

ORIGINAL RESEARCH ARTICLE

Digital voice biomarkers for Parkinson's disease: A study on sustained vowel analysis in the Russian population

Ildar Khasanov^{1,2}  and Diana Khasanova^{2*} ¹Department of Biomedical Engineering and Artificial Intelligence in Biotechnical Systems, Institute of Artificial Intelligence, Robotics, and Systems Engineering (AIRSE), Kazan Federal University, Kazan, Tatarstan Republic, Russia²Department of Digital Technologies in Healthcare, Faculty of General Medicine, Kazan State Medical University, Kazan, Tatarstan Republic, Russia

Abstract

Voice abnormalities are common in Parkinson's disease (PD), but the extent to which language-robust acoustic markers capture PD dysphonia in real-world clinical recording conditions and whether they are confounded by sex, language background, or medication state remains uncertain. This study aims to quantify PD-controlled differences in sustained-vowel acoustics in a Russian cohort, evaluate sex and language effects (Russian, Russian-Tatar bilinguals, and exploratory Tatar subgroup), and assess the robustness to clinical covariates. Cross-sectional data from the BRAINPHONE project were analyzed ($n = 201$; PD = 109; controls = 92). Participants produced sustained/a:/vowels in routine clinics (≥ 16 kHz, 32-bit, wav). Acoustic features included perturbation (jitter and shimmer), cepstral/noise measures (cepstral peak prominence; harmonic-to-noise ratio; glottal-to-noise excitation ratio), and pitch metrics. Group contrasts used the Mann-Whitney U test and false discovery rate (FDR). Robust models adjusted for age, sex, and language; prespecified interactions probed diagnosis \times sex/language. Spearman correlations related acoustics to Movement Disorder Society-Unified Parkinson's Disease Rating Scale III, Hoehn and Yahr, disease duration, and medication variables. PD showed higher perturbation and lower cepstral/noise measures than controls (all $q \leq 0.01$ effects were consistent in females and males and replicated in Russian monolinguals and Russian-Tatar bilinguals with the Tatar monolingual subgroup being directionally similar. Covariate-adjusted models retained significant PD effects. Acoustic-clinical correlations were small ($|\rho| \leq 0.21$) and did not survive FDR. In real-world clinical recordings of sustained vowels, CPPS, GNE, and shimmer provide robust, language-tolerant, medication-insensitive markers of PD dysphonia, supporting use as a complementary digital biomarker for telemedicine and longitudinal monitoring.

Keywords: Parkinson's disease; Voice; Sustained vowel; Digital biomarker; Language influence; Acoustic features; Sex differences; Bilingualism

***Corresponding author:**Khasanova Diana
(diana.khasanova1987@gmail.com)

Citation: Khasanov I, Khasanova D. Digital voice biomarkers for Parkinson's disease: A study on sustained vowel analysis in the Russian population. *Adv Neurol.* 2026;5(2):025180042.
doi: 10.36922/AN025180042

Received: April 30, 2025**Revised:** October 2, 2025**Accepted:** October 27, 2025**Published online:** November 11, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Parkinson's disease (PD) is one of the most common neurodegenerative disorders, affecting 2–3% of people aged ≥ 65 years worldwide, with prevalence expected to rise.^{1,2} In Russia, regional studies estimate 120–200 cases/100,000 inhabitants, but the absence of a national registry limits the accuracy of epidemiological data.^{3–6} Access to specialized movement disorder care also remains uneven, particularly in rural areas. Diagnosis today relies mainly on clinical assessment using the International Parkinson and Movement Disorder Society (MDS) Clinical Diagnostic Criteria,^{7,8} which, while validated, depend on subjective expert evaluation. This reliance introduces variability in early or atypical presentations, while advanced diagnostic tools, such as dopamine transporter imaging and cerebrospinal fluid assays, remain costly and limited in availability.

Digital biomarkers offer a solution for scalable, cost-effective PD detection and monitoring, particularly in resource-constrained healthcare systems. Defined as characteristics measured through digital health technologies to indicate biological or pathogenic processes, digital biomarkers enable remote assessment, reducing reliance on sporadic clinical visits.^{9,10} Among them, voice analysis is especially promising. It is non-invasive, requires minimal infrastructure, and can be integrated into telemedicine platforms.^{11,12} Vocal impairments in PD, collectively termed hypokinetic dysarthria, result from motor symptoms such as rigidity and bradykinesia, manifesting as reduced pitch variability, phonation instability, monotony, and breathiness.¹³ Acoustic analysis of sustained vowels has consistently revealed differences between PD patients and controls, with parameters such as jitter, shimmer, cepstral measures, and spectral features showing diagnostic value across languages.^{14–16} For example, Rusz *et al.*¹⁶ demonstrated that even in early, untreated PD in Czech speakers, fundamental frequency variation and other acoustic measures can reliably differentiate patients from controls. More recent work has extended to other non-English contexts. Dudek *et al.*¹⁷ showed that in the Polish cohort, spectral characteristics, mel-frequency cepstral coefficients, and shimmer are among the most informative features, with machine learning models achieving high classification performance. Likewise, Sousa *et al.*,¹⁸ in a presumably mixed-language European sample, showed that prosodic, respiratory, and spectral domains change with medication state, underscoring that subtle speech changes are measurable even beyond native English contexts. Cross-language model generalization has been explored recently. For example, Siniukov *et al.*¹⁹ showed that models trained with self-supervised learning,

adapting across multiple datasets in different languages (English, Italian, and Spanish), can maintain sensitivity and specificity above 90%.

Despite this growing body of cross-linguistic work, Russian-speaking patients remain underrepresented. Russia's linguistic landscape, including Russian (East Slavic) and Tatar (Turkic) languages, with many bilingual speakers in regions such as Tatarstan, may modulate vocal biomarkers due to distinct phonetic structures.²⁰ For example, Tatar's vowel harmony and phonetic inventory could influence laryngeal control differently than Russian's prosodic patterns. Bilingualism may further affect vocal characteristics. These linguistic differences, combined with bilingualism itself, may influence the acoustic manifestation of PD-related dysarthria, but this remains unexplored.

This study aims to address this gap by evaluating sustained vowel phonation in Russian- and Tatar-speaking PD patients, including bilinguals. We focused on acoustic measures, such as fundamental frequency, perturbation indices, and spectral/cepstral features, and examined their relationships with clinical severity scores, such as Unified Parkinson's Disease Rating Scale (UPDRS) Part III (motor) score. To our knowledge, this is the first systematic investigation of Russian-speaking patients and exploratory data on Tatar speakers, contributing to the development of inclusive, language-sensitive digital biomarkers for PD that are relevant both locally and globally.

2. Materials and methods

2.1. Participants

This investigation is part of the ongoing longitudinal BRAINPHONE project, a multicenter digital biomarker initiative focusing on neurodegenerative disorders in the Russian population.^{21–23} The project systematically collects speech samples from patients with movement disorders and healthy controls. The present analysis used baseline cross-sectional data obtained during routine outpatient visits from the project's expanding database.

Data were obtained from participants recruited at the Movement Disorders Center in Kazan, Tatarstan. PD diagnosis was confirmed by a movement disorder specialist with over 12 years of experience, using the MDS Clinical Diagnostic Criteria for PD.⁷ All participants were native speakers of Russian, Tatar, or both (bilingual). Language status was self-reported, with bilinguals indicating daily use of both languages. No standardized proficiency testing was performed, which represents a methodological limitation.

All included PD patients were diagnosed as Hoehn and Yahr stages 3–4, reflecting the advanced outpatient

profile within the recruitment site. Early-stage patients are underrepresented in this cohort but are a focus of ongoing recruitment in the BRAINPHONE project. Both patients with established and those with *de novo* PD were eligible. All patients reported their PD status: year of disease onset, duration of illness, medications taken, dose, and time since last administration. When applicable, the motor state at the time of testing was recorded as “on” or “off.” Exclusion criteria were atypical or secondary Parkinsonism, speech disorders unrelated to PD (e.g., post-stroke aphasia, non-Parkinsonian dysarthria, and vocal cord dystonia), recent surgery (<6 months), severe hearing impairment, and acute respiratory infections.

The final cohort comprised Russian-speaking PD patients ($n = 50$) and controls ($n = 52$), Tatar-speaking PD patients ($n = 7$) and controls ($n = 5$), and bilingual Russian–Tatar PD patients ($n = 52$) and controls ($n = 35$). Given the limited number of Tatar-speaking participants, analyses in this subgroup are presented as exploratory.

All participants provided written informed consent before enrollment, in accordance with the Declaration of Helsinki.²⁴ Data were anonymized and stored securely following institutional guidelines.

2.2. Procedures

Participants were instructed to produce a sustained/a:/vowel for at least 5 s at a comfortable pitch and loudness. Recordings were conducted in routine clinical settings using standard microphones (minimum sampling rate: 16 kHz; bit depth: 32-bit) and saved in uncompressed.wav format. This approach prioritized ecological validity while ensuring sufficient audio quality for analysis.

Motor symptom severity was assessed with MDS-UPDRS Part III during routine clinical visits. Most assessments were performed in the “on-medication” state, though the interval since the last levodopa dose varied among patients. This variability was documented (in the forms of dose and time since intake) but not standardized, representing both a limitation and a reflection of real-world clinical practice.

Voice samples were analyzed using Praat software (version 6.2.23). Acoustic features were extracted across four domains: (i) fundamental frequency characteristics (mean, minimum, and maximum pitch in Hz, and pitch variability through standard deviation and semitone range); (ii) perturbation measures (jitter local and five-point Period Perturbation Quotient [PPQ5] variants, and shimmer local and decibel [dB] variants); (iii) spectral and cepstral parameters (cepstral peak prominence [CPPS], harmonic-to-noise ratio [HNR], HNR in dB, spectral tilt measures [H1–H2 difference], and glottal-to-noise

excitation ratio [GNE]); and (iv) additional features (including spectral slope parameters [e.g., floor, ceiling, slope, and tilt], high-frequency noise energy, power spectral density distribution, and voice break fraction).

The sustained vowel/a:/task was chosen because it provides a controlled phonatory signal that minimizes the confounding influence of articulation and sentence-level prosody, thereby offering a stable basis for acoustic analysis.^{25,26} While this task reduces cross-linguistic variability compared to connected speech, vowel production still reflects language-specific phonetic patterns (e.g., vowel harmony or typical pitch range). Prior studies have demonstrated that sustained vowels are highly sensitive to PD-related dysphonia, particularly through perturbation and cepstral measures.^{26,27} This makes the task both a robust starting point for universal biomarker development and a useful window into how linguistic background may modulate vocal impairment.

All methods and procedures were developed and standardized within the framework of the BRAINPHONE project.

2.3. Statistical analysis

All analyses were conducted in Python 3.9 (NumPy, Pandas, SciPy, statsmodels, and matplotlib). Two-sided tests were used throughout with an α of 0.05. Continuous variables were summarized as mean \pm standard deviation (SD); categorical variables as n (%). Missing data were not imputed; tests used pairwise complete observations.

Given non-normal distributions on most acoustic variables (Shapiro–Wilk), PD versus control comparisons for continuous measures used the Mann–Whitney U test. Effect sizes are reported as the rank-biserial correlation (r_{rb}) computed from U ; 95% confidence intervals (CIs) are provided where indicated. Categorical variables (e.g., sex) were compared with χ^2 tests. For multi-level factors (e.g., language strata), we used Kruskal–Wallis tests, with Dunn’s pairwise contrasts where relevant.

To limit false positives while preserving readability for clinicians, we controlled the false discovery rate (FDR) using the Benjamini–Hochberg procedure within prespecified feature families:

- Perturbation: Jitter (local, PPQ5), shimmer (local, dB)
- Cepstral/noise: CPPS, HNR (dB), HNR, GNE
- Pitch: Mean F0, short-term F0 variability (SD), semitone range (spectral features, if reported, were adjusted within a spectral family)

We reported raw p - and q -values (FDR-adjusted): $q \leq 0.05$ was considered significant; $0.05 < q \leq 0.10$ was treated as exploratory/suggestive (used only to frame

trends, not as confirmatory evidence). Family definitions and the use of within-family FDR are stated *a priori* to align with reviewer guidance.

To examine sex and language effects without inflating type I error, we repeated PD–control contrasts within the sex (female and male) and within the language strata (Russian, Russian–Tatar bilingual, and Tatar). For each stratum, we used Mann–Whitney *U* tests with within-stratum FDR across the same core features. Given the limited number of participants in the Tatar subgroup (7 PD and 5 controls), its results are labeled as exploratory.

As a primary robustness check, we fitted ordinary least squares models with heteroscedasticity-consistent (HC3) standard errors for each acoustic outcome, with diagnosis (PD vs. control) as the primary predictor, and age, sex, and language as covariates.

We report adjusted β for diagnosis with 95% CIs and *p*-values. Prespecified interactions (diagnosis \times sex, diagnosis \times language) were probed in sensitivity models; simple effects were interpreted cautiously if interactions were non-significant. As an extended sensitivity analysis, we added Hoehn and Yahr stage, disease duration, levodopa equivalent daily dose (LEDD), individual levodopa dose, and time since last dose to the model; because these are PD-specific, results are presented as conservative checks rather than primary inference.

To address variability in medication state, we repeated the PD vs. healthy control contrasts, restricting PD patients to 90–180 min post-levodopa (typical “ON” window), again using Mann–Whitney *U* tests with the same feature families and FDR policy. In a complementary approach, we refitted the demographic-adjusted models, adding LEDD and time since last dose as covariates to test whether dopaminergic exposure/timing accounts for the observed group differences.

Within the PD group, associations between acoustic features and clinical measures (MDS-UPDRS Part III, MDS-UPDRS III without tremor, Hoehn and Yahr stage, disease duration, LEDD, individual dose, and time since last dose) were evaluated with Spearman's ρ (pairwise complete data). FDR (with the Benjamini–Hochberg procedure) was applied across all feature–clinical pairs. As a sensitivity analysis, we estimated rank-based partial Spearman correlations by rank-residualizing each acoustic and clinical measure for age, sex, LEDD, and time since last dose, and then correlating the residuals. For this prespecified small set of measures (CPPS, GNE, and shimmer vs. MDS-UPDRS Part III \pm tremor), we report unadjusted *p*-values.

Primary numeric results are reported in the main tables. Violin/box plots illustrate key effects by diagnosis, sex, and language, and a correlation heatmap summarizes

acoustic–clinical relationships in PD. Exact *n* for each test is shown in tables/figure captions; all tests are two-tailed.

3. Results

3.1. Clinical results

3.1.1. PD and control

The study cohort comprised 201 participants: 109 patients with clinically confirmed PD and 92 age- and sex-matched healthy controls. Patients were at advanced stages of the disease, with Hoehn and Yahr stage 3 (*n* = 87, 80%) or stage 4 (*n* = 22, 20%). The mean age of the PD group was 68.1 ± 7.5 years compared with 66.7 ± 11.2 years in controls. This difference was not statistically significant (Mann–Whitney *U* = 4,722, *p* = 0.29, rank-biserial *r* = -0.07 , 95% CI = -0.20 – 0.06). Sex distribution did not differ between groups (χ^2 = 0.84, *p* = 0.36). Within the PD cohort, there were no significant differences in demographic or clinical characteristics between Hoehn and Yahr stages 3 and 4 (all *p* > 0.05). Motor symptom severity was quantified with the MDS-UPDRS Part III, where PD patients showed a mean score of 48.0 ± 15.2 , consistent with advanced disease.

Pharmacological treatment was also characterized: LEDD averaged 677.1 ± 268.9 mg. The mean individual levodopa dose most recently taken before assessment was 183.5 ± 51.5 mg, with a mean elapsed time of 331 ± 236 min. These data reflect heterogeneous treatment states at the time of testing, which were recorded for use in adjusted analyses. Overall demographic and clinical characteristics are summarized in Table 1.

3.1.2. Sex differences

Across the whole cohort, there were 97 males (48%) and 104 females (52%), with similar distributions in PD patients (57/52) and controls (40/52; χ^2 = 0.84, *p* = 0.36). Male and female controls were comparable in age (males: 66.1 ± 10.7 years vs. females: 67.2 ± 11.5 years; *U* = 971, *p* = 0.58, *r* = -0.05). Likewise, within the PD group, age did not differ significantly between sexes (males: 68.9 ± 6.8 years vs. females: 67.2 ± 8.1 years; *U* = 1,402, *p* = 0.21, *r* = -0.12 , 95% CI = -0.29 – 0.05).

In the PD group, disease duration did not differ by sex (males: 7.5 ± 4.9 years vs. females: 6.9 ± 4.3 years; *U* = 1,426, *p* = 0.33, *r* = -0.09). Motor symptom severity, as measured using MDS-UPDRS Part III, showed no sex-based differences (males: 47.3 ± 15.4 vs. females: 48.8 ± 15.0 ; *U* = 1,469, *p* = 0.42, *r* = -0.07). LEDD was comparable between sexes (males: 690.4 ± 271.1 mg vs. females: 662.8 ± 267.3 mg; *U* = 1,473, *p* = 0.45, *r* = -0.06). No differences were detected for individual levodopa dose or time since last dose (all *p* > 0.20).

Table 1. Demographic and clinical characteristics of participants

Parameter	Group	n (%)	Mean±SD	95% CI
Age (years)	PD	109 (54.2)	68.12±7.48	66.70–69.54
	Control	92 (45.8)	66.66±11.16	64.35–68.97
	PD males	39 (35.8)	67.00±8.52	64.24–69.76
	PD females	70 (64.2)	68.74±6.81	67.12–70.37
	Control males	33 (35.9)	66.52±12.39	62.12–70.91
	Control females	59 (64.1)	66.75±10.52	64.00–69.49
	PD Russian	50 (45.9)	68.94±7.42	66.83–71.05
	PD bilingual	52 (47.7)	67.17±7.67	65.04–69.31
	PD Tatar	7 (6.4)	69.29±6.47	63.30–75.27
	Control Russian	52 (56.5)	68.27±10.63	65.31–71.23
	Control bilingual	35 (38.0)	63.89±12.04	59.75–68.02
	Control Tatar	5 (5.5)	69.40±7.16	60.51–78.29
PD duration (years)	PD	109	7.23±4.61	6.35–8.11
	PD males	39	7.77±5.35	6.03–9.51
	PD females	70	6.93±4.15	5.94–7.92
	PD Russian	50	7.46±5.26	5.96–8.96
	PD bilingual	52	7.42±4.05	6.30–8.55
	PD Tatar	7	4.14±2.34	1.98–6.31
LEDD	PD	109	677.10±268.92	621.70–732.49
	PD males	39	711.43±264.93	620.36–802.50
	PD females	70	656.38±271.48	584.97–727.79
	PD Russian	50	656.55±251.66	578.11–734.99
	PD bilingual	52	683.42±265.07	604.64–762.21
	PD Tatar	7	791.50±449.98	232.86–1,350.14
Levodopa dose before assessment (mg)	PD	109	183.54±51.54	172.98–194.10
	PD males	39	187.50±48.60	170.79–204.21
	PD females	70	181.19±53.47	167.26–195.13
	PD Russian	50	175.31±52.47	159.16–191.47
	PD bilingual	52	188.59±48.72	174.11–203.07
	PD Tatar	7	207.90±66.23	125.67–290.13
Time from the last dose (minutes)	PD	109	331.41±236.33	281.33–381.49
	PD males	39	334.26±249.95	242.59–425.93
	PD females	70	329.86±230.85	268.58–391.14
	PD Russian	50	292.97±209.81	224.94–361.01
	PD bilingual	52	362.44±256.43	285.35–439.54
	PD Tatar	7	357.00±251.19	110.80–603.20
MDS-UPDRS, Part III	PD	109	48.04±15.17	45.06–51.02
	PD males	39	48.95±15.94	43.63–54.26
	PD females	70	47.52±14.81	43.85–51.19
	PD Russian	50	48.23±15.00	43.66–52.79
	PD bilingual	52	49.06±15.85	44.60–53.52
	PD Tatar	7	39.43±8.42	31.64–47.22

Notes: Proportions are calculated relative to total PD for subgroups; full data available upon request.

Abbreviations: CI: Confidence interval; SD: Standard deviation; LEDD: Levodopa equivalent daily dose; MDS-UPDRS, Part III: Unified Parkinson's Disease Rating Scale Part III (motor examination); PD: Parkinson's disease.

Taken together, male and female PD patients were clinically similar in age, disease duration, severity, and treatment exposure.

3.1.3. Linguistic differences

Linguistic diversity was self-reported, comprising 102 Russian monolinguals, 12 Tatar monolinguals, and 87 Russian–Tatar bilinguals. Analyses were stratified by diagnosis (PD vs. control) within each language group.

There were notable patterns in clinical presentation and treatment among PD patients. Russian-speaking PD patients ($n = 50$, 45.9%) had a mean age of 68.9 ± 7.4 years and disease duration of 7.5 ± 5.3 years, comparable to Russian–Tatar bilinguals ($n = 52$, 47.7%; age: 67.2 ± 7.7 years, duration: 7.4 ± 4.1 years). Tatar monolinguals ($n = 7$, 6.4%) had a mean age of 69.3 ± 6.5 years and shorter disease duration (4.1 ± 2.3 years), although the small sample size limits the reliability of this observation, and the difference was not significant. Medication patterns differed across groups, with Tatar speakers receiving higher LEDD (791 ± 450 mg) compared to Russian speakers (657 ± 252 mg) and bilinguals (683 ± 265 mg). Kruskal–Wallis tests showed no significant differences in LEDD between groups ($p=0.56$). Motor symptom severity was lowest in Tatar speakers (MDS-UPDRS Part III: 39.4 ± 8.4) compared with Russian (48.2 ± 15.0) and bilingual (49.1 ± 15.9) patients, also without a significant difference ($p=0.29$).

Taken together, Russian-speaking, bilingual, and Tatar-speaking participants were clinically similar with respect to demographic characteristics, disease severity, and treatment exposure. The limited size of the Tatar subgroup limits statistical power, and therefore, findings in this group are considered exploratory.

3.2. Acoustic results

3.2.1. PD and control

Acoustic analysis of sustained vowel phonation revealed robust group differences between PD patients ($n = 109$) and healthy controls ($n = 92$) across perturbation, cepstral, and noise-related domains (Table 2, Figure 1). Group comparisons were performed using Mann–Whitney U tests due to non-normal distributions (Shapiro–Wilk $p < 0.05$ for most variables). Effect sizes are reported as rank-biserial correlations with 95% CIs, and q -values reflect FDR correction within feature families.

In perturbation measures, both jitter and shimmer were consistently elevated in PD. Jitter local was higher in PD patients (0.81 ± 0.55 vs. 0.65 ± 0.56 ; $U = 6,341$, $p=0.0008$, $r = 0.28$, $q = 0.0011$), as was jitter PPQ5 (0.48 ± 0.36 vs. 0.38

± 0.36 ; $U = 6,454$, $p=0.0003$, $r = 0.30$, $q = 0.0005$). Shimmer local (8.52 ± 3.69 vs. 6.67 ± 3.86 ; $U = 6,697$, $p < 0.0001$, $r = 0.35$, $q = 0.0001$) and shimmer dB (0.78 ± 0.33 vs. 0.60 ± 0.34 ; $U = 6,722$, $p < 0.0001$, $r = 0.35$, $q = 0.0001$) showed large and significant differences.

Cepstral and noise measures were also affected. CPPS was markedly reduced in PD (10.35 ± 2.61 dB vs. 12.35 ± 2.89 dB; $U = 2,989$, $p < 0.0001$, $r = -0.40$, $q < 0.001$). HNR was lower in PD (12.85 ± 4.36 dB vs. 14.89 ± 5.14 dB; $U = 3,561$, $p=0.0006$, $r = -0.28$, $q = 0.0013$), as was HNR (dB) (28.92 ± 5.45 dB vs. 30.99 ± 6.20 dB; $U = 3,968$, $p=0.011$, $r = -0.21$, $q = 0.011$). GNE showed the largest effect (0.65 ± 0.17 vs. 0.76 ± 0.15 ; $U = 2,912$, $p < 0.0001$, $r = 0.42$, $q < 0.001$).

The voices of PD patients were characterized by elevated perturbation (jitter and shimmer), reduced CPPS, and deterioration of noise measures (HNR and GNE). These acoustic alterations were robust to multiple-comparison correction and exhibited moderate to large effect sizes.

By contrast, fundamental frequency and most spectral measures did not differ reliably between groups. Mean pitch did not differ significantly between groups (PD: 169.95 ± 46.65 Hz vs. controls: 173.29 ± 53.38 Hz; $U = 4,760$, $p=0.61$, $r = 0.04$, $q = 0.61$). Pitch variability showed a weak reduction in PD (8.44 ± 14.48 Hz vs. controls: 8.67 ± 15.89 Hz; $U = 5,730$, $p=0.062$, $r = -0.15$, $q = 0.12$) and did not survive correction. Semitone range likewise showed no significant difference ($p=0.67$). Spectral slope was steeper in PD voices (-19.85 ± 4.01 vs. -18.50 ± 3.94 ; $p=0.017$, $r = -0.20$), whereas the H1–H2 difference only showed a trend (1.57 ± 7.55 vs. -0.39 ± 7.05 ; $p=0.058$). Neither effect survived FDR correction. Voice break fraction and power spectral density metrics showed no PD-vs.-control differences (all $q > 0.20$) and are not discussed further.

These results suggest that perturbation and cepstral/noise features, rather than pitch-based measures, provide the most reliable acoustic markers of PD in our cohort.

3.2.2. Sex differences

Our detailed analysis of voice parameters across sex and diagnostic groups revealed pronounced sexual dimorphism in vocal characteristics, which persisted in PD but with notable disease-specific modifications (Figure 2).

The most pronounced sex differences were found in fundamental frequency measures. As expected, control females exhibited significantly higher pitch values than control males across all measures. Mean pitch was 54% higher in females (198.38 ± 45.54 Hz) than in males (128.44 ± 31.86 Hz; $p < 0.001$, $r = 0.64$, $q < 0.001$). This dimorphism was attenuated but preserved in PD patients, with females still showing a 22% higher mean

Table 2. Core acoustic features of sustained vowel phonation in Parkinson's disease and control groups stratified by sex and native language

Strata parameter	Feature	PD			HC			<i>U</i>	<i>p</i>	<i>r</i>	<i>q</i>
		<i>n</i>	Mean	SD	<i>n</i>	Mean	SD				
Total	sv_cppts	109	10.345	2.614	92	12.347	2.888	2,989.5	<0.001	0.404	<0.001
	sv_gne	109	0.645	0.168	92	0.757	0.151	2,912.0	<0.001	0.419	<0.001
	sv_shimmer LocaldB	108	0.781	0.331	92	0.602	0.336	6,722.0	<0.001	−0.353	<0.001
	sv_shimmerLocal	108	8.516	3.690	92	6.669	3.859	6,697.5	<0.001	−0.348	<0.001
	sv_jitterLocal	108	0.809	0.554	92	0.650	0.556	6,341.0	0.0008	−0.276	0.0011
	sv_HNR	108	12.850	4.358	92	14.89	5.142	3,561.0	<0.001	0.283	0.0013
	sv_HNR (dB)	109	28.924	5.448	92	30.992	6.197	3,968.5	0.011	0.209	0.011
	sv_meanPitch	108	169.947	46.649	92	173.290	53.386	4,760.0	0.611	0.042	0.611
	sv_sdPitch	108	8.441	14.483	92	8.674	15.890	5,730.0	0.062	−0.153	0.124
Female	sv_cppts	70	10.135	2.652	59	12.341	2.835	1,171.0	<0.001	0.433	<0.001
	sv_gne	70	0.654	0.172	59	0.774	0.135	1,140.5	<0.001	0.448	<0.001
	sv_shimmerLocaldB	70	0.798	0.335	59	0.564	0.315	2,998.0	<0.001	−0.452	<0.001
	sv_jitterLocal	70	0.844	0.614	59	0.594	0.461	2,809.0	<0.001	−0.360	0.001
	sv_hnrd	70	29.560	5.967	59	33.107	5.752	1,334.5	0.001	0.354	0.001
Male	sv_cppts	39	10.722	2.465	33	12.358	2.937	418.0	0.011	0.350	0.027
	sv_gne	39	0.627	0.163	33	0.726	0.173	397.0	0.005	0.383	0.027
	sv_shimmerLocaldB	38	0.750	0.326	33	0.669	0.365	739.0	0.198	−0.179	0.331
	sv_jitterLocal	38	0.744	0.423	33	0.750	0.681	709.5	0.344	−0.132	0.430
	sv_hnrd	39	27.781	4.199	33	27.211	5.122	693.5	0.576	−0.078	0.576
Russian	sv_cppts	50	10.632	2.511	52	12.414	2.939	827.0	0.002	0.364	0.004
	sv_gne	50	0.663	0.152	52	0.767	0.141	730.0	<0.001	0.438	0.001
	sv_shimmerLocaldB	49	0.748	0.334	52	0.564	0.290	1,723.5	0.007	−0.353	0.004
	sv_jitterLocal	49	0.769	0.402	52	0.625	0.558	1,687.5	0.005	0.295	0.006
	sv_hnrd	50	28.607	5.738	52	31.008	6.274	993.5	0.041	0.236	0.041
Russian and Tatar	sv_cppts	52	10.040	2.749	35	12.344	2.910	453.5	<0.001	0.427	0.006
	sv_gne	52	0.630	0.178	35	0.753	0.161	480.5	0.001	0.393	0.007
	sv_shimmerLocaldB	52	0.809	0.335	35	0.646	0.389	1,060.5	0.007	−0.339	0.017
	sv_jitterLocal	52	0.827	0.650	35	0.683	0.568	979.0	0.083	−0.236	0.073
	sv_hnrd	52	28.841	5.338	35	31.040	6.282	593.5	0.098	0.251	0.071
Tatar	sv_cppts	7	10.563	1.708	5	11.674	1.474	12.0	0.527	0.314	0.755
	sv_gne	7	0.624	0.178	5	0.670	0.130	14.0	0.615	0.200	0.755
	sv_shimmerLocaldB	7	0.811	0.213	5	0.688	0.281	25.5	0.222	−0.457	0.755
	sv_jitterLocal	7	0.953	0.621	5	0.672	0.350	22.5	0.464	−0.238	0.755
	sv_hnrd	7	31.806	1.430	5	30.492	3.594	20.0	0.755	−0.143	0.755

Notes: $q \leq 0.05$ indicates significant; $0.05 < q \leq 0.10$ indicates exploratory.

Abbreviations: cppts: Cepstral peak prominence; gne: Glottal-to-noise excitation ratio; HC: Healthy controls; HNR: Harmonic-to-noise ratio;

HNR (dB): Harmonic-to-noise ratio in dB; *p*: Two-sided unadjusted *P*-value; PD: Parkinson's disease; *q*: False discovery rate-adjusted *P*-value

(Benjamini–Hochberg); *r*: rank-biserial correlation computed from *U*, interpreted as small (≈ 0.1), medium (≈ 0.3), or large (≥ 0.5) effect; sv: Sustained vowel, *U*: Mann–Whitney statistic; SD: Standard deviation.

pitch (181.39 ± 50.75 Hz vs. 148.86 ± 26.41 Hz in males; $p < 0.001$, $r = 0.42$, $q < 0.001$). Maximum pitch followed the same pattern, with a 57% sex difference in controls

(females: 217.06 ± 44.13 Hz vs. males: 137.92 ± 34.82 Hz) that narrowed to 24% in PD patients (females: 199.46 ± 52.22 Hz vs. males: 160.93 ± 26.50 Hz). Thus, PD affects

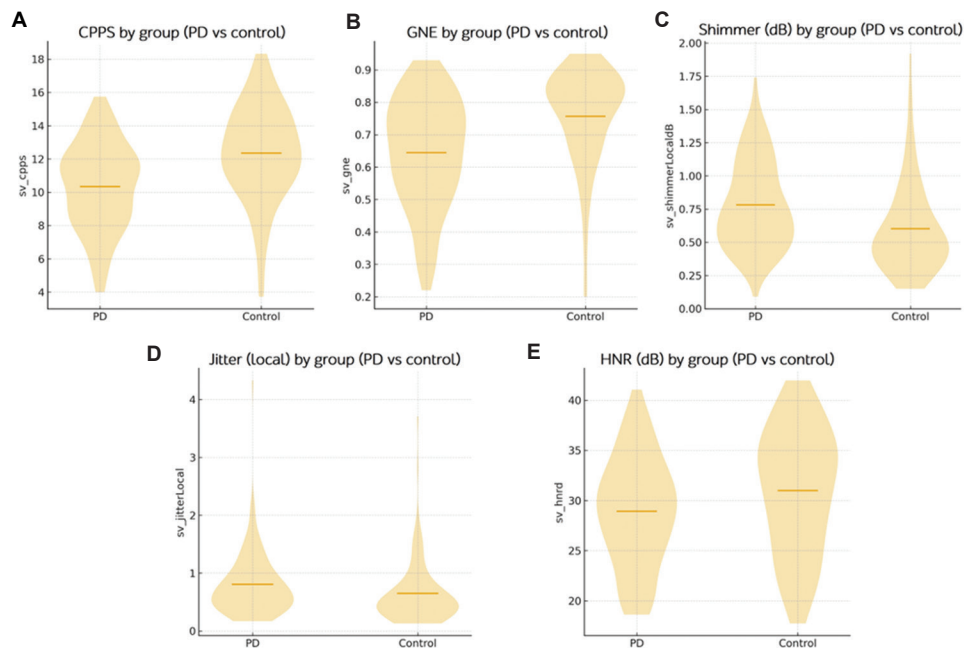


Figure 1. Violin plots of acoustic features in PD and control groups: (A) CPPS, (B) GNE, (C) shimmer (dB), (D) jitter (local), and (E) HNR (dB). Abbreviations: Control: Healthy controls; CPPS: Cepstral peak prominence; GNE: Glottal-to-noise excitation ratio; HNR (dB): Harmonic-to-noise ratio in dB; PD: Parkinson's disease; sv: Sustained vowel.

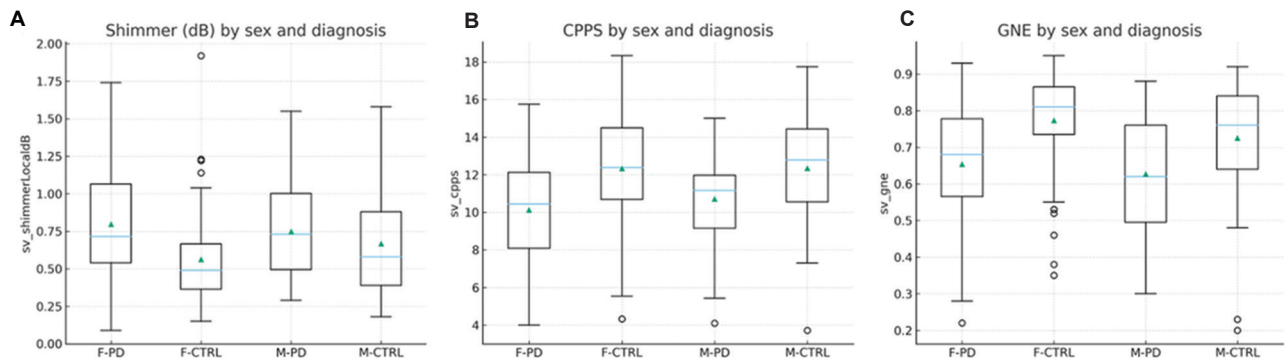


Figure 2. Sex-stratified boxplots by sex and diagnosis: (A) shimmer (dB), (B) CPPS, and (C) GNE. Abbreviations: CTRL: Healthy controls; CPPS: Cepstral peak prominence; F: Female; GNE: Glottal-to-noise excitation ratio; M: Male; PD: Parkinson's disease; sv: Sustained vowel.

vocal pitch but does not eliminate fundamental sex-based frequency differences.

The perturbation analyses revealed particularly striking results. PD-induced instability in vocal fold vibration was particularly evident in females. Jitter local increased by 42% in PD females (0.844 ± 0.61) compared with control females (0.59 ± 0.46 ; $p < 0.001$, $r = 0.36$, $q = 0.001$), whereas male differences were smaller and non-significant (PD: 0.74 ± 0.42 vs. controls: 0.75 ± 0.68 ; $p = 0.344$). Similarly, shimmer local was 39% higher in PD females (8.70 ± 3.72 vs. controls: 6.26 ± 3.66 ; $p < 0.001$, $r = 0.42$, $q = 0.003$) but showed only an 11% increase in males (PD: 8.18 ± 3.57 vs.

controls: 7.40 ± 4.03 ; $p = 0.240$). Shimmer (dB) confirmed the same pattern (females: +41%, $p < 0.001$; males: +12%, $p = 0.198$). These results suggest greater vulnerability of female voices to PD-related perturbation, with shimmer particularly sensitive.

The spectral analysis revealed complex patterns of sex and disease interaction. Sex differences in spectral quality were evident in controls but diminished in PD. Control females had superior HNR (dB) compared to males (females: 33.11 ± 5.75 vs. males: 27.21 ± 5.04 ; $p < 0.001$), yet this advantage disappeared in PD (29.56 ± 5.92 vs. 27.78 ± 4.14 ; $p = 0.072$). CPPS showed minimal

sex difference in controls (females: 12.34 ± 2.84 vs. males: 12.36 ± 2.94 ; $p=0.978$) but was significantly reduced in both PD groups (females: 10.14 ± 2.65 ; males: 10.72 ± 2.47 ; both $p<0.01$ vs. controls, $q < 0.03$). GNE also declined significantly in PD, with reductions of 16% in females (PD: 0.65 ± 0.17 vs. controls: 0.77 ± 0.13 ; $p<0.001$, $q < 0.005$) and 14% in males (PD: 0.63 ± 0.16 vs. controls: 0.73 ± 0.17 ; $p=0.005$, $q = 0.027$). These findings may indicate that while shimmer/jitter effects are female-specific, CPPS and GNE represent robust, sex-neutral biomarkers.

Analysis of pitch variability revealed interesting disease effects. Control females displayed greater pitch variability than males (females: 11.8 ± 18.7 Hz vs. males: 3.1 ± 4.9 Hz; $p=0.006$), but this difference was no longer significant in PD (females: 9.1 ± 15.7 Hz vs. males: 2 ± 11.7 Hz; $p=0.185$). Similarly, semitone range differences between sexes disappeared in PD. Voice break rates and spectral slope/tilt showed trends toward sex effects but did not reach significance after correction.

The spectral slope analysis revealed that control males had steeper slopes than females (-20.22 ± 3.09 vs. -17.53 ± 4.00 ; $p=0.001$), a pattern that persisted in PD patients (-20.35 ± 4.36 vs. -19.56 ± 3.75 ; $p=0.146$). Spectral tilt showed similar patterns, with control males having more negative values than females (-4.48 ± 2.75 vs. -3.27 ± 3.30 ; $p=0.068$), and PD patients showing intermediate values (-2.40 ± 3.12 vs. -3.21 ± 2.72 ; $p=0.406$). These findings suggest that spectral characteristics may be less affected by PD than time-based perturbation measures.

These analyses confirm that sexual dimorphism in voice is preserved but reshaped in PD. Female patients showed greater perturbation-related deterioration, particularly in shimmer measures, whereas cepstral and noise-based parameters (CPPS, GNE, and HNR) were consistently altered in both sexes. As clinical variables were comparable between males and females (Section 3.1.2), these differences likely reflect genuine sex-specific

vulnerability of the vocal apparatus to PD pathology rather than demographic or treatment confounds.

3.2.3. Linguistic differences

When stratified by language background, acoustic profiles of PD patients and controls were broadly similar across Russian-speaking and bilingual subgroups, with exploratory data available for the small Tatar-speaking group (Figure 3).

Mean pitch was highest in Tatar speakers (PD: 214.89 ± 15.07 Hz; controls: 176.26 ± 40.81 Hz), slightly lower in bilinguals (PD: 168.16 ± 44.29 Hz; controls: 179.62 ± 50.77 Hz), and lowest in Russian speakers (PD: 165.42 ± 48.22 Hz; controls: 174.80 ± 55.54 Hz). However, these differences were modest and non-significant after correction (Kruskal–Wallis $H = 2.14$, $p=0.34$, $q = 0.42$). Pitch variability followed a similar trend, with bilinguals and Tatars showing slightly larger ranges, but again without statistical significance.

Perturbation measures (jitter and shimmer) were consistently elevated in PD patients across all language groups. Russian-speaking PD patients demonstrated significantly higher jitter local (0.77 ± 0.40 vs. controls: 0.62 ± 0.56 ; $p=0.005$, $r = 0.30$, $q = 0.009$) and shimmer local (8.11 ± 3.68 vs. controls: 6.25 ± 3.25 ; $p=0.003$, $r = 0.54$, $q = 0.004$). Comparable effects were seen in bilinguals, with jitter (0.83 ± 0.65 vs. controls 0.68 ± 0.57 ; $p=0.083$, $q = 0.036$) and shimmer (8.85 ± 3.78 vs. controls 7.15 ± 4.57 ; $p=0.006$, $q = 0.004$) elevated in PD. The small Tatar subgroup showed the same direction of effect (e.g., shimmer local PD: 8.86 ± 2.38 vs. controls: 7.64 ± 3.24), but these differences did not reach significance, attributable to its limited sample size.

In contrast, CPPS was significantly lower in both Russian (10.63 ± 2.51 vs. controls: 12.41 ± 2.94 ; $p=0.0021$, $q = 0.006$) and bilingual PD patients (10.04 ± 2.75 vs. controls: 12.34 ± 2.91 ; $p<0.001$, $q = 0.001$), with large effect sizes ($r = -0.65$ to -0.81). GNE also showed robust

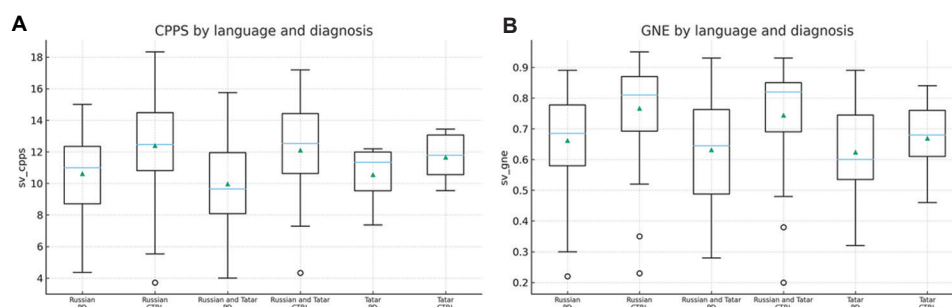


Figure 3. Language-stratified boxplots by language and diagnosis: (A) CPPS and (B) GNE

Abbreviations: CTRL: Healthy controls; CPPS: Cepstral peak prominence; GNE: Glottal-to-noise excitation ratio; PD: Parkinson's disease; sv: Sustained vowel.

reductions in Russian (0.66 ± 0.15 vs. controls: 0.77 ± 0.14 ; $p < 0.001$, $q < 0.001$) and bilingual PD patients (0.63 ± 0.18 vs. controls: 0.75 ± 0.16 ; $p = 0.001$, $q < 0.001$). Again, the Tatar subgroup showed numerically similar differences (e.g., CPPS: PD: 10.56 ± 1.71 vs. controls: 11.67 ± 1.47 ; GNE: PD: 0.62 ± 0.18 vs. controls: 0.67 ± 0.13), but these results should be considered preliminary due to their limited power.

Overall, the acoustic alterations associated with PD were consistent across Russian-speaking and bilingual participants, with jitter, shimmer, CPPS, and GNE emerging as reliable markers independent of language background. The exploratory data from the Tatar subgroup followed the same pattern but cannot be interpreted with confidence, given its limited sample size (7 PD and 5 controls). Furthermore, bilingual status was self-reported, without a formal assessment of proficiency, dominance, or frequency of use, which may have influenced outcomes. These limitations constrain the strength of conclusions regarding linguistic effects but support the cross-linguistic robustness of perturbation and cepstral/noise-based features as markers of PD-related dysphonia.

3.2.4. Covariate-adjusted analysis

Findings were robust in covariate-adjusted models. After controlling for age, sex, and language, PD remained associated with lower CPPS ($\beta = -1.93$, $p < 0.001$) and GNE ($\beta = -0.105$, $p < 0.001$), and higher shimmer ($\beta = +1.74$ local, $p = 0.0017$; $\beta = +0.169$ dB, $p = 0.0005$), while pitch measures remained non-significant. No diagnosis \times sex or diagnosis \times language interactions were detected after adjustment. In a conservative sensitivity model additionally including stage, duration, and medication covariates (all of which were not defined in controls), effect sizes were attenuated, but GNE remained significant ($\beta = -0.296$, $p = 0.014$). Full results are provided in Appendix Table A1.

These analyses demonstrate that the main acoustic markers of PD (jitter, shimmer, CPPS, and GNE) are robust to potential demographic and clinical confounds, strengthening their candidacy as reliable digital biomarkers of PD-related dysphonia.

3.2.5. Medication sensitivity analysis

To assess whether medication state could account for the observed acoustic differences, we performed two complementary sensitivity analyses: (i) we repeated PD versus control comparisons, restricting the PD group to participants assessed 90–180 min after their last levodopa dose; and (ii) We refitted covariate-adjusted models to include time since last dose and LEDD as covariates (in addition to age, sex, and language).

Within the defined window ($n = 14$ PD), all effects trended in the same direction as in the full sample but did not reach statistical significance, consistent with the small subsample. Pitch measures remained non-informative (mean pitch $p = 0.535$, pitch SD $p = 0.077$). These windowed results suggest that timing within a typical “on” interval does not abolish the PD–control differences; rather, power is insufficient in this small subset.

We next included LEDD and time since last dose alongside age, sex, and language in robust ordinary least squares models (HC3) to predict each acoustic outcome. The PD main effect remained significant for the core voice-quality markers (only significant results): CPPS: $\beta_{PD} = -2.61$ (95% CI = -3.81 – -1.42), $p < 0.001$; GNE: $\beta_{PD} = -0.153$ (95% CI = -0.232 – -0.075), $p = 0.0001$; shimmer (local): $\beta_{PD} = +2.20$ (95% CI = $+0.46$ – $+3.94$), $p = 0.013$; shimmer (dB): $\beta_{PD} = +0.216$ (95% CI = $+0.062$ – $+0.369$), $p = 0.006$; and HNR: $\beta_{PD} = -2.39$ (95% CI = -4.48 – -0.30), $p = 0.025$. Pitch measures again showed no significant adjusted PD effect (both $p > 0.50$; not shown).

Across both sensitivity approaches, the pattern and magnitude of PD-related changes were stable. Within the 90–180 min window, effects pointed the same way but were underpowered ($n = 14$). In covariate-adjusted models that explicitly accounted for time since last administration and LEDD, CPPS and GNE remained significantly reduced, and shimmer increased in PD, with HNR also significantly lower. These findings indicate that medication timing and overall dopaminergic load are not the primary drivers of the group differences observed in this cohort. Clinically, this is consistent with prior reports that PD dysphonia is less levodopa-responsive than limb motor symptoms, supporting the utility of voice measures as a complementary, relatively stable biomarker domain.^{28–30}

3.3. Correlations

Spearman's rank correlations were used to assess relationships between acoustic parameters and clinical measures (MDS-UPDRS Part III, LEDD, disease duration, and Hoehn and Yahr stage) in PD patients ($n = 109$). Correlations between acoustic and clinical measures were generally weak. The strongest (nominal) associations were (i) lower CPPS/HNR (dB) with greater motor severity when tremor was removed from MDS-UPDRS Part III (e.g., CPPS vs. MDS-UPDRS Part III without tremor: $\rho = -0.16$ to -0.19 ; HNR [dB] $\rho = -0.18$; $p = 0.07$ – 0.12), but none survived FDR; (ii) shimmer (local/dB) showed modest negative associations with disease duration and Hoehn and Yahr stage ($\rho = -0.21$; $p = 0.027$ – 0.031), again not significant after FDR; and (iii) medication variables (LEDD, last dose, and time since last dose) showed minimal relationships

with acoustic features ($|p|$ typically < 0.20 ; all $q > 0.55$). The correlation matrices are presented in Figure 4.

After rank-residualizing acoustic and clinical measures on age, sex, LEDD, and time since last dose, CPPS retained a small negative association with MDS-UPDRS Part III (without tremor) (partial $\rho = -0.185$; $p=0.092$), while GNE and shimmer associations remained weak and non-significant. These patterns support the view that voice quality degradation relates weakly to axial/bradykinetic severity, but not to tremor or medication timing *per se*.

In the real world, advanced-PD, outpatient cohort, acoustic markers (especially CPPS, GNE, and shimmer) distinguish PD from controls, but they correlate only weakly with clinical scales and medication exposure. Voice, therefore, appears to index a partly independent symptom domain, consistent with prior reports that PD dysphonia is not strongly levodopa-responsive. These results motivate longitudinal tracking (progression sensitivity) and multidomain modeling rather than relying on cross-sectional correlations with MDS-UPDRS Part III.

4. Discussion

In our real-world, advanced-PD, outpatient cohort, sustained-vowel analysis revealed a coherent and biologically plausible dysphonia profile: increased perturbation (jitter and shimmer), reduced CPPS, and deterioration of noise-related measures (GNE and HNR), whereas fundamental frequency metrics were comparatively preserved. These

differences remained after FDR control and persisted in covariate-adjusted models (age, sex, and language), were robust to sensitivity analyses for medication timing and dopaminergic load, and correlated only weakly with concurrent motor-severity indices. Collectively, the results suggest that CPPS, GNE, and shimmer are reliable and scalable markers of PD-related phonatory impairment that are not reducible to demographics, language background, or immediate levodopa state, in line with reports across other languages and cohorts.^{12,13,16,26}

The clustering of elevated jitter/shimmer with lower CPPS/HNR/GNE recapitulates prior clinical-acoustic observations in PD and matches the physiology of hypokinetic dysarthria: increased laryngeal rigidity, reduced subglottal pressure, and incomplete glottal closure, yielding unstable vocal-fold vibration and increased aperiodic energy.^{16,30,31} CPPS and GNE summarize harmonic organization and glottal noise excitation, respectively, and thus behave as sensitive indices of impaired phonation stability. By contrast, the mean fundamental frequency is dominated by anatomic and communicative determinants (laryngeal size, sex, and task intent), explaining its limited discriminative value here and in several prior reports.³¹ Spectral trends (steeper spectral slope and H1–H2 tendencies) were weaker, unsurprising given the additional variability from vocal-tract filtering and task execution; resolving these effects requires larger samples and connected-speech tasks.³⁰

Sex-stratified analyses clarify two points. First, sexual dimorphism in fundamental frequency measures is preserved in PD, although attenuated, consistent with persistent anatomical differences in the larynx.³² Second, perturbation-heavy markers, especially shimmer, were more abnormal in PD females than in PD males, whereas CPPS and GNE were comparably abnormal in both sexes. As male and female patients were well-matched on age, disease duration, MDS-UPDRS Part III, and dopaminergic treatment, these contrasts are unlikely to reflect clinical confounding. Mechanistically, sex-linked differences in vocal-fold tissue (mucosal thickness and viscoelastic properties) and hormonal milieu provide plausible explanations for greater shimmer vulnerability in females.^{32,33} Our results, consistent with prior studies, suggest sex-sensitive thresholds for perturbation indices, anchored by sex-neutral cepstral/noise markers (CPPS and GNE).^{11,13} Notably, formal diagnosis \times sex interactions did not reach significance after covariate control, implying that sex differences are feature-specific rather than global. Larger sample sizes powered for interactions and direct laryngeal assessments (e.g., videostroboscopy) could clarify the mechanisms and effect sizes.³⁴

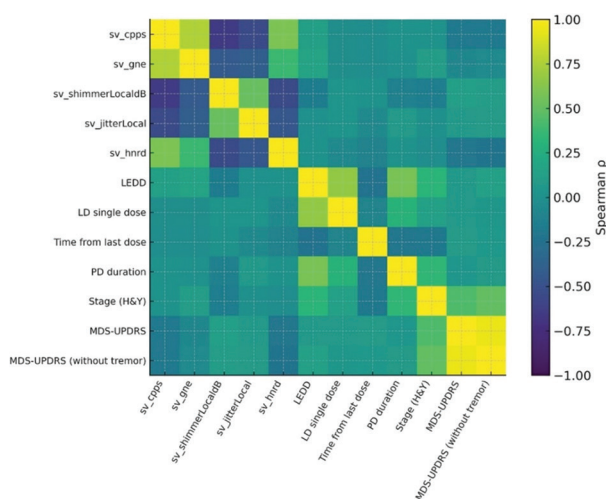


Figure 4. Correlation matrix for acoustic and clinical parameters. Abbreviations: cpps: Cepstral peak prominence; gne: Glottal-to-noise excitation ratio; H&Y: Hoehn and Yahr; hnrd: Harmonic-to-noise ratio in dB; LD single dose: Dose of levodopa intake before assessment; LEDD: Levodopa equivalent daily dose; MDS-UPDRS: Unified Parkinson's Disease Rating Scale Part III (motor examination); p: Two-sided unadjusted p -value; PD: Parkinson's disease; sv: Sustained vowel.

Using a task that minimizes articulatory and prosodic confounds (sustained/a:/), we observed near-identical PD signatures in Russian monolinguals and Russian-Tatar bilinguals: jitter/shimmer increases and CPPS/GNE reductions with comparable effect sizes, consistent with language-robust phonatory markers reported in other languages.^{18,27,34} At the same time, we refrain from claiming full invariance: the Tatar monolingual subgroup was limited (7 PD and 5 controls), rendering its findings exploratory, and bilingual status was based on self-report without standardized proficiency/dominance/usage metrics. Both factors can modulate speech motor control and may obscure subtle language-linked effects.³⁴

Methodologically, the sustained vowel was intentional; it isolates phonation to test cross-language generalizability. The present results support phonatory robustness (CPPS, GNE, and perturbation). They do not preclude language-linked differences in prosody or articulation, which require connected-speech protocols. Future work should therefore combine sustained vowels with standardized reading and spontaneous speech and incorporate formal bilingual profiling.

Acoustic features showed only weak associations with MDS-UPDRS Part III (including tremor-subtracted scores), Hoehn and Yahr stage, disease duration, and medication exposure; none survived FDR. Covariate-adjusted partial correlations retained at most a modest link between lower CPPS and greater axial/bradykinetic burden, implying specific but limited coupling rather than redundancy with limb motor ratings, echoing prior work.^{28,35} Medication sensitivity analyses further support relative levodopa insensitivity: limiting analyses to a typical “ON” window (90–180 min post-dose) preserved effect directionality (underpowered subset), and models explicitly adjusting for LEDD and time since last dose retained significant PD effects for CPPS, GNE, and shimmer (and HNR).^{28,35,36} Clinically, these findings align with evidence that speech/voice is less consistently levodopa-responsive than limb motor signs, reinforcing voice as a complementary, relatively stable biomarker domain suitable for remote monitoring.²⁸

These findings have important implications for understanding PD-related voice changes. The dissociation between vocal parameters and medication status suggests that voice symptoms may be less responsive to dopaminergic therapy than other motor symptoms, potentially reflecting involvement of non-dopaminergic pathways in PD-related dysphonia. The results support using multiparameter voice assessment in PD, as different acoustic measures appear to capture distinct yet interrelated aspects of vocal impairment. The limited correlation with

motor scores further reinforces the value of voice analysis as an independent biomarker domain in PD assessment.

These findings underline the robustness and sensitivity of acoustic parameters as digital biomarkers, not only distinguishing PD from healthy states but also reflecting subtle differences linked to sex and native language. This points toward potential personalized voice-based monitoring strategies for PD progression and intervention effectiveness, notably through remote, accessible digital health platforms. Future research should use longitudinal designs to quantify progression sensitivity, larger minority-language cohorts with formal bilingual profiling, and task diversification (sustained vowels + connected speech).

Several limitations should be considered when interpreting these findings, particularly with respect to language-specific vocal characteristics in PD:

- (i) The sample size of the Tatar-speaking subgroup was limited (7 PD and 5 controls). This severely limits statistical power, widens CIs, and increases the risk of both false negatives (failure to detect true effects) and unstable effect-size estimates. All Tatar-specific results should therefore be regarded as exploratory. Similarly, sample-size imbalance across language categories (Russian > bilingual > Tatar) may have reduced the sensitivity to modest language effects and complicated interaction testing (diagnosis × language), even after we applied non-parametric methods and FDR control
- (ii) The study was cross-sectional. We cannot infer trajectories of voice change or determine whether language background, sex, or clinical variables (e.g., axial symptoms) differentially influence rates of deterioration. Longitudinal data are required to assess progression, medication-state dynamics within individuals, and temporal coupling between acoustic measures and clinical outcomes
- (iii) Our bilingual characterization relied on self-report. We did not implement standardized assessments of language proficiency, dominance, age of acquisition, or frequency of use. These factors can modulate speech motor control and phonation stability, thereby attenuating or obscuring language-linked differences
- (iv) Recordings were obtained in routine clinical settings, not in soundproof environments. Although we enforced minimum sampling/bit depth and used features known to be relatively robust, ambient noise may still affect certain parameters, especially noise-sensitive metrics (HNR and GNE). This choice improves ecological validity but may introduce measurement noise that biases estimates
- (v) We focused on a single phonation task (sustained/a:/). This task intentionally isolates laryngeal function and

minimizes articulatory/prosodic confounds; hence, it is advantageous for cross-linguistic comparisons. However, it cannot capture language-dependent prosody, coarticulation, or segmental features that may emerge more clearly in connected speech. As such, our findings speak primarily to phonatory biomarkers rather than the full spectrum of speech impairment in PD

- (vi) The medication state was heterogeneous. Although we recorded the last levodopa dose and time since intake and performed medication-sensitivity analyses, we did not experimentally standardize states. Subtle pharmacodynamic effects or delayed responses could therefore elude detection in a cross-sectional snapshot
- (vii) While we controlled the FDR within feature families, multiple testing across numerous acoustic variables still raises the possibility of type I error. Conversely, given subgroup sizes (notably the Tatar group), type II error (false negatives) is also a concern, particularly for interaction tests (diagnosis \times sex, diagnosis \times language)
- (viii) Although we used robust estimators and sensitivity models, our analytical choices (e.g., feature extraction settings and reliance on a single sustained vowel) may influence absolute values of certain features (pitch trackers and perturbation algorithms are known to vary by implementation). Harmonization across software versions and test–retest reliability assessments were beyond the scope of this study and should be addressed in future work.

In summary, the small Tatar subgroup, self-reported bilingual status without proficiency/dominance testing, cross-sectional design, clinic-based recording conditions, and single-task protocol are the principal constraints. These limitations reinforce our cautious interpretation of language effects as exploratory and motivate longitudinal, multilingual studies with standardized bilingual profiling, connected-speech tasks, and environment/device harmonization to validate and extend the present findings.

5. Conclusion

In summary, sustained-vowel acoustics captured robust, medication-insensitive differences between PD and controls in our multilingual cohort, with CPPS, GNE, and shimmer emerging as pragmatic markers that generalize across sex and (within the limits of our sample) language background. Their weak ties to limb motor ratings emphasize that voice reflects a partly independent dimension of PD. The physiological plausibility, statistical robustness, and ecological feasibility support their integration into inclusive, language-aware clinical workflows and remote monitoring pipelines.

This study aimed to bridge critical gaps in PD research by validating sustained vowel acoustics as digital biomarkers in the Russian population that is underrepresented in scientific research. By combining rigorous acoustic analysis with clinical correlations, we strive to advance equitable diagnostic tools that align with linguistic and healthcare realities. Our findings may not only enhance local PD management but also inform global efforts to harness voice as a biomarker in diverse linguistic and cultural settings. This study underscores the potential of voice analysis as a scalable PD biomarker in multilingual populations, with implications for inclusive, language-sensitive diagnostic tools.

Acknowledgments

None.

Funding

The Foundation for Assistance to Small Innovative Enterprises funded initial research on 03.07.2023 (grant number: 4986FC1/85525).

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: All authors

Formal analysis: All authors

Investigation: All authors

Methodology: All authors

Writing—original draft: All authors

Writing—review & editing: All authors

Ethics approval and consent to participate

The study protocol was reviewed and approved by the ethics committee of Kazan State Medical University (approval no.: # 6 [18.06.2024]). Written form of consent was obtained from each of the subjects to participate in the study.

Consent for publication

All participants gave written informed consent for participation in the study and for the use of anonymized voice sample data for publication.

Availability of data

The acoustic feature dataset that supports the findings of this study is available from the corresponding author upon reasonable request. However, the original audio recordings cannot be shared due to participant privacy and consent constraints.

Further disclosure

Parts of the findings have been published as abstracts at the Movement Disorders Congress 2024 on September 27 - October 1, 2024 at Philadelphia, PA, USA under the titles "Efficiency of AI-Based Service for Screening Diagnostics of Parkinson's Disease (Brainphone Project)" (<https://www.mdsabstracts.org/abstract/efficiency-of-ai-based-service-for-screening-diagnostics-of-parkinsons-disease-brainphone-project/>) and "Gender specificity in AI-Based Screening Diagnostics of Parkinson's Disease (Brainphone Project)" (<https://www.mdsabstracts.org/abstract/gender-specificity-in-ai-based-screening-diagnostics-of-parkinsons-disease-brainphone-project/>); and at AD/PD™ 2024 International Conference on Alzheimer's and Parkinson's Diseases and related neurological disorders on March 5–9, 2024 at Lisbon, Portugal, under the titles "VHI-10 And Cognitive Dysfunction-Relationship And Appropriateness Of Use In Patients With Parkinson's Disease" (DOI:10.13140/RG.2.2.35570.52162) and "Artificial Intelligence Opportunities For Voice Diagnostics Of Parkinson's Disease (Brainphone Project)" (DOI:10.13140/RG.2.2.11243.55843).

References

- World Health Organization. *Parkinson Disease Fact Sheet*. Available from: <https://www.who.int/news-room/fact-sheets/detail/parkinson-disease> [Last accessed on 2025 Nov 10].
- Poewe W, Seppi K, Tanner CM, *et al*. Parkinson disease. *Nat Rev Dis Primers*. 2017;3:17013.
doi: 10.1038/nrdp.2017.13
- Kalashnikova OS, Munasipova SE, Khasanova DM, Latypova GR, Zalyalova ZA. Epidemiology of Parkinson's disease in the Republic of Tatarstan. *Pract Med*. 2011;7(55):210-211.
- Rozhdestvensky AS, Delov RA, Marks EA, Gaponenko IA, Khanokh EV. Clinical and epidemiological aspects of Parkinson's disease in the South of Western Siberia. *Front Neurol*. 2020;11:538782.
doi: 10.3389/fneur.2020.538782
- Razdorskaya VV, Voskresenskaya ON, Yudina GK. Parkinson's disease in Russia: Prevalence and incidence. *Saratov J Med Sci Res*. 2016;12(3):379-384.
- Zalyalova ZA, Kalashnikov OS, Abduglimova DM. Parkinsonism in the Republic of Tatarstan according to the data of the Republican clinic-diagnostic center extrapyramidal pathology and Botulinotherapy. *Neurol Bull*. 2011;XLIII(2):92-96.
doi: 10.17816/nb13706
- Postuma RB, Berg D, Stern M, *et al*. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord*. 2015;30(12):1591-1601.
doi: 10.1002/mds.26424
- Ministry of Health of the Russian Federation. *Clinical Guidelines Parkinson's Disease, Secondary Parkinsonism and Other Diseases Manifested By Parkinsonism Syndrome*; 2021.
- FDA-NIH (Food and Drug Administration-National Institutes of Health). *Biomarker Working Group. BEST (Biomarkers, Endpoints, and other Tools) Resource*. Silver Spring, MD: Food and Drug Administration; 2016.
- Vasudevan S, Saha A, Tarver ME, Patel B. Digital biomarkers: Convergence of digital health technologies and biomarkers. *NPJ Digit Med*. 2022;5:36.
doi: 10.1038/s41746-022-00583-z
- Cao F, Vogel AP, Gharahkhani P, Renteria ME. Speech and language biomarkers for Parkinson's disease prediction, early diagnosis and progression. *NPJ Parkinsons Dis*. 2025;11:57.
doi: 10.1038/s41531-025-00913-4
- Naeem I, Ditta A, Mazhar T, Anwar M, Saeed MM, Hamamet H. Voice biomarkers as prognostic indicators for Parkinson's disease using machine learning techniques. *Sci Rep*. 2025;15:12129.
doi: 10.1038/s41598-025-96950-3
- Atalar MS, Oguz O, Genc G. Hypokinetic dysarthria in Parkinson's disease: A narrative review. *Sisli Etfal Hastan Tip Bul*. 2023;57(2):163-170.
doi: 10.14744/SEMB.2023.29560
- Xavier D, Felizardo V, Ferreira B, *et al*. Voice analysis in Parkinson's disease - a systematic literature review. *Artif Intell Med*. 2025;163:103109.
doi: 10.1016/j.artmed.2025.103109
- Suppa A, Costantini G, Asci F, *et al*. Voice in Parkinson's disease: A machine learning study. *Front Neurol*. 2022;13:831428.
doi: 10.3389/fneur.2022.831428
- 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*. 2011;129:350-367.
doi: 10.1121/1.3514381
- Dudek M, Hemmerling D, Kaczmarek M, *et al*. Analysis of voice, speech, and language biomarkers of Parkinson's disease collected in a mixed reality setting. *Sensors (Basel)*. 2025;25(8):2405.
doi: 10.3390/s25082405
- Sousa M, Krýže P, Amstutz D, *et al*. Digital speech biomarkers can measure acute effects of levodopa in Parkinson's disease. *NPJ Parkinsons Dis*. 2025;11:184.

- doi: 10.1038/s41531-025-01045-5
19. Siniukov M, Xing E, Attaripour Isfahani S, Soleymani M. Towards a Generalizable Speech Marker for Parkinson's Disease. *ArXiv:2501.03581v2* [eess.AS].
doi: 10.48550/arXiv.2501.03581
 20. Conklin J, Dmitrieva O. Vowel acoustics of Volga Tatar. In: *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019)*; 2019. p. 1604-1608.
 21. Khasanova D, Khasanov I, Ilina G, Zalyalova Z. Gender specificity in AI-Based screening diagnostics of Parkinson's disease (brainphone project). *Mov Disord*. 2024;39:S574-S575.
 22. Zalyalova Z, Khasanova D, Khasanov I, Ilina G. Artificial intelligence opportunities for voice diagnostics of Parkinson's disease (BRAINPHONE project). AD/PD 2024. Advances in Science and Therapy. In: *Conference: International Conference on Alzheimer's and Parkinson's Diseases and Related Neurological Disorders (ADPD 2024)*. Lisbon; 2024. p. 1147.
doi: 10.32000/2072-1757-2024-1-70-76
 23. Khasanova D, Khasanov I, Zalyalova Z. VHI-10 and cognitive dysfunction - relationship and appropriateness of use in patients with Parkinson's disease. In: *Conference: International Conference on Alzheimer's and Parkinson's Diseases and Related Neurological Disorders (ADPD 2024)*, Lisbon, 2024. p. 1412.
 24. World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191-2194.
doi: 10.1001/jama.2013.281053
 25. Arora S, Visanji NP, Mestre TA, *et al*. Investigating voice as a biomarker for leucine-rich repeat kinase 2-associated Parkinson's disease. *J Parkinsons Dis*. 2018;8(4):503-510.
doi: 10.3233/JPD-181389
 26. Wang M, Zhao X, Li F, *et al*. Using sustained vowels to identify patients with mild Parkinson's disease in a Chinese dataset. *Front Aging Neurosci*. 2024;16:1377442.
doi: 10.3389/fnagi.2024.1377442
 27. Tsanas T, Arora S. Data driven subtyping of parkinson's using acoustic analysis of sustained vowels and cluster analysis: Findings in the Parkinson's voice initiative study. *SN Comput Sci*. 2022;3:232.
doi: 10.1007/s42979-022-01123-y
 28. Tykalova T, Novotny M, Ruzicka E, Dusek P, Rusz J. Short-term effect of dopaminergic medication on speech in early-stage Parkinson's disease. *npj Parkinsons Dis*. 2022;8(1):22.
doi: 10.1038/s41531-022-00286-y
 29. Rusz J, Tykalova T, Novotny M, Zogala D, Ruzicka E, Dusek P. Automated speech analysis in early untreated Parkinson's disease: Relation to gender and dopaminergic transporter imaging. *Eur J Neurol*. 2021;29:81-90.
doi: 10.1111/ene.15099
 30. Asci F, Costantini G, Saggio G, Suppa A. Fostering voice objective analysis in patients with movement disorders. *Mov Disord*. 2021;36:1041.
doi: 10.1002/mds.28537
 31. Cernak M, Orozco-Arroyave JR, Rudzicz F, Christensend H, Vasquezb JK, Noth E. Characterisation of voice quality of Parkinson's disease using differential phonological posterior features. *Comput Speech Lang*. 2017;46:196-208.
doi: 10.1016/j.csl.2017.06.004
 32. Hertrich I, Ackermann H. Gender-specific vocal dysfunctions in Parkinson's disease: Electroglottographic and acoustic analyses. *Ann Otol Rhinol Laryngol*. 1995;104(3):197-202.
doi: 10.1177/000348949510400304
 33. Wu KH, Tobias ML, Kelley DB. Estrogen receptor expression in laryngeal muscle in relation to estrogen-dependent increases in synapse strength. *Neuroendocrinology*. 2003;78(2):72-80.
doi: 10.1159/000071962
 34. Rusz J, Krack K, Tripoliti E. From prodromal stages to clinical trials: The promise of digital speech biomarkers in Parkinson's disease. *Neurosci Biobehav Rev*. 2024(167):105922.
doi: 10.1016/j.neubiorev.2024.105922
 35. Silva JMSD, Gomes AOC, Coriolano MDGWS, *et al*. Oropharyngeal geometry and acoustic parameters of voice in healthy and Parkinson's disease subjects. *Codas*. 2023;35(2):e20210304.
doi: 10.1590/2317-1782/20232021304pt
 36. Ho AK, Bradshaw JL, Iansek R. For better or worse: The effect of levodopa on speech in Parkinson's disease. *Mov Disord*. 2008;23(4):574-580.
doi: 10.1002/mds.21899

Appendix

Table A1. Covariate-adjusted models

Feature	Beta_DX	Confidential interference		<i>p</i>
		Low	High	
sv_cppps	−1.9271	−2.7121	−1.1422	0.0
sv_gne	−0.1054	−0.15	−0.0608	0.0
sv_shimmerLocaldB	0.1694	0.0735	0.2652	0.0005
sv_shimmerLocal	1.7378	0.6538	2.8217	0.0017
sv_hnr	−1.9332	−3.3023	−0.5641	0.0056
sv_hnrd	−2.0402	−3.6383	−0.442	0.0123