Müller, N. Consonant articulation acoustics and intelligibility in Swedish speakers with Parkinson's disease: a pilot study. *Clin. Linguist. Phon.* **37,** 845–865 (2023).
[27] Tracey, B. et al. Towards interpretable speech biomarkers: exploring MFCCs. *Sci. Rep.* **13,** 22787 (2023).
[28] Favaro, A. et al. Unveiling early signs of Parkinson's disease via a longitudinal analysis of celebrity speech recordings. *npj Parkinsons Dis.* **10,** 207 (2024).
[29] Cai, W. et al. Subthalamic nucleus-language network connectivity predicts dopaminergic modulation of speech function in Parkinson's disease. *Proc. Natl Acad. Sci. USA* **121,** e2316149121 (2024).
[30] Contreras-Ruston, F. et al. Sensory feedback in Parkinson disease voice production: a systematic review. *J. Voice* https://doi.org/10.1016/j.jvoice.2025.02.039 (2025).
[31] Thies, T. et al. Automatic speech analysis combined with machine learning reliably predicts the motor state in people with Parkinson's disease. *npj Parkinsons Dis.* **11,** 105 (2025).
[32] World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* **310,** 2191–2194 (2013).

[33] Postuma, R. B. et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov. Disord.* **30,** 1591–1601 (2015).
[34] Folstein, M. F., Folstein, S. E. & McHugh, P. R. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12,** 189–198 (1975).
[35] Jost, S. T. et al. Levodopa dose equivalency in Parkinson's disease: updated systematic review and proposals. *Mov. Disord.* **38,** 1236–1252 (2023).
[36] Hoehn, M. M. & Yahr, M. D. Parkinsonism: onset, progression and mortality. *Neurology* **17,** 427–442 (1967).
[37] Fahn, S. & Elton, R. L. Unified Parkinson's Disease Rating Scale. In *Recent Developments in Parkinson's Disease* Vol. 2 (eds Fahn, S., Marsden, C. D., Calne, D. B. & Goldstein, M.) 153–163, 293–304. (Macmillan Healthcare Information, 1987).
[38] Jacobson, B. H. et al. The Voice Handicap Index (VHI): development and validation. *Am. J. Speech Lang. Pathol.* **6,** 66–70 (1997).
[39] Enderby, P. Frenchay Dysarthria Assessment. *Br. J. Disord. Commun.* **15,** 165–173 (1980).
[40] Yesavage, J. A. et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J. Psychiatr. Res.* **17,** 37–49 (1982).
[41] Goetz, C. G. et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov. Disord.* **23,** 2129–2170 (2008).
[42] Hamilton, M. The assessment of anxiety states by rating. *Br. J. Med. Psychol.* **32,** 50–55 (1959).
[43] Hamilton, M. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* **23,** 56–62 (1960).
[44] Rosen, C. A., Lee, A. S., Osborne, J., Zullo, T. & Murry, T. Development and validation of the Voice Handicap Index-10. *Laryngoscope* **114,** 1549–1556 (2004).
[45] Nasreddine, Z. S. et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53,** 695–699 (2005).
[46] Schmidt, M. *Rey Auditory Verbal Learning Test: A Handbook* (Western Psychological Services, 1996).
[47] Shulman, K. I. Clock-drawing: is it the ideal cognitive screening test? *Int. J. Geriatr. Psychiatry* **15,** 548–561 (2000).
[48] Reitan, R. M. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept. Mot. Skills* **8,** 271–276 (1958).
[49] Patel, R. R. et al. Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *Am. J. Speech Lang. Pathol.* **27,** 887–905 (2018).
[50] Svec, J. G. & Granqvist, S. Guidelines for selecting microphones for human voice production research. *Am. J. Speech Lang. Pathol.* **19,** 356–368 (2010).
[51] Boersma, P. Praat, a system for doing phonetics by computer. *Glot Int.* **5,** 341–345 (2001).
[52] Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).
[53] MacKinnon, J. G. & White, H. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *J. Econometrics* **29,** 305–325 (1985).
[54] Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (Chapman & Hall, 1993).
[55] Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12,** 2825–2830 (2011).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx