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.

Associations of Spontaneous Speech prosody with Cognitive and Motor function in Parkinson's disease

Gianna Kuhles

g.kuhles@fz-juelich.de

Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf

Tabea Thies

Faculty of Arts and Humanities, IfL – Phonetics, University of Cologne

Michael T. Barbe

Faculty of Medicine and University Hospital Cologne, Department of Neurology, University of Cologne

Simon B. Eickhoff

Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf

Julia A. Camilleri

Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf

Susanne Weis

Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf

Article

Keywords:

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11757771/v1

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

Additional Declarations: Competing interest reported. None of the authors have conflicts of interest to disclose in relationship with the current work. However, T.T. is an employee of ki:elements GmbH.

Associations of Spontaneous Speech prosody with Cognitive and Motor function in Parkinson's disease

Gianna Kuhles[1,2,*], Tabea Thies[3,4], Michael T. Barbe[4], Simon B. Eickhoff[1,2], Julia A. Camilleri[1,2], & Susanne Weis[1,2]

* Corresponding author g.kuhles@fz-juelich.de

1 Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

2 Institute of Neuroscience and Medicine (INM-7: Brain and Behaviour), Research Centre Jülich, Jülich, Germany

3 Faculty of Arts and Humanities, IfL – Phonetics, University of Cologne, Köln, Germany

4 Faculty of Medicine and University Hospital Cologne, Department of Neurology, University of Cologne, Köln, Germany

ABSTRACT

Cognitive changes are common in Parkinson's disease (PD), yet efficient digital assessment markers remain limited. Speech is easily recorded and automatically analysed, but whether it provides practical markers for cognitive function requires further investigation. This study examined associations of spontaneous speech prosody with cognitive performance (MoCA, PANDA) and motor symptom severity (UPDRS-III) in 176 individuals with PD. Prosodic components improved estimation of all three measures beyond sex, age, and depressive symptoms, most strongly for MoCA. Correlation analyses showed that better cognitive performance was associated with louder, steadier, more pitch-varied speech with shorter pauses, and with more variable formant bandwidths, whereas associations with motor severity were weak and did not survive adjustment. In cross-validation on unseen individuals, prosody improved prediction beyond the covariates for MoCA and PANDA, whereas UPDRS-III could not be predicted. These findings suggest that prosody is associated with cognitive performance, highlighting its potential as a functional digital marker for low-burden monitoring.

INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterised by motor and non-motor symptoms. Traditionally, the clinical assessment of the condition has focused heavily on motor aspects. However, cognitive changes also form part of the clinical spectrum and may already be relevant in the early stages of the disease (Aarsland et al., 2021; Hely et al., 2008). Frequent monitoring of cognitive performance over time typically requires repeated

assessments by specialised personnel, which can be challenging to implement in routine clinical practice. This highlights the need for additional, objective, and sensitive markers that can be obtained quickly, repeatedly and with little burden, ideally in real-life settings, especially since changes in cognitive function can occur as early as the prodromal phase of PD and may even predate motor symptoms (Cao et al., 2025; Fengler et al., 2017; Sugimura et al., 2026). Among the potential candidates for such markers, speech has emerged as a particularly promising area of investigation.

Speech is frequently affected in PD (Frota et al., 2021) with abnormalities often emerging early and reflecting multiple aspects of the disease (Rusz et al., 2013). Unlike neuroimaging or biofluid markers, linguistic features can be readily extracted from speech recordings without specialised clinical infrastructure, making their assessment highly practical, low-effort, and scalable. This accessibility enhances their potential for the repeated monitoring of cognitive performance, including in everyday settings. In particular, spontaneous speech offers a window into natural but cognitively demanding speech production whilst encompassing a broad range of linguistic and paralinguistic information. Entirely free speech, however, allows speakers to choose topics of differing cognitive difficulty. Picture description represents a compromise as it elicits connected, self-generated speech while constraining its content, thereby standardising the cognitive demand across speakers, and it has proven informative for cognitive impairment in PD (Andrade et al., 2023; Smith et al., 2025). Automatic analysis of such speech samples could thus provide objective quantitative markers that are both efficient and scalable for clinical use (Cao et al., 2025; Pizzimenti et al., 2026). However, the literature on speech biomarkers is heterogeneous, which limits the generalisability of its findings. Studies vary in their datasets, recording protocols and analysis methods (Ngo et al., 2022). Additionally, speech features can be influenced by factors such as the recording device or the speaker's state at the time of recording (Cummins et al., 2025), and the value of extracted features can depend on the extraction method (Dineley et al., 2024). Many commonly used features also show low repeatability across recordings (Stegmann et al., 2020), and many constructs lack standardised operationalisation (Berisha & Liss, 2024). In addition, reported accuracies tend to decrease as sample size increases, suggesting overfitting in small samples (Berisha et al., 2022). Consequently, comparisons across studies remain challenging, and the identification of speech characteristics robustly and specifically associated with PD and its clinical features remains elusive, hindering clinical translation. Addressing these challenges requires standardised speech elicitation, established and automatically extracted feature sets, sufficiently large samples relative to the number of features, and validation in individuals not used for model fitting.

The prosodic aspect of speech appears to be of particular relevance in PD and refers to suprasegmental properties of speech that shape how utterances are organised temporally and acoustically, including features such as fundamental frequency (F0), intensity, speech rate, and pauses. Prosodic encoding operates at both cognitive and motor levels, requiring speakers to cognitively select which elements to highlight through prosodic modulation while coordinating the speech motor system to execute these changes. Consequently, prosody contributes to expression, structuring, and intelligibility, and requires the coordinated organisation of speech. Studies show that prosody is frequently altered in PD, however, across studies, only F0 variability and pause duration consistently differed in PD, whereas evidence for changes in other prosodic measures was limited or absent (Fumel et al., 2024).

Prosodic changes in PD have mostly been interpreted as manifestations of impaired motor control (Martens et al., 2015; Skodda et al., 2011b). However, the temporal organisation of speech also requires planning across utterances (Ferreira, 1991), and features such as pausing reflect both motor and cognitive influences (Smith & Caplan, 2018). Prosodic features may therefore reflect the combined effects of both motor execution and cognitive control, suggesting that they could relate to cognitive performance as well as motor symptom severity. Since speech engages multiple cognitive processes, it is plausible that linguistic features may also contain information about an individual's cognitive performance. In PD, changes in various cognitive-linguistic domains have been described (Cao et al., 2025; Fagherazzi et al., 2021; Fumel et al., 2024; Rektorova et al., 2016), with particular emphasis on the link between executive function impairments and speech production (Herbig et al., 2025; Thies et al., 2020; Wei et al., 2025). Accordingly, speech characteristics, including prosodic features, are increasingly being investigated as potential markers of neurodegenerative disease (García et al., 2023; Hecker et al., 2022; Hernandez et al., 2026; König et al., 2026; Kovac et al., 2025; Pérez-Toro et al., 2025; Stegmann et al., 2022). In PD, features such as speaking rate and the timing of pauses between utterances during a picture description task have been associated with cognitive performance and mild cognitive impairment, but not with motor symptom severity (Andrade et al., 2023). Furthermore, in early-stage PD, speech difficulties have been linked to more rapid cognitive decline, but not to more severe motor progression (Polychronis et al., 2019).

When investigating these associations, it is important to consider factors that may influence both prosodic characteristics and clinical outcomes, as these can create spurious associations. Age and sex are known to affect both speech characteristics and cognitive performance (Harada et al., 2013; Skodda et al., 2011a; Torre & Barlow, 2009), while depressive symptoms can affect both cognitive functioning and speech production (Pérez-Toro et al., 2022). This is particularly relevant in PD, where depression is among the most

frequent non-motor symptoms, with clinically relevant depressive symptoms reported in approximately one-third of individuals (Reijnders et al., 2008). Beyond its prevalence, depression has been linked to changes in speech (Pérez-Toro et al., 2022), and it can lower performance on cognitive tests. Therefore, for prosody markers to be of practical clinical value, they need to provide information beyond available demographic and clinical variables. At the same time, speech characteristics related to disease duration and motor severity may themselves carry clinically relevant information about disease progression and symptom burden, rather than representing confounding influences that should be accounted for.

The present study addressed the methodological requirements outlined above in 176 individuals with PD, using a standardised picture description task, an established set of automatically extracted prosodic features, and assessment of speech and motor severity off medication. Specifically, we investigated 1) whether prosody explains variance in cognitive performance (Montreal Cognitive Assessment, MoCA and Parkinson Neuropsychometric Dementia Assessment, PANDA) and motor symptom severity (Unified Parkinson's Disease Rating Scale, UPDRS-III) beyond sex, age and depressive symptoms, 2) which prosodic characteristics are most strongly associated, using component-level correlations, and 3) whether this gain generalises to held-out individuals, assessed through cross-validated prediction. Together, these analyses test whether prosody provides information about cognitive performance in PD that could support low-burden, repeatable monitoring.

RESULTS

Sample characteristics and clinical heterogeneity in PD

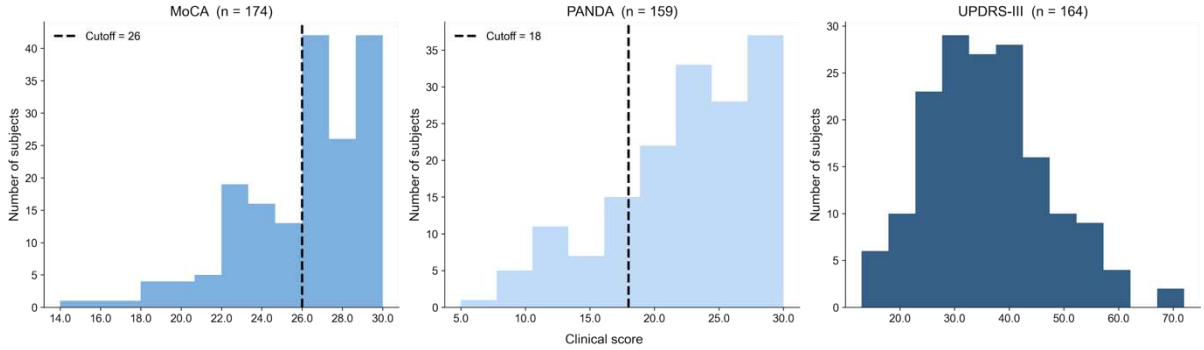
Speech recordings and clinical assessments were obtained from 176 individuals with PD. All individuals with PD provided complete prosodic feature data. The three analysis samples differ in size depending on the availability of clinical measures: The analyses included 174 individuals in the MoCA models (1.1% of the sample excluded), 159 in the PANDA models (9.7% excluded), and 164 in the UPDRS-III models (6.8% excluded). Retained and excluded individuals did not differ on any clinical variable in any of the three comparisons (Appendix A, Table A1).

Individuals' cognitive performance averaged 26.00 on the MoCA (SD=3.25, min=14, max=30) and 22.38 on the PANDA (SD=5.78, min=5, max=30). Sixty-four individuals (36.8%) scored below the MoCA cutoff of 26 points, and 27 (17.0%) scored below the PANDA cutoff of 18 points, suggesting possible clinically relevant cognitive impairment. Motor performance ranged from mild to severe impairment, with UPDRS-III scores averaging 35.99 (SD=11.19, min=13, max=72). The distributions of all three clinical scores are shown in Figure 1. In

addition, depressive symptoms assessed with the Beck Depression Inventory–II (BDI-II, Beck et al., 1996) averaged 9.36 (SD=6.78, min=0, max=34). Although the mean score fell within the minimal range (Beck et al., 1996), 37 individuals scored 14 or higher, indicating at least mild depressive symptoms (mild: n = 20, moderate: n = 15, severe: n = 2). The two cognitive measures were moderately correlated (r = 0.48), and both correlated weakly with motor symptom severity (both r = −0.30). BDI-II scores were weakly associated with MoCA (r = −0.18) and PANDA (r = −0.21) but not with UPDRS-III scores (Appendix A, Table A5).

Figure 1

Histograms showing clinical heterogeneity in the PD cohort across MoCA, PANDA, and UPDRS-III



Prosodic profiles in PD

Prosodic speech patterns, extracted from a 90-second picture description (see Methods) were characterised by the 88 features of the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS, Eyben et al., 2016) grouped into frequency, energy/amplitude, spectral and temporal domains (see Methods). Descriptive values for all 88 prosodic features within the PD group are shown in Appendix A, Figure A1, and listed in Appendix A, Table A2.

To characterise how the covariates relate to prosody itself, each of the 88 features was examined for associations with sex, age and BDI-II score (Appendix B, Table B6). Of the 88 features 46 showed significant sex differences (Appendix B, Table B6), mainly in the frequency and spectral domains, with higher fundamental and formant frequencies in women, whereas no feature was associated with age or, after correction, with BDI-II score. In addition to the within-group patterns described above, the PD group was compared with an age- and sex-matched reference group (Appendix A, Table A3). As openSMILE does not provide normative

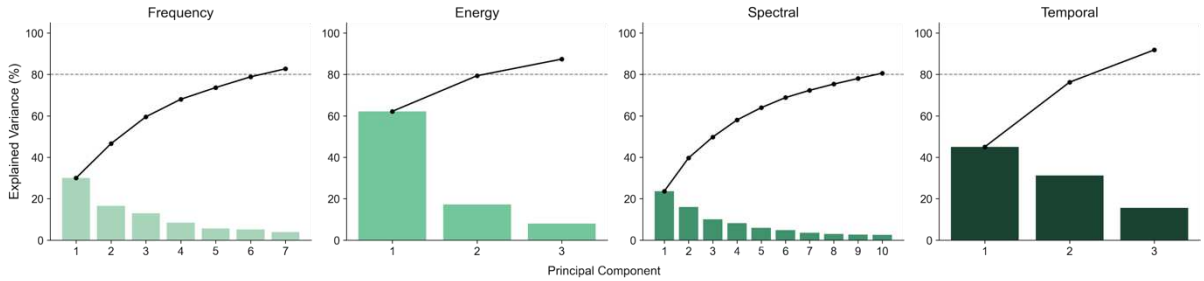
reference values for the eGeMAPS features (Eyben et al., 2016), this comparison offers a study-specific benchmark for contextualising the direction and magnitude of the observed prosodic patterns. After domain-specific false discovery rate (FDR) correction, 64 of the 88 prosodic features differed significantly between the groups (Appendix A, Table A3). As recording conditions were not fully standardised across groups, the comparison should be interpreted with caution, as they may, at least partly, reflect site-related effects.

Domain-specific principal component analysis

The prosodic features were moderately intercorrelated within domains (mean absolute correlation 0.21 to 0.50, Appendix A, Table A4). To reduce their number relative to the sample size, principal component analyses (PCA) were performed separately within each prosodic domain, retaining components until 80% of the variance was explained (see Methods). PCA reduced the 88 prosodic features to 23 principal components (PC) while retaining a minimum of 80% of the variance within each domain: seven frequency, three energy/amplitude, ten spectral, and three temporal components displayed in Figure 2. The three highest-loading features for each component are reported in Appendix A, Table A6.

Figure 2

PCA scree plots per prosodic domain



The distribution of explained variance differed across domains. The first energy/amplitude component accounted for 62.2% of the variance, compared with 45.0% in the temporal, 30.0% in the frequency, and 23.7% in the spectral domain. The spectral domain showed the least concentrated variance structure, requiring ten components to reach the 80% threshold, with the tenth component accounting for 2.6%. Thus, spectral variation was distributed across several distinct dimensions rather than dominated by a small number of components. The Kaiser–Meyer–Olkin (KMO) values for the energy/amplitude, frequency, and spectral

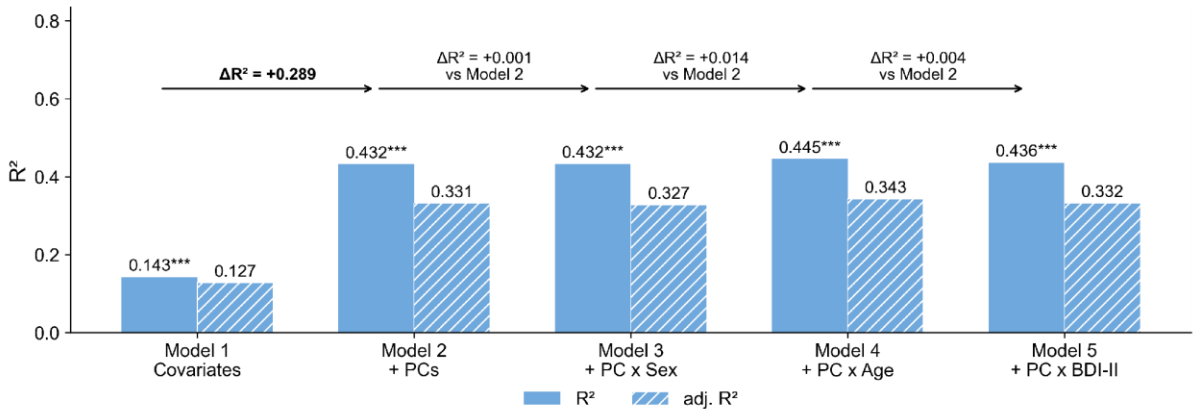
domains exceeded the conventional threshold of 0.50, with sampling adequacy highest for the energy/amplitude domain, followed by the frequency and spectral domains, reported in Appendix A, Table A4. In contrast, the temporal domain had a KMO value below 0.50, indicating limited suitability for dimensionality reduction. Results involving the temporal components were therefore interpreted with caution. At the component level, 11 of the 23 components differed between men and women, most strongly Frequency PC1 ($r = -0.93$, higher in women), which reflects overall voice pitch, whereas no component was associated with age or BDI-II score (Appendix B, Figure B1). Since the PCA was conducted separately per domain, components were uncorrelated within domains but correlated across domains (median $|r| = 0.12$, maximum $|r| = 0.76$, variance inflation factors in Appendix A, Table A7).

In-sample estimation of cognitive performance (MoCA)

Using hierarchical regression, we compared a covariate-only model (Model 1: sex, age, BDI-II score) with a model additionally including the 23 PC (Model 2), tested moderation by each covariate (Models 3–5), and assessed each component individually by correlation (see Methods).

Figure 3

Explained variance of the regression models for MoCA performance



Note. Model 1: covariates (age, sex, BDI-II), Model 2: + 23 PCs (principal components) and Models 3–5: Model 2 + interaction of Spectral PC10 with sex, age or BDI-II, respectively.

R^2: proportion of variance in MoCA scores explained by the model, adj R2: adjusted R^2, corrected for the number of predictors. Arrows show the change in R^2 (ΔR^2) from Model 1 to Model 2 and from Model 2 to Models 3–5. Asterisks mark the significance of the overall model F-test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Model 1 explained 14.3% of the variance in MoCA scores (adjusted $R^2 = 0.127$, $p < 0.001$). Lower MoCA scores were associated with older age ($r = -0.32$) and higher BDI-II scores ($r = -0.18$), whereas sex was not associated with MoCA performance (Table 1). Adding the prosodic components in Model 2 increased the explained variance to 43.2% (adjusted $R^2 = 0.331$), a difference of $\Delta R^2 = 0.289$ ($p < 0.001$), corresponding to a large effect (Cohen's $f^2 = 0.51$). In Model 2, age remained the only significant predictor. Among the 23 prosodic components, only Spectral PC10 reached the uncorrected significance threshold (p uncorrected = 0.048), but it did not survive FDR correction (p = 0.431, Appendix B, Table B1). This component mainly captures the variability of voice quality during speech (H1H2_SD, mfcc2V_SD and slpV5_15_SD, Appendix A, Table A6).

Table 1

Associations of the covariates with the clinical outcome

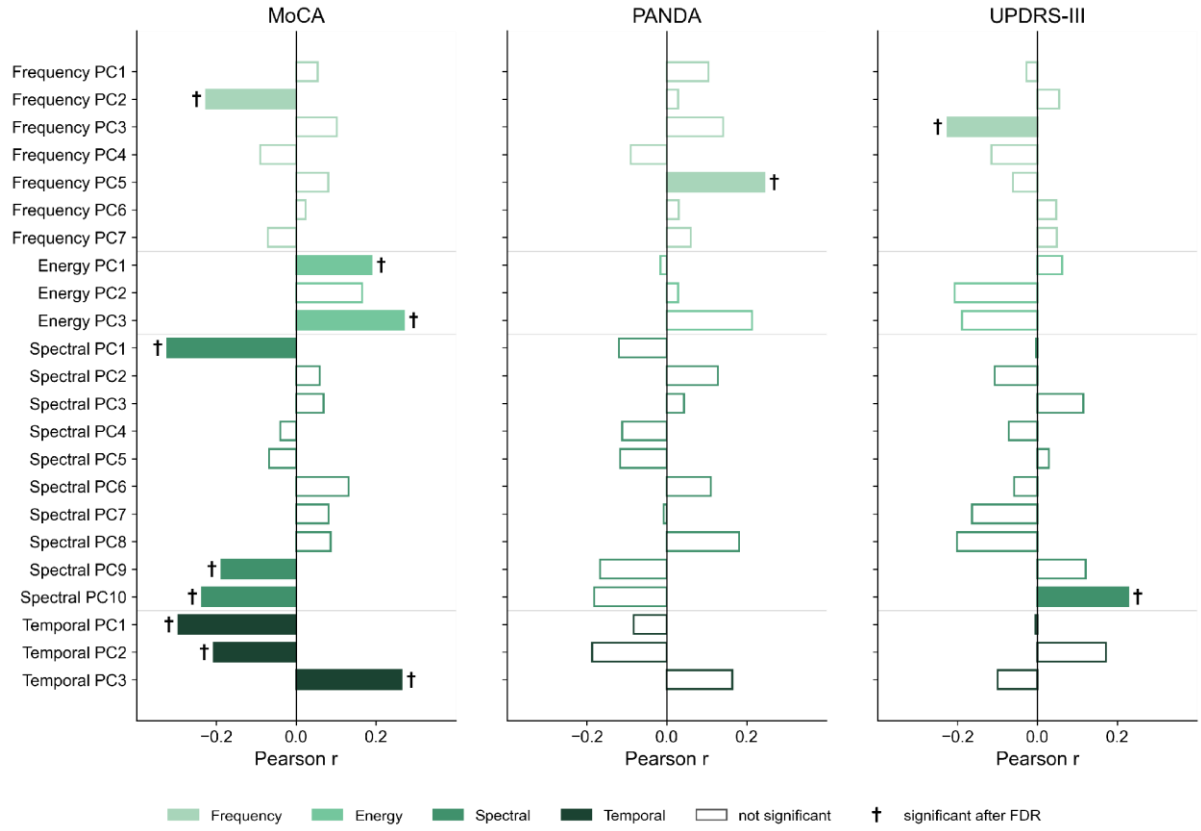
	MoCA	**PANDA**	**UPDRS-III**
Bivariate associations			
Age, r (*p* FDR)	**−0.319 (< 0.001)**	−0.141 (0.116)	0.087 (0.266)
BDI-II, r (*p* FDR)	**−0.178 (0.028)**	**−0.209 (0.025)**	0.150 (0.088)
Sex, rank-biserial r (*p* FDR)	−0.051 (0.581)	−0.134 (0.160)	0.179 (0.088)
Model 1 Covariates			
R^2 Model 1	0.143	0.080	0.044
Age, b [95% CI], *p* FDR	**−0.119 [−0.172, −0.066], < 0.001**	−0.089 [−0.192, 0.015], 0.098	0.089 [−0.123, 0.301], 0.408
BDI-II, b [95% CI], *p* FDR	**−0.088 [−0.155, −0.021], 0.016**	**−0.172 [−0.299, −0.046], 0.024**	0.242 [−0.008, 0.493], 0.172
Sex (male vs female), b [95% CI], *p* FDR	−0.567 [−1.521, 0.387], 0.242	−1.544 [−3.378, 0.290], 0.098	2.892 [−0.707, 6.490], 0.172

Note. p FDR = corrected across the three covariates within each outcome, separately for the bivariate associations, Model 1: linear regression of each outcome on age, sex and BDI-II.

Models 3 to 5 extended Model 2 by separately including the interaction of Spectral PC10 with sex (Model 3), age (Model 4), and BDI-II score (Model 5). Adding the interaction terms did not significantly improve model fit (Figure 3). Thus, there was no evidence that the association between Spectral PC10 and MoCA performance differed by sex, age, or depressive symptom severity. Supplementary analyses tested interaction effects for all prosodic components and all three outcomes. No interaction was significant, and no omnibus block test approached significance (Appendix B, Table B4). Although no individual prosodic component was identified as a significant predictor, the addition of the prosodic components to the covariate-only model increased the explained variance by 28.9 percentage points (ΔR^2=0.289). We therefore examined the association between each component and MoCA performance without adjustment for the other components. Nine of the 23 components were significantly associated with MoCA scores, as shown in Figure 4. Controlling for age, sex and BDI-II score changed the nine correlations only marginally, and eight of the nine remained significant (Appendix B, Table B5).

Figure 4

Correlations of the 23 prosodic components with MoCA, PANDA and UPDRS-III



Note. † significant after FDR correction across the 23 components of each outcome.

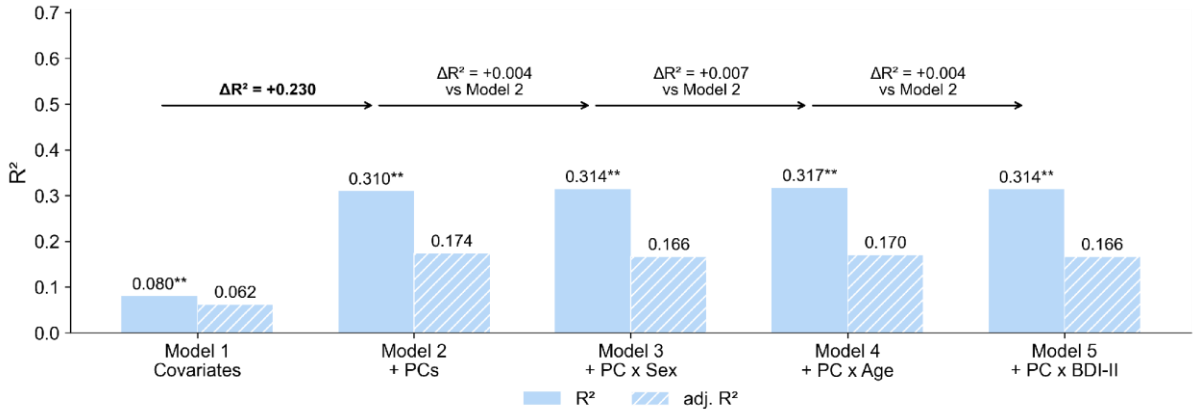
In-sample estimation of cognitive performance (PANDA)

Model 1 explained 8.0% of the variance in PANDA scores (adjusted $R^2 = 0.062$, $p = 0.005$). Lower PANDA scores were associated with higher BDI-II scores ($r = -0.21$), whereas age and sex were not associated with PANDA performance (Table 1). Adding the prosodic components in Model 2 increased the explained variance to 31.0% (adjusted $R^2 = 0.174$), a difference of $\Delta R^2 = 0.230$ ($p = 0.018$), corresponding to a medium effect (Cohen's $f^2 = 0.33$; Figure 5). No predictor was significant in Model 2. Among the prosodic components, Frequency PC5 and Spectral PC6 reached the uncorrected significance threshold (Appendix B, Table B2). Frequency PC5 mainly captures F0 rise characteristics and the variability of the second formant bandwidth (F0s_SDRise, F0s_mRise, F2BW_SD), Spectral PC6 the mean and variability of the third cepstral coefficient (mfcc3_SD, mfcc3_m, mfcc3V_m, Appendix A, Table

A6). Models 3 to 5 extended Model 2 by separately including the interactions of these two components with the covariates. None of the interaction terms was significant (Figure 5). At the component level, Frequency PC5 was the only component associated with PANDA performance (r = 0.244, p = 0.046, Figure 4), an association carried mainly by the variability of the second and third formant bandwidths (F2BW_SD, F3BW_SD, Appendix B, Table B7). This association remained significant when controlling for sex, age and BDI-II score (p = 0.016, Appendix B, Table B5).

Figure 5

Explained variance of the regression models for PANDA performance



Note. Model 1: covariates (age, sex, BDI-II), Model 2: + 23 PCs (principal components) and Models 3–5: Model 2 + interaction of Frequency PC5 and Spectral PC6 with sex, age or BDI-II, respectively. R^2: proportion of variance in PANDA scores explained by the model, adj R2: adjusted R^2, corrected for the number of predictors. Arrows show the change in R^2 (ΔR^2) from Model 1 to Model 2 and from Model 2 to Models 3–5. Asterisks mark the significance of the overall model F-test (* p < 0.05, ** p < 0.01, *** p < 0.001).

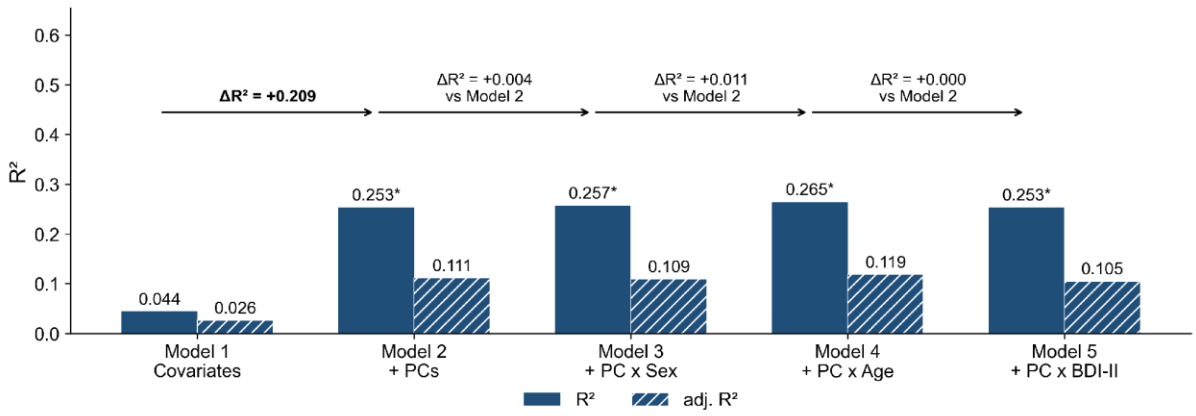
In-sample estimation of motor symptom severity (UPDRS-III)

For UPDRS-III scores, Model 1 did not reach significance (R^2 = 0.044, adjusted R^2 = 0.026, p = 0.064), and none of the covariates was associated with motor symptom severity (Table 1). Adding the prosodic components in Model 2 increased the explained variance to 25.3%

(adjusted R^2 = 0.111), a difference of ΔR^2 = 0.209 (p = 0.039), corresponding to a medium effect (Cohen's f^2 = 0.28; Figure 6). No predictor was significant in Model 2. Among the prosodic components, only Spectral PC3 reached the uncorrected significance threshold (Appendix B, Table B3). This component mainly captures the mean spectral slope and the first cepstral coefficient (slpV5_15_m, mfcc1_m, slpV5_15_SD, Appendix A, Table A6). Models 3 to 5 extended Model 2 by separately including the interaction of Spectral PC3 with the covariates. Adding the interaction terms did not significantly improve model fit (Figure 6). At the component level, Spectral PC10 (r = 0.227) and Frequency PC3 (r = −0.225) were associated with UPDRS-III scores (both p = 0.044), as shown in Figure 4. The association of Spectral PC10 was carried mainly by the variability of voice quality and of spectral flux (H1H2_SD, specFlux_SD), that of Frequency PC3 by the mean formant bandwidths and the F0 fall (F2BW_m, F1BW_m, F0s_mFall, Appendix B, Table B7). Neither association remained significant when controlling for age, sex and BDI-II score (both p = 0.081, Appendix B, Table B5).

Figure 6

Explained variance of the regression models for UPDRS-III score



Note. Model 1: covariates (age, sex, BDI-II), Model 2: + 23 PCs (principal components) and Models 3–5: Model 2 + interaction of Spectral PC3 with sex, age or BDI-II, respectively.

R^2: proportion of variance in UPDRS-III scores explained by the model, adj R2: adjusted R^2, corrected for the number of predictors. Arrows show the change in R^2 (ΔR^2) from Model 1 to Model 2 and from Model 2 to Models 3–5. Asterisks mark the significance of the overall model F-test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Cross-validated prediction

Predictive performance was assessed using cross-validated Ridge regression, comparing a covariate-only model with a model additionally including the prosodic components (see Methods). Cross-validated prediction performance is reported in Table 2. Figure 7 shows the distribution of out-of-fold R^2 values across 100 random partitions for the covariate-only model and the model including both covariates and PCs, together with their difference (ΔR^2) and the results for the fixed partition. Models based on the prosodic components alone showed a similar pattern, also when the covariates were regressed out of the components within each training fold (Appendix C, Table C2), indicating that prediction did not rely on information shared with the covariates.

Table 2

Cross-validated prediction of the clinical outcomes

	MoCA	**PANDA**	**UPDRS-III**
Pooled out-of-fold R^2			
Covariates only, R^2 (median)	0.104 (0.101)	0.003 (0.016)	−0.005 (−0.006)
Covariates + prosody, R^2 (median)	0.255 (0.206)	0.048 (0.073)	−0.011 (−0.027)
Prosody beyond covariates			
ΔR^2 (median)	**0.151 (0.105)**	**0.045 (0.053)**	−0.005 (−0.024)
p FDR (permutation test)	**0.003**	**0.008**	0.353

Note. R^2 = out-of-fold R^2 of the Ridge model on the fixed random partition used for inference. Covariates: age, sex and BDI-II. ΔR^2 = R^2 (covariates + prosody) − R^2 (covariates only). *p* FDR = permutation test of ΔR^2 (1000 permutations of the prosodic components), corrected across the three outcomes.

For MoCA, adding the prosodic components to the covariate-only model increased performance to $R^2 = 0.255$, corresponding to a gain of $\Delta R^2 = 0.151$ ($p = 0.003$). To identify the predictors contributing most to predictive performance, the model including covariates and PCs was refitted on the full sample for MoCA and PANDA (Figure 8). For MoCA, the largest coefficients besides age were observed for Temporal PC2, Spectral PC10, Temporal PC1, Spectral PC1 and Spectral PC9. As in the correlation analysis, higher MoCA scores were associated with a higher rate of voiced segments and loudness peaks (Temporal PC2), shorter

pauses with less variability in pause duration (Temporal PC1), lower variability in voice quality, reflected in H1–H2 difference and spectral flux (Spectral PC10 and PC9), and higher formant amplitudes with less variability (Spectral PC1, Appendix B, Table B7).

For PANDA, the covariates alone did not predict performance ($R^2 = 0.003$), whereas the model including the prosodic components reached an R^2 of 0.048 ($\Delta R^2 = 0.045$, $p = 0.008$). Although small, this gain was consistent across partitions. The largest coefficients besides BDI-II score were observed for Frequency PC5, Frequency PC6, Spectral PC4 and Temporal PC2. Frequency PC5, the only component associated with PANDA in the correlation analysis, contributed in the same direction as in that analysis. Its association with PANDA was mainly driven by variability in the second and third formant bandwidths (F2BW_SD and F3BW_SD, Appendix B, Table B7).

For UPDRS-III, neither model outperformed the sample mean across the majority of partitions (covariate-only model: $R^2 = -0.005$, model including prosodic components: $R^2 = -0.011$), indicating an absence of predictive power for motor symptom severity. As neither model showed valid predictive performance for UPDRS-III, component importance was not investigated.

Figure 7

Cross-validated prediction of MoCA, PANDA & UPDRS-III

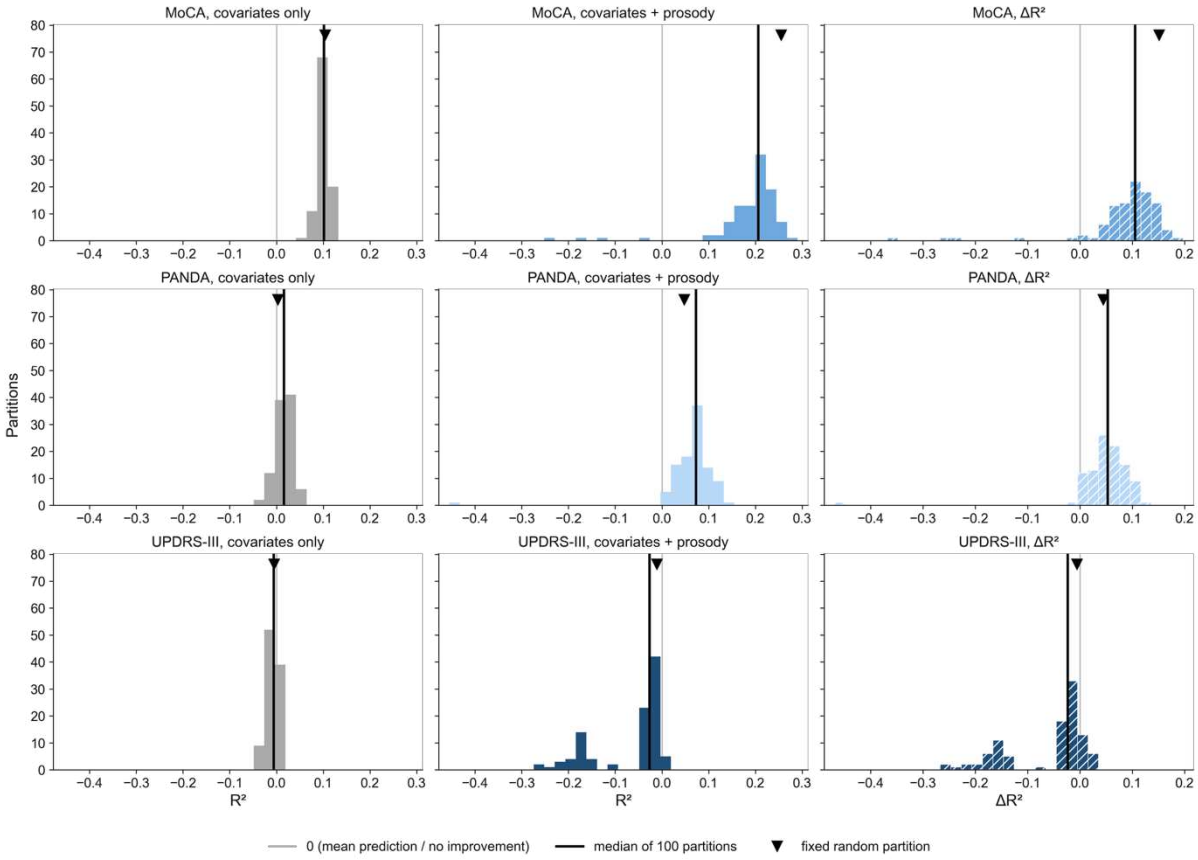
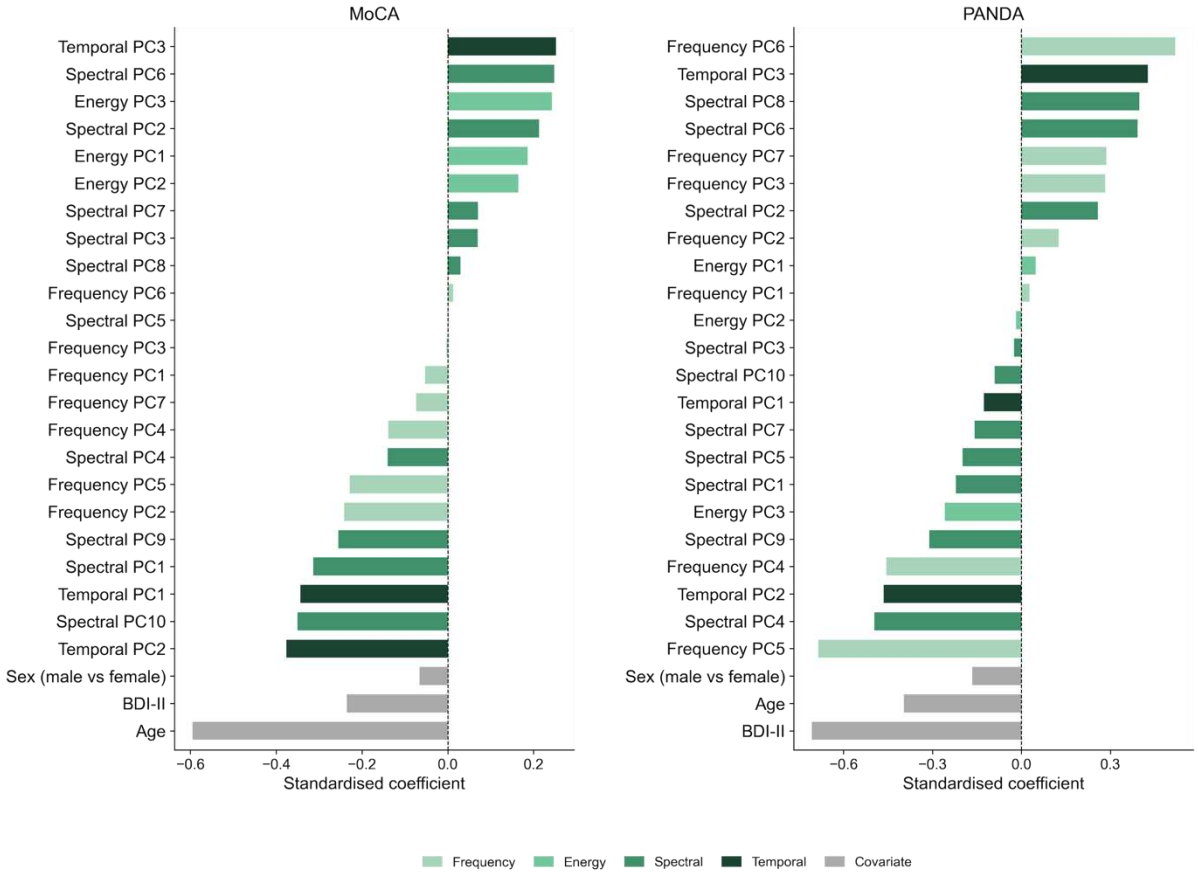

MoCA, covariates only
MoCA, covariates + prosody
MoCA, ΔR²
PANDA, covariates only
PANDA, covariates + prosody
PANDA, ΔR²
UPDRS-III, covariates only
UPDRS-III, covariates + prosody
UPDRS-III, ΔR²
Partitions
R²
R²
ΔR²
0 (mean prediction / no improvement)
median of 100 partitions
fixed random partition

Figure 8

Component importance of MoCA and PANDA performance



Note. Coefficients of the covariate and prosody model refitted on the full sample for MoCA and PANDA. Because the PCA is refitted within the prediction pipeline, component signs are arbitrary and can be reversed relative to the in-sample analysis.

DISCUSSION

In summary, in 176 individuals with PD, prosodic features were shown to be associated with cognitive performance beyond that explained by sex, age, or depressive symptoms, whereas associations with motor symptom severity were weak in-sample and did not generalise to unseen individuals.

First, the hierarchical regression tested whether prosodic features as a whole explained additional variance in each clinical measure beyond sex, age and depressive symptoms. Second, individual regression coefficients estimated each prosodic component's contribution, although strong collinearity limited their interpretation. Therefore, component-level correlations assessed associations with the outcome and allowed for identification of relevant prosodic components. Finally, cross-validated Ridge regression assessed whether the prosodic components improved prediction beyond the covariates in unseen individuals.

Overall, the covariate-only models explained a modest proportion of the variance in cognitive performance and did not reach significance for motor symptom severity. Adding the prosodic components increased the explained variance for all three outcomes, but only for the two cognitive measures did this gain carry over to held-out individuals, where prosody improved prediction beyond sex, age and depressive symptoms. Since this was held for both the MoCA, a screening of global cognition, and the PANDA, which was developed specifically for PD, the association does not appear to depend on a single cognitive screening test. No individual prosodic component remained significant as a partial coefficient in the regression model after correction, potentially due to collinearity, which inflates the variance of regression coefficients (O'Brien, 2007). Consequently, the unique contribution of individual components to cognitive performance could not be reliably identified. However, the subsequent correlation analyses revealed significant associations with MoCA performance for nine prosodic components across all four domains and with PANDA performance for one component, providing insights into which prosodic features were associated with cognitive performance. Specifically, all three temporal components were associated with MoCA performance, as were two of the three energy/amplitude components, three of the ten spectral components, and one of the seven frequency-related components. Thus, MoCA performance was related to all temporal and most energy components, i.e. to the timing of speech and to vocal intensity, but only to some of the spectral and frequency-related components.

Interpreting these associations requires considering the acoustic properties captured by the underlying features and components: The frequency domain captures fundamental frequency (F0, perceived as pitch), its variability and perturbation (jitter), as well as formant frequencies and bandwidths. The energy domain captures loudness, its variability and amplitude perturbation (shimmer and harmonics-to-noise ratio). The spectral domain describes the distribution of acoustic energy across frequencies, reflecting voice quality and articulation (e.g. spectral slope, H1–H2 difference, formant amplitudes, cepstral coefficients and spectral flux). The temporal domain captures speech timing, including the rate of loudness peaks and the duration and rate of voiced and unvoiced segments. Each component summarises correlated features within its respective domain. Appendix B, Table B7 shows which features contributed to the association between each component and the respective outcome. According to the correlation results, better MoCA performance was associated with louder speech overall (Energy PC1) and with less variability in loudness combined with a higher lower-percentile loudness (Energy PC3). Better MoCA performance was further linked to a higher rate of loudness peaks (Temporal PC3), to shorter and less variable pauses (Temporal PC1) and to a higher rate of voiced segments and loudness peaks (Temporal PC2). Poorer MoCA performance, in contrast, went along with lower F0 variability and lower jitter (Frequency PC2),

with lower and more variable formant amplitudes (Spectral PC1, an association spread across many features) and with greater variability of the H1–H2 difference and of spectral flux (Spectral PC10, Spectral PC9, Appendix B, Table B7). Moreover, spectral PC10 was the only component also associated with motor symptom severity, in the same direction, but unlike its association with MoCA, this association did not survive adjustment for age, sex and depressive symptoms.

For PANDA, the only associated component, Frequency PC5, was defined primarily by F0 rise characteristics, but its association was carried mainly by the variability of the second- and third-formant bandwidths, whereas the pitch-related features contributed less (Appendix B, Table B7). Frequency PC5 was also among the largest coefficients of the cross-validated PANDA model, with the same direction of association. As formant bandwidths reflect acoustic damping in the vocal tract, which depends on articulatory configuration and phonation, more variable bandwidths may indicate a more dynamic vocal tract configuration. Since this association rests on a single component, this interpretation remains tentative. The components contributing most to the cross-validated MoCA prediction were among those identified by the correlation analysis and carried the same sign, so that the in-sample and out-of-sample analyses point to the same speech characteristics.

Our results suggest that in this sample, better cognitive performance in PD was associated with speech that was louder, steadier in intensity, more varied in pitch, more stable in voice quality and characterised by shorter pauses and a higher rate of voiced segments. Conversely, poorer cognitive performance in PD was associated with a speech profile characterised by reduced intensity, greater monotony, longer pauses, and less stable voice quality. Notably, these features align with hypokinetic dysarthria (Darley et al., 1969), a speech impairment typically reflected in altered speech rate, pause duration, speech intensity, and pitch variation (Rusz et al., 2011; Skodda & Schlegel, 2008). Accordingly, most of the features underlying the associated components also differed between the PD group and the age- and sex-matched reference group (Appendix A, Table A3), although not always in the same direction, indicating that cognitive performance relates largely to prosodic characteristics typical of PD rather than to incidental variation. At the same time, several of these measures, including the variability of voiced-segment length and of the lower cepstral coefficients have also been associated with clinically relevant outcomes in other conditions assessed with the same feature set (Bamberger et al., 2026; Haider et al., 2020), suggesting that their clinical relevance extends beyond PD.

The observed associations between prosodic features and cognitive performance may reflect two partly overlapping mechanisms. First, some prosodic features may capture cognitive

processes involved in speech production. Spontaneous speech requires planning, lexical retrieval and monitoring, drawing on cognitive resources such as working memory, attention and executive control, and its temporal organisation depends on these processes (Ferreira, 1991; Smith & Caplan, 2018). The associations between temporal components and MoCA performance are consistent with this account, although the impact of the temporal components should be interpreted with caution given the limited sampling adequacy of this domain. Similar findings have been reported in PD, where speaking rate and pausing were related to cognitive performance but not to motor symptom severity (Andrade et al., 2023). Beyond PD, longer and more frequent pauses and slower speech have been associated with mild cognitive impairment (Tóth et al., 2018) and early tau burden in cognitively unimpaired adults (Young et al., 2024), while pause-related information has improved the automated detection of Alzheimer's disease (Yuan et al., 2021). However, pauses are not specific to cognitive processes. They may also reflect prosodic structuring, lexical retrieval difficulties, increased cognitive load or respiratory and motor constraints (Smith & Caplan, 2018). Second, features related to phonation and voice quality, such as variability in the H1–H2 difference and in formant amplitudes, are less likely to reflect cognitive operations directly, whereas F0 variability may reflect both, as pitch modulation also serves prosodic phrasing. The associations of phonatory features with cognition may instead arise from shared neurodegenerative processes affecting both cognitive functioning and the stability of speech production. That these phonatory components were not robustly related to UPDRS-III scores argues against an explanation in terms of general motor severity, although speech-specific motor impairment cannot be ruled out, since the UPDRS-III captures speech only marginally. Taken together, the observed pattern may reflect contributions from both the cognitive control of speech and the integrity of speech production.

In addition to this interpretation, Pizzimenti and colleagues (2026) recently proposed a framework that assigns prosodic features to the system primarily producing them, with overlapping cardio-respiratory, voice/phonatory, speech/articulatory and cognitive/language levels. Within this framework, the three temporal components, which capture speech rate and the timing of voiced and unvoiced segments, belong to the speech/articulatory level, although pauses may also reflect respiratory support. Components relating to F0 variability, jitter and harmonic differences (Frequency PC2, Spectral PC9, Spectral PC10) belong to the voice/phonatory level. Loudness and formant amplitudes are not assigned to a specific level, so the energy components and Spectral PC1 cannot be placed unambiguously. Cognitive performance is thus associated with features at both the speech/articulatory and the voice/phonatory level.

Furthermore, associations with motor symptom severity were weaker and not robust. Although the prosodic components explained additional variance in UPDRS-III scores in-sample, neither model predicted motor severity in held-out individuals, and the two components associated with UPDRS-III did not remain significant when controlling for age, sex and depressive symptoms. One reason may lie in the outcome measure itself. Of the 33 scored items of the UPDRS- III, only one rates speech (item 3.1) and one facial expression (item 3.2), so that speech contributes at most 4 of 132 points, whereas rigidity, bradykinesia and tremor of the limbs dominate the total score (Goetz et al., 2008). The total score may therefore be relatively insensitive to the speech-motor impairments that prosodic features capture, and the absence of a robust association should not be read as evidence that prosody is unrelated to motor function. Notably, speech and motor severity were both assessed off medication, whereas cognition was assessed on medication, so a shared medication state cannot explain why prosody related more closely to cognition.This is consistent with the view that prosody reflects not only the motor execution of speech but also its cognitive control (see above), and with the finding that speech difficulties in early PD predict faster cognitive decline but not motor progression (Polychronis et al., 2019). In contrast to our findings, Wiesman and colleagues reported that rated prosody impairment actually related to motor symptom severity, but not to MoCA performance, while only articulation related to cognition (2023). Two methodological differences may help explain this discrepancy. First, their prosody measure was based on a comparative perceptual rating of four repeated sentences by non-expert raters, which captures overall prosodic impairment rather than the specific temporal dimensions associated with cognitive performance in our study. Second, their individuals were assessed on medication, which further limits direct comparison. These findings suggest that different aspects of prosody may relate differently to motor and cognitive functioning, with perceived prosodic impairment potentially reflecting motor dysfunction and temporal speech characteristics being more closely associated with cognitive performance.

Our central question was how much prosody contributes to the estimation of clinical outcomes beyond sex, age and depressive symptoms, which may be related to both prosody and the clinical outcomes. For both cognitive measures, the covariates alone explained a modest share of the variance, and adding prosody increased it. In held-out individuals, prosody likewise improved prediction beyond the covariates, including for the PANDA, which the covariates alone did not predict. Prosody thus provided information about cognition beyond sex, age and depressive symptoms.
The covariates themselves related to prosody and cognition in distinct ways. Sex was associated with a large proportion of the prosodic features, with 46 of the 88 features and 11 of the 23 components differing between women and men, consistent with anatomical

differences in vocal-fold mass and vocal-tract length and with sex-related patterns of dysprosody in PD (Skodda et al., 2011a), but it was not associated with any of the three outcomes. Age, in contrast, was associated with MoCA performance but with no prosodic feature or component, consistent with evidence that the core prosodic abnormalities of PD occur independently of age at onset (Rusz et al., 2021a). Depressive symptoms were weakly associated with both cognitive measures, but with no prosodic feature or component after correction, although speech markers of depression have been reported in PD (Pérez-Toro et al., 2022). Depressive symptoms nevertheless warrant particular attention, because depression is among the most frequent non-motor symptoms of PD (Reijnders et al., 2008) and can lower cognitive test performance, making it a plausible confound of speech–cognition associations. Importantly, the cognitive associations were largely unaffected by controlling for sex, age and depressive symptoms, and none of these factors moderated them, suggesting that they are not explained by demographic or affective differences between individuals. The associations with motor severity, in contrast, did not survive this adjustment. As most individuals reported minimal depressive symptoms, this conclusion applies to the range observed here and should be tested in samples with more severe depression. Prosodic measures should therefore be interpreted alongside sex, age and depressive symptoms rather than in isolation, and future studies should report a covariate-only baseline model, both in-sample and in cross-validation, to quantify the contribution of prosodic information beyond these factors.

Our findings have implications for the use of speech markers in clinical diagnostics. Speech can be recorded quickly and non-invasively, and prosodic features can be extracted automatically, which makes them promising for a low-burden and repeatable assessment of cognitive performance in PD. The picture description task appears well suited for this purpose, since it elicits spontaneous speech that requires planning while remaining short and standardised, and it has already proven informative for cognitive impairment in PD (Andrade et al., 2023; Smith et al., 2025) and Alzheimer's disease (Bang et al., 2024). Which individual features carry the most diagnostic value, however, remains to be established, as prosodic components summarise several correlated features. Future studies should therefore test specific hypotheses on single, automatically extracted features, in particular those that carried the associations observed here, such as pause duration and variability, the rate of voiced segments, loudness and the variability of harmonic differences to combine scalability with interpretability.

Following the recent consensus definitions and recommendations for speech-based measures (Cummins et al., 2026; Pizzimenti et al., 2026), the prosodic features reported here are

measures with a demonstrated association with cognitive performance, but not validated biomarkers, so that several limitations should be considered. First, although cross-validated prediction of cognitive performance was successful, the gains were modest for MoCA and small for PANDA, which limits the direct clinical applicability of the models at the level of the individual patient. In addition, the limited sample size restricts the generalisability of our findings. Second, the prosodic components were strongly collinear, which makes the partial coefficients uninterpretable. No conclusion about the unique contribution of a single component can be drawn from our data, and this also applies to the Ridge coefficients. Third, the temporal domain showed limited suitability for dimensionality reduction, and its unvoiced-segment measures loaded positively on both Temporal PC1 and Temporal PC3, which correlated with MoCA performance in opposite directions. The temporal components therefore reflect pausing behaviour less clearly than the energy and spectral components reflect intensity and voice quality. This is relevant, since two of the five most heavily weighted predictors come from the temporal domain. Fourth, no test-retest data were collected. We can therefore not report intraclass correlations, standard errors of measurement or minimal detectable changes, although these are needed before a change in a measure can be interpreted clinically (Cummins et al., 2026).

In conclusion, our findings suggest that prosodic measures are more robustly associated with cognitive performance than with motor symptom severity. To our knowledge, this is the first study to systematically examine automatically extracted prosodic features from spontaneous speech in a large sample of individuals with PD in the medication-off state. By relating these features to both cognitive performance and motor symptom severity in the same individuals, we demonstrated their association with cognitive performance and showed that they improve the prediction of cognitive performance beyond sex, age and depressive symptoms in unseen individuals. Our findings highlight the potential of prosody as a marker of cognitive performance, although further validation is needed before conclusions about individual patients can be drawn. In addition to longitudinal data and substantially larger samples, standardised elicitation and recording settings, demonstrated test-retest reliability, and validation within a defined context of use are required before any prosodic feature can be considered a reliable biomarker.

METHODS

Participants and study design

Individuals with Parkinson's Disease (PD) were recruited at the Department of Neurology of the University Hospital Cologne between May 2021 and March 2026. Diagnosis was established by a neurologist according to the International Parkinson and Movement Disorder Society (MDS) clinical diagnostic criteria for PD (Postuma et al., 2015).The study was approved by the local ethics committee of the University Hospital Cologne under the registration numbers 23-1461 and 23-1355 and all participants provided written informed consent. Participants were native German speakers living in North Rhine-Westphalia, Germany.

In total, 176 subjects (61 female & 115 male) with a mean age of 62.68 years (SD=8.78) were assessed. Disease duration averaged 8.4 years (SD=4.58), and motor severity ranged from Hoehn & Yahr stage 1-4. Our sample was not restricted to early-stage individuals or a narrow severity range but covered a broad clinical spectrum. Since only individuals able to complete the study protocol were included, the most severely affected individuals had to be excluded by design. The neuropsychological assessments reported in this study were conducted in a regular ON medication state, while speech and motor severity assessments were conducted in an OFF-medication state, with dopaminergic medication withheld for at least 12 hours.

Clinical assessments

Global cognitive performance was assessed with the MoCA (Nasreddine et al., 2005). Scores range from 0 to 30, with scores below the conventional cutoff of 26 indicating cognitive impairment. Additionally, the PANDA (Kalbe et al., 2008) was administered. PANDA scores below the recommended cutoff of 18 likewise indicate possible cognitive impairment. Whereas the MoCA screens global cognitive performance, the PANDA was developed specifically for PD and focuses on the cognitive functions typically affected in the disease, such as executive functions and working memory. Using both allowed us to examine whether associations with prosody hold for a general and a disease-specific measure of cognition. Motor symptom severity was quantified by a certified rater using UPDRS-III (Goetz et al., 2008) and depressive symptoms were assessed with the BDI-II (Beck et al., 1996) completed by the patients. All instruments were administered in their validated German versions. The two cutoffs mentioned above are reported for descriptive purposes only, in order to convey where the present sample is located relative to established clinical thresholds. All four measures entered the statistical analyses as continuous variables; no dichotomisation was performed at any point.

Speech recording and prosodic feature extraction

Spontaneous speech was elicited by asking participants to describe a picture from the Bogenhausener Dysarthrieskalen (BoDyS, Ziegler et al., 2018) in as much detail as possible

for 90 seconds. Recordings were made in a quiet room. The analysis window began with the participant's first utterance and comprised the following 90 seconds. Speech beyond this window was not analysed. Recordings were made in a quiet environment and spontaneous speech was recorded using a condenser headset microphone (AKG C 520; 44.1 kHz, 16-bit), ensuring a consistent mouth-to-microphone distance across participants and recording sessions. In accordance with recommendations from the relevant literature (Patel et al., 2018; Rusz et al., 2021b), the microphone was placed 5 cm from the participant's lips and oriented at a 45° angle away from the mouth. The microphone features a cardioid polar pattern, a frequency range of 50–20,000 Hz, and a flat frequency response above 200 Hz, thereby reducing unwanted room reflections. The gain setting remained unchanged across recording sessions and experimental conditions.

Prosodic features were extracted with the toolbox open-Source Media Interpretation by Large feature-space Extraction (openSMILE, Eyben et al., 2010), version 2.5.0. For extraction, the eGeMAPS (Eyben et al., 2016) was chosen, which comprises 88 features. The benefit of this toolbox lies in its standardised, fully automatic computation of a fixed feature set, which addresses the lack of comparability that has repeatedly been observed for manually derived prosodic measures (Hecker et al., 2022). Following the classification underlying the eGeMAPS parameter set, the 88 features were assigned to the four prosodic domains that structure the entire analysis: 24 frequency related features, 15 energy/amplitude related features, 43 spectral features, and 6 temporal features. Frequency-related parameters comprised measures of F0, jitter, formant frequencies, and formant bandwidths.

Energy/amplitude-related parameters included loudness, shimmer, harmonics-to-noise ratio, and equivalent sound level. Spectral parameters comprised spectral flux, mel-frequency cepstral coefficients (MFCC), spectral balance measures, spectral slopes, harmonic differences, and formant amplitude ratios. Temporal parameters captured the rate of loudness peaks and voiced segments, alongside the mean duration and variability of voiced and unvoiced segments. This grouping defines the units within which dimension reduction was performed and within which FDR correction was applied at the feature level. Table 3 presents each feature together with its corresponding set of abbreviated feature variables and a description of the feature. The original openSMILE toolbox feature names corresponding to all abbreviated feature variables are provided in Appendix A, Table A3.

Table 3

Prosodic features extracted with openSMILE using eGeMAPS.

Feature	Feature Variables	Description
Frequency related parameters		
F0 (semitone scale)	F0s_m, F0s_SD, F0s_pctl20, F0s_pctl50, F0s_pctl80, F0s_rng, F0s_mRise, F0s_SDRise, F0s_mFall, F0s_SDFall	Pitch, logarithmic F0 on a semitone frequency scale starting at 27.5 Hz
Jitter	jitter_m, jitter_SD	Deviations in individual consecutive F0 period lengths
Formant frequencies 1-3	F1_m, F1_SD, F2_m, F2_SD, F3_m, F3_SD	Centre frequency of the first, second and third formant
Formant bandwidths 1-3	F1BW_m, F1BW_SD, F2BW_m, F2BW_SD, F3BW_m, F3BW_SD	Bandwidth of the first, second and third formant
Energy / amplitude related parameters		
Loudness	loud_m, loud_SD, loud_pctl20, loud_pctl50, loud_pctl80, loud_rng, loud_mRise, loud_SDRise, loud_mFall, loud_SDFall	Estimate of perceived signal intensity from an auditory spectrum
Shimmer	shimmer_m, shimmer_SD	Difference in peak amplitudes of consecutive F0 periods
Harmonics-to-noise ratio	HNR_m, HNR_SD	Relation of energy in harmonic components to energy in noise-like components
Equivalent sound level	SPL	Sound pressure level with the same total energy as the actual fluctuating signal
Spectral parameters		
Spectral flux	specFlux_m, specFlux_SD, specFluxV_m, specFluxV_SD, specFluxUV_m	Difference between the spectra of two consecutive frames
Mel Frequency Cepstral Coefficients 1-4	mfcc1_m, mfcc1_SD, mfcc2_m, mfcc2_SD, mfcc3_m, mfcc3_SD, mfcc4_m, mfcc4_SD, mfcc1V_m, mfcc1V_SD, mfcc2V_m, mfcc2V_SD, mfcc3V_m, mfcc3V_SD, mfcc4V_m, mfcc4V_SD	Perceived pitch of the frequency spectrum, overall and for voiced regions
Alpha ratio	alphaRV_m, alphaRV_SD, alphaRUV_m	Ratio of summed energy from 50-1000 Hz to 1-5 kHz

Hammarberg index	hammV_m, hammV_SD, hammUV_m	Ratio of the strongest energy peak in the 0-2 kHz region to that in the 2-5 kHz region
Spectral slopes	slpV0_5_m, slpV0_5_SD, slpV5_15_m, slpV5_15_SD, slpUV0_5_m, slpUV5_15_m	Linear regression slope of the logarithmic power spectrum in the specified bands
Harmonic differences	H1H2_m, H1H2_SD, H1A3_m, H1A3_SD	Energy ratio of the first F0 harmonic to the second harmonic, and to the highest harmonic in the third formant range
Formant amplitudes 1-3	F1amp_m, F1amp_SD, F2amp_m, F2amp_SD, F3amp_m, F3amp_SD	Relative energy of the first, second and third formant
Temporal parameters		
Loudness peaks per second	loudPPS	Number of loudness peaks per second
Voiced segments	VoicedPS, VSegLen_m, VSegLen_SD	Rate and length of continuously voiced regions
Unvoiced segments	UVSegLen_m, UVSegLen_SD	Length of continuously unvoiced regions

Note. m = arithmetic mean, SD = normalised standard deviation, pctl = percentile, rng = 20th-80th percentile range, mRise/mFall = mean rising/falling slope, SDRise/SDFall = standard deviation of rising/falling slope, V = voiced regions, UV = unvoiced regions.

Statistical analysis

Data management and analysis were performed using Python 3, version 3.10.7. Analyses proceeded in mainly three parts: the prosodic features were first reduced to domain-specific components, these components were then related to the clinical measures within the PD sample, and a cross-validated prediction analysis finally tested whether this information transfers to unseen individuals with PD.

Domain-specific principal component analysis

Given the sample size and the pronounced intercorrelations within the eGeMAPS feature set (Appendix A, Table A4), we performed separate PCA within each of the four prosodic domains. Separate decompositions keep each component interpretable within one prosodic domain, retain the domain structure of the feature set, and prevent the spectral domain, which contributes 43 of the 88 features, from dominating a joint decomposition. Features were z-scored prior to decomposition. Within each prosodic domain, components were retained until

the cumulative explained variance reached 80%, allowing components explaining smaller proportions of variance to capture potentially relevant domain-specific acoustic information. The resulting component scores were combined into one predictor set, with each component retaining its prosodic domain label. Sampling adequacy for dimensionality reduction was assessed separately for each prosodic domain using the KMO measure and the mean absolute intercorrelation (Appendix A, Table A4). Components are described by their three highest-loading features. Because PCA was conducted separately for each domain, the resulting components were orthogonal within domains but could remain correlated across domains. As components from all four domains were subsequently entered into the same regression models, variance inflation factors were calculated to assess potential cross-domain multicollinearity (Appendix A, Table A7). Component-level tests were corrected across all 23 components of a given outcome, pooling the domains, since the components form a single predictor set.

Within-sample association analysis

Within the PD sample, the components were related to each clinical measure using ordinary least squares regression models estimated in a stepwise manner. Model 1 served as a covariate-only baseline model and included sex, age, and BDI-II total score. Given the established associations of demographic factors and depressive symptoms with both prosodic characteristics and clinical performance, these variables were included to account for their potential contribution to the association of interest. Model 2 added the prosodic PC scores to the covariates, thereby testing the incremental contribution of prosodic information to the outcome beyond that explained by sex, age, and depressive symptoms. Accordingly, the change in explained variance from Model 1 to Model 2 quantifies the additional variance accounted for by the prosodic components. Models 3 to 5 extended Model 2 to examine whether associations between prosodic components and the outcome varied as a function of sex, age, or depressive symptoms. For each moderator, interaction terms were added between the respective variable and each prosodic component that had reached the uncorrected significance threshold in Model 2, thereby focusing these models on the components showing evidence of an association with the outcome and keeping the ratio of predictors to observations balanced. Analyses testing all prosodic component × moderator interactions, together with omnibus tests of whether each block of interactions improved model fit, are reported in Appendix B, Table B4. Model fit was evaluated using R^2, adjusted R^2 and Cohen's f^2. R^2 indicates the proportion of variance in outcome scores explained by the model. Because R^2 increases with every added predictor even in the absence of a true association, adjusted R^2 corrects R^2 for the number of predictors relative to the sample size. Differences

between nested models were quantified by the change in explained variance (ΔR^2), tested by the F-test for the increment, which compares the gain in explained variance per added predictor against the variance still unexplained. Since a large number of coefficients was tested within each model, p-values were corrected using the Benjamini-Hochberg procedure controlling the FDR at $\alpha = 0.05$. Correction was applied within families of tests: across the coefficients of a given model, across the three covariates within each outcome for the covariate-outcome associations, across the 23 component-level correlations and partial correlations for a given outcome, across the interaction tests of a given moderator, across the three outcomes for the F-test of the prosodic increment, and, at the feature level, within each of the four prosodic domains separately. The omnibus F-test of a model constitutes a single test and was not corrected. Unless stated otherwise, all p-values reported in the Results are FDR-corrected. All tests were two-sided, except for the F-tests of model fit and the permutation tests of predictive performance, which are one-sided by construction.

To further characterise the associations between prosody, the clinical measures and the covariates, several complementary analyses were conducted. Sex differences were tested throughout with Mann–Whitney U tests, using the rank-biserial correlation as effect size (positive values indicating higher values in men). All other associations were assessed with Pearson correlations. First, the clinical measures and covariates were intercorrelated (Appendix A, Table A5). Second, each of the 88 untransformed prosodic features was related to sex, age and BDI-II score (Appendix B, Table B6). In a supplementary comparison, the features of the PD group were compared with those of the matched reference group using Mann–Whitney U tests complemented by Cohen's d (Appendix A, Table A3). Third, the same associations with sex, age and BDI-II score were computed for each of the 23 prosodic components (Appendix B, Figure B1). Fourth, the covariates were related to each outcome, bivariately and jointly in Model 1 (Table 1). Finally, each prosodic component was correlated with each outcome, both without adjustment and controlling for age, sex and BDI-II score (partial correlations, Appendix B, Table B5). To identify which features carried a significant component-level association, each correlation was decomposed post hoc into feature contributions (loading × feature–outcome correlation, scaled so that the contributions sum to the component correlation, Appendix B, Table B7).

Cross-validated prediction

To assess the predictive performance of the prosodic features in unseen individuals, we additionally evaluated them in a cross-validated prediction framework. Analyses were performed with julearn (Hamdan et al., 2024), version 0.3.4, which builds on scikit-learn and

enforces that every preprocessing step is fitted within the training partition of each cross-validation fold. Within each training fold, the 88 prosodic features were z-scored and decomposed by separate PCA per prosodic domain retaining components until 80% of the domain variance was reached. Because the decomposition is refitted on each training partition, the number of retained components can differ slightly between folds. The prosodic components were passed to the model and the fitted transformations were then applied unchanged to the held-out fold. Mirroring the hierarchical regression, two models were compared for each outcome on the same folds: a covariate-only model including sex, age and BDI-II score, and a model including the covariates together with the prosodic components. Covariates and components were z-scored within the training fold and entered the same Ridge model. The difference in out-of-fold R^2 between the two models (ΔR^2) quantifies how much the prosodic components improve prediction in unseen individuals beyond the covariates. In addition, models with the prosodic components alone were evaluated with and without regressing the covariates out of the components within each training fold. The latter serves as a sanity check for confound leakage (Appendix C, Table C2). Model performance was assessed by 10-fold cross-validation with individuals randomly assigned to folds under a fixed seed. The sample is divided into ten parts of equal size, and each part serves once as the held-out test set while the model is fitted on the remaining nine. Every patient therefore belongs to the test set once and receives one prediction, by a model that was fitted without them. Across the ten folds this yields one prediction for every patient in the analysis sample. Performance is reported as the pooled out-of-fold R^2, computed once over the complete vector of predictions, with the mean of the whole sample as the reference against which the model is judged. Inference was based on one fixed random partition of individuals into folds. To assess how strongly performance depends on this assignment, the cross-validation was repeated across 100 random partitions. The median pooled R^2 and ΔR^2 across partitions are reported (Table 2), and their distributions are shown in Figure 7. Additionally, the mean absolute error (MAE) is reported in Appendix C, Table C1. Ridge regression was used as the estimator for all prediction models. Ridge adds a penalty proportional to the squared magnitude of the coefficients, which shrinks them towards zero and trades a small amount of bias for a large reduction in variance (Hastie et al., 2009; Hoerl & Kennard, 1970). It is therefore the appropriate linear estimator when predictors are collinear and the number of observations per predictor is modest. The Ridge penalty was tuned within the training data and never on the fold used for evaluation. Each training fold was divided into three parts: the model was fitted with each of seven candidate penalties, logarithmically spaced from 0.01 to 10,000, on two of those parts and scored on the third. The penalty performing best on average was then refitted on the whole training fold and applied once to the held-out individuals.

Statistical significance was assessed by permutation. In each permutation, the rows of the prosodic feature block were shuffled while the outcome and the covariates kept their original alignment. This preserves the relationship between covariates and outcome while removing any association between the prosodic features and the outcome. The full pipeline was rerun on each of 1000 permuted datasets. For the comparison of the two models, the prosodic block was shuffled in the covariate-plus-prosody model, and the observed ΔR^2 was compared with the resulting null distribution of ΔR^2. For the models based on the prosodic components alone (Appendix C, Table C2), the observed pooled R^2 was compared with its null distribution. P-values were computed as $(k + 1)/(N + 1)$, where k is the number of permutations reaching the observed value and N the number of permutations (North et al., 2002), and were FDR corrected across the three outcomes, separately for each model comparison.

Finally, component importance was quantified using the signed coefficients of the covariate-plus-prosody Ridge model refitted on the full sample after cross-validation. Because the components and covariates were z-scored before the estimator, the coefficients share a common scale. Their magnitudes can therefore be compared directly, and their signs indicate the direction of association, as in the regression models above. On the other hand, since the PCA is refitted within the prediction pipeline, component signs are arbitrary and may be reversed relative to the in-sample analysis.

REFERENCES

1. Aarsland, D., Batzu, L., Halliday, G. M., Geurtsen, G. J., Ballard, C., Ray Chaudhuri, K., & Weintraub, D. (2021). Parkinson disease-associated cognitive impairment. *Nature Reviews Disease Primers, 7*(1), 47. https://doi.org/10.1038/s41572-021-00280-3
2. Andrade, E. I. N., Manxhari, C., & Smith, K. M. (2023). Pausing before verb production is associated with mild cognitive impairment in Parkinson's disease. *Frontiers in Human Neuroscience, 17*, 1102024. https://doi.org/10.3389/fnhum.2023.1102024
3. Bamberger, R., Kuhles, G., Lotter, L. D., Dukart, J., Konrad, K., Günther, T., Siniatchkin, M., Fuchs, M., & von Polier, G. (2026). Voice features in paediatric ADHD: Baseline group differences and longitudinal change under methylphenidate. *European Child & Adolescent Psychiatry*. Advance online publication. https://doi.org/10.1007/s00787-026-03213-1
4. Bang, J.-U., Han, S.-H., & Kang, B.-O. (2024). Alzheimer's disease recognition from spontaneous speech using large language models. *ETRI Journal, 46*(1), 96–105. https://doi.org/10.4218/etrij.2023-0356
5. Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory–II*. Psychological Corporation.
6. Berisha, V., Krantsevich, C., Stegmann, G., Hahn, S., & Liss, J. (2022). Are reported accuracies in the clinical speech machine learning literature overoptimistic? In *Proceedings of Interspeech 2022* (pp. 2453–2457). https://doi.org/10.21437/Interspeech.2022-691
7. Berisha, V., & Liss, J. M. (2024). Responsible development of clinical speech AI: Bridging the gap between clinical research and technology. *npj Digital Medicine, 7*, 208. https://doi.org/10.1038/s41746-024-01199-1
8. Camilleri, J. A., Volkening, J., Heim, S., Mochalski, L. N., Neufeld, H., Schlothauer, N., Eickhoff, S. B., Kuhles, G. & Weis, S. (2024). SpEx: A German-language dataset of speech and executive function performance. *Scientific reports, 14*(1), 9431. https://doi.org/10.1038/s41598-024-58617-3
9. Cao, F., Vogel, A. P., Gharahkhani, P., & Rentería, M. E. (2025). Speech and language biomarkers for Parkinson's disease prediction, early diagnosis and

progression. *npj Parkinson's Disease, 11*(1), 57. https://doi.org/10.1038/s41531-025-00913-4

10. Cummins, N., Ramanarayanan, V., Low, D., Catania, F., Ng, S.-I., Botelho, C., Dineley, J., Elbeji, A., Liss, J. M., Pérez-Toro, P. A., Berisha, V., & Quatieri, T. (2026). *Towards clinical adoption of voice and speech as measures of health: The need for harmonization* (arXiv:2609.28894). arXiv. https://doi.org/10.48550/arXiv.2609.28894

11. Cummins, N., White, L. L., Rahman, Z., Lucas, C., Pan, T., Carr, E., Matcham, F., Downs, J., Dobson, R., Quatieri, T. F., & Dineley, J. (2025). Collection and analysis of repeated speech samples: Methodological framework and example protocol. *JMIR Research Protocols, 14*, e69431. https://doi.org/10.2196/69431

12. Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12*(2), 246–269. https://doi.org/10.1044/jshr.1202.246

13. Dineley, J., Carr, E., White, L. L., Lucas, C., Rahman, Z., Pan, T., Matcham, F., Downs, J., Dobson, R. J., Quatieri, T. F., & Cummins, N. (2024). Variability of speech timing features across repeated recordings: A comparison of open-source extraction techniques. In *Proceedings of Interspeech 2024* (pp. 2015–2019). https://doi.org/10.21437/Interspeech.2024-1074

14. Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., & Truong, K. P. (2016). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing, 7*(2), 190–202. https://doi.org/10.1109/TAFFC.2015.2457417

15. Eyben, F., Wöllmer, M., & Schuller, B. (2010). openSMILE: The Munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM International Conference on Multimedia* (pp. 1459–1462). ACM. https://doi.org/10.1145/1873951.1874246

16. Fagherazzi, G., Fischer, A., Ismael, M., & Despotovic, V. (2021). Voice for health: The use of vocal biomarkers from research to clinical practice. *Digital Biomarkers, 5*(1), 78–88. https://doi.org/10.1159/000515346

17. Fengler, S., Liepelt-Scarfone, I., Brockmann, K., Schäffer, E., Berg, D., & Kalbe, E. (2017). Cognitive changes in prodromal Parkinson's disease: A review. *Movement Disorders, 32*(12), 1655–1666. https://doi.org/10.1002/mds.27135

18. Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*(2), 210–233. https://doi.org/10.1016/0749-596X(91)90004-4

19. Frota, S., Cruz, M., Cardoso, R., Guimarães, I., Ferreira, J. J., Pinto, S., & Vigário, M. (2021). (Dys)prosody in Parkinson's disease: Effects of medication and disease duration on intonation and prosodic phrasing. *Brain Sciences, 11*(8), 1100. https://doi.org/10.3390/brainsci11081100

20. Fumel, J., Bahuaud, D., Weed, E., Fusaroli, R., & Basirat, A. (2024). A systematic review and Bayesian meta-analysis of acoustic measures of prosody in Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 67*(8), 2548–2564. https://doi.org/10.1044/2024_JSLHR-23-00588

21. García, A. M., de Leon, J., Tee, B. L., Blasi, D. E., & Gorno-Tempini, M. L. (2023). Speech and language markers of neurodegeneration: A call for global equity. *Brain, 146*(12), 4870–4879. https://doi.org/10.1093/brain/awad253

22. Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martínez-Martín, P., Poewe, W., Sampaio, C., Stern, M. B., Dodel, R., Dubois, B., Holloway, R. G., Jankovic, J., Kulisevsky, J., Lang, A. E., Lees, A. J., Leurgans, S. E., LeWitt, P. A., Nyenhuis, D. L., … LaPelle, N. R. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders, 23*(15), 2129–2170. https://doi.org/10.1002/mds.22340

23. Haider, F., de la Fuente, S., & Luz, S. (2020). An assessment of paralinguistic acoustic features for detection of Alzheimer's dementia in spontaneous speech. *IEEE Journal of Selected Topics in Signal Processing, 14*(2), 272–281. https://doi.org/10.1109/JSTSP.2019.2955022

24. Hamdan, S., More, S., Sasse, L., Komeyer, V., Patil, K. R., Raimondo, F., & Alzheimer's Disease Neuroimaging Initiative. (2024). Julearn: An easy-to-use library for leakage-free evaluation and inspection of ML models. *GigaByte, 2024*, gigabyte113. https://doi.org/10.46471/gigabyte.113

25. Harada, C. N., Natelson Love, M. C., & Triebel, K. L. (2013). Normal cognitive aging. *Clinics in Geriatric Medicine, 29*(4), 737–752. https://doi.org/10.1016/j.cger.2013.07.002

26. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

27. Hecker, P., Steckhan, N., Eyben, F., Schuller, B. W., & Arnrich, B. (2022). Voice analysis for neurological disorder recognition–A systematic review and perspective on emerging trends. *Frontiers in Digital Health, 4*, 842301. https://doi.org/10.3389/fdgth.2022.842301

28. Hely, M. A., Reid, W. G. J., Adena, M. A., Halliday, G. M., & Morris, J. G. L. (2008). The Sydney multicenter study of Parkinson's disease: The inevitability of dementia at 20 years. *Movement Disorders, 23*(6), 837–844. https://doi.org/10.1002/mds.21956

29. Herbig, E., Mücke, D., Barbe, M. T., & Thies, T. (2025). Executive dysfunctions impair and levodopa improves articulatory timing in Parkinson's disease. *Frontiers in Human Neuroscience, 19*, 1580376. https://doi.org/10.3389/fnhum.2025.1580376

30. Hernandez, H., Pedraza, L. K., Santamaria-Garcia, H., Moguilner, S., Legaz, A., Prado, P., Cuadros, J., Amoruso, L., Gonzalez, L., Dellavale, D., Espinoza-Puelles, J. P., Palma Espinosa, J., Jarne, C., Mattiussi, F., Caccia, M., Sosa Welford, A., Pelella, N., Inchauspe, J., Ferrante, F. J., … Ibanez, A. (2026). Speech clocks decode dementia phenotypes, social exposome, and biological aging. *Science Advances, 12*(40), eaef9864. https://doi.org/10.1126/sciadv.aef9864

31. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

32. Kalbe, E., Calabrese, P., Kohn, N., Hilker, R., Riedel, O., Wittchen, H.-U., Dodel, R., Otto, J., Ebersbach, G., & Kessler, J. (2008). Screening for cognitive deficits in Parkinson's disease with the Parkinson neuropsychometric dementia assessment (PANDA) instrument. *Parkinsonism & Related Disorders, 14*(2), 93–101. https://doi.org/10.1016/j.parkreldis.2007.06.008

33. König, A., Tröger, J., Mallick, E., Linz, N., Ritchie, C., Gregory, S., Hunter, M., Johnson, K., Sánchez-Benavides, G., Grau-Rivera, O., Rădoi, A., Porta-Mas, C., Köhler, S., Teipel, S., Hansson, O., Tideman, P., Wuestefeld, A., & Palmqvist, S. (2026). Speech-based digital cognitive assessment for clinical trials: Detecting cognitive impairment stages and AD biomarker relations across European cohorts. *Alzheimer's & Dementia, 22*(5), e71462. https://doi.org/10.1002/alz.71462

34. Kovac, D., Novakova, L., Mekyska, J., Novotny, K., Brabenec, L., Klobusiakova, P., & Rektorova, I. (2025). Digital speech biomarkers for assessing cognitive decline across neurodegenerative conditions. *Computers in Biology and Medicine, 198*, 111251. https://doi.org/10.1016/j.compbiomed.2025.111251

35. Martens, H., Van Nuffelen, G., Dekens, T., Hernández-Díaz Huici, M., Kairuz Hernández-Díaz, H. A., De Letter, M., & De Bodt, M. (2015). The effect of intensive speech rate and intonation therapy on intelligibility in Parkinson's disease. *Journal of Communication Disorders, 58*, 91–105. https://doi.org/10.1016/j.jcomdis.2015.10.004

36. Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

37. Ngo, Q. C., Motin, M. A., Pah, N. D., Drotár, P., Kempster, P., & Kumar, D. K. (2022). Computerized analysis of speech and voice for Parkinson's disease: A systematic review. *Computer Methods and Programs in Biomedicine, 226*, 107133. https://doi.org/10.1016/j.cmpb.2022.107133

38. North, B. V., Curtis, D., & Sham, P. C. (2002). A note on the calculation of empirical P values from Monte Carlo procedures. *American Journal of Human Genetics, 71*(2), 439–441. https://doi.org/10.1086/341527

39. O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

40. Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology, 27*(3), 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009

41. Pérez-Toro, P. A., Arias-Vergara, T., Klumpp, P., Vásquez-Correa, J. C., Schuster, M., Nöth, E., & Orozco-Arroyave, J. R. (2022). Depression assessment in people with Parkinson's disease: The combination of acoustic features and natural language processing. *Speech Communication, 145*, 10–20. https://doi.org/10.1016/j.specom.2022.09.001

42. Pérez-Toro, P. A., Ferrante, F. J., Pérez, G., Tee, B. L., de Leon, J., Nöth, E., Schuster, M., Maier, A., Slachevsky, A., Gorno-Tempini, M. L., Ibáñez, A., Orozco-Arroyave, J. R., & García, A. (2025). Automated speech markers of Alzheimer dementia: Test of cross-linguistic generalizability. *Journal of Medical Internet Research, 27*, e74200. https://doi.org/10.2196/74200

43. Pizzimenti, M., Kalia, A., Toghranegar, J. A., Ebraheem, M., Cummins, N., Ghosh, S. S., Anibal, J. T., Au, R., Azarang, A., Bahr, R. H., Barvaux, S., Bedrick, S. D., Botha, H., Coleman, O. C., Elbeji, A., Kourtis, L. C., Rameau, A., Sara, J. D. S., Watts, S. W., … Fagherazzi, G. (2026). Consensus-based definitions for vocal biomarkers: The international VOCAL initiative. *Digital Biomarkers, 10*(1), 178–190. https://doi.org/10.1159/000553327

44. Polychronis, S., Niccolini, F., Pagano, G., Yousaf, T., & Politis, M. (2019). Speech difficulties in early de novo patients with Parkinson's disease. *Parkinsonism & Related Disorders, 64*, 256–261. https://doi.org/10.1016/j.parkreldis.2019.04.026

45. Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A. E., Halliday, G., Goetz, C. G., Gasser, T., Dubois, B., Chan, P., Bloem, B. R., Adler, C. H., & Deuschl, G. (2015). MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders, 30*(12), 1591–1601. https://doi.org/10.1002/mds.26424

46. Reijnders, J. S., Ehrt, U., Weber, W. E., Aarsland, D., & Leentjens, A. F. (2008). A systematic review of prevalence studies of depression in Parkinson's disease. *Movement disorders, 23*(2), 183-189. https://doi.org/10.1002/mds.21803

47. Rektorova, I., Mekyska, J., Janousova, E., Kostalova, M., Eliasova, I., Mrackova, M., Berankova, D., Necasova, T., Smekal, Z., & Marecek, R. (2016). Speech prosody impairment predicts cognitive decline in Parkinson's disease. *Parkinsonism & Related Disorders, 29*, 90–95. https://doi.org/10.1016/j.parkreldis.2016.05.018

48. Rusz, J., Čmejla, R., Růžičková, H., & Růžička, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *The Journal of the Acoustical Society of America, 129*(1), 350–367. https://doi.org/10.1121/1.3514381

49. Rusz, J., Cmejla, R., Tykalova, T., Ruzickova, H., Klempir, J., Majerova, V., Picmausova, J., Roth, J., & Ruzicka, E. (2013). Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task. *The Journal of the Acoustical Society of America, 134*(3), 2171–2181. https://doi.org/10.1121/1.4816541

50. Rusz, J., Tykalová, T., Novotný, M., Růžička, E., & Dušek, P. (2021a). Distinct patterns of speech disorder in early-onset and late-onset de-novo Parkinson's disease. *npj Parkinson's Disease, 7*, 98. https://doi.org/10.1038/s41531-021-00243-1

51. Rusz, J., Tykalova, T., Ramig, L. O., & Tripoliti, E. (2021b). Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders. *Movement Disorders, 36*(4), 803–814. https://doi.org/10.1002/mds.28465

52. Skodda, S., & Schlegel, U. (2008). Speech rate and rhythm in Parkinson's disease. *Movement Disorders, 23*(7), 985–992. https://doi.org/10.1002/mds.21996

53. Skodda, S., Visser, W., & Schlegel, U. (2011a). Gender-related patterns of dysprosody in Parkinson disease and correlation between speech variables and motor symptoms. *Journal of Voice, 25*(1), 76–82. https://doi.org/10.1016/j.jvoice.2009.07.005

54. Skodda, S., Visser, W., & Schlegel, U. (2011b). Vowel articulation in Parkinson's disease. *Journal of Voice, 25*(4), 467–472. https://doi.org/10.1016/j.jvoice.2010.01.009

55. Smith, K. M., & Caplan, D. N. (2018). Communication impairment in Parkinson's disease: Impact of motor and cognitive symptoms on speech and language. *Brain and Language, 185*, 38–46. https://doi.org/10.1016/j.bandl.2018.08.002

56. Smith, K. M., Williamson, J. R., & Quatieri, T. F. (2025). Speech acoustic markers can detect mild cognitive impairment in Parkinson's disease. *IEEE Journal of Selected Topics in Signal Processing, 19*(5), 731–740. https://doi.org/10.1109/JSTSP.2025.3620716

57. Stegmann, G., Hahn, S., Bhandari, S., Kawabata, K., Shefner, J., Duncan, C. J., Liss, J., Berisha, V., & Mueller, K. (2022). Automated semantic relevance as an indicator of cognitive decline: Out-of-sample validation on a large-scale longitudinal dataset. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 14*(1), e12294. https://doi.org/10.1002/dad2.12294

58. Stegmann, G. M., Hahn, S., Liss, J., Shefner, J., Rutkove, S. B., Kawabata, K., Bhandari, S., Shelton, K., Duncan, C. J., & Berisha, V. (2020). Repeatability of commonly used speech and language features for clinical applications. *Digital Biomarkers, 4*(3), 109–122. https://doi.org/10.1159/000511671

59. Sugimura, Y., Baba, T., Totsune, T., Oizumi, H., Hasegawa, T., Suzuki, K., & Takeda, A. (2026). Temporal patterns of motor and nonmotor symptom emergence in

Parkinson's disease: A cluster analysis. *Clinical Parkinsonism & Related Disorders, 14*, 100423. https://doi.org/10.1016/j.prdoa.2026.100423

60. Thies, T., Mücke, D., Lowit, A., Kalbe, E., Steffen, J., & Barbe, M. T. (2020). Prominence marking in parkinsonian speech and its correlation with motor performance and cognitive abilities. *Neuropsychologia*, *137*, 107306.

61. Torre, P., III, & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders, 42*(5), 324–333. https://doi.org/10.1016/j.jcomdis.2009.03.001

62. Tóth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatlóczki, G., Bánréti, Z., Pákáski, M., & Kálmán, J. (2018). A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Current Alzheimer Research, 15*(2), 130–138. https://doi.org/10.2174/1567205014666171121114930

63. Wei, H. T., Kulzhabayeva, D., Erceg, L., Rose, M. K., Spencer, K. A., Robin, J., Bialystok, E., & Meltzer, J. A. (2025). Natural speech analysis can reveal individual differences in executive function across the adult lifespan. *Journal of Speech, Language, and Hearing Research, 68*(12), 5708–5726. https://doi.org/10.1044/2025_JSLHR-24-00268

64. Wiesman, A. I., Donhauser, P. W., Degroot, C., Diab, S., Kousaie, S., Fon, E. A., Klein, D., Baillet, S., PREVENT-AD Research Group, & Quebec Parkinson Network. (2023). Aberrant neurophysiological signaling associated with speech impairments in Parkinson's disease. *npj Parkinson's Disease, 9*(1), 61. https://doi.org/10.1038/s41531-023-00495-z

65. Young, C. B., Smith, V., Karjadi, C., Grogan, S. M., Ang, T. F. A., Insel, P. S., Henderson, V. W., Sumner, M., Poston, K. L., Au, R., & Mormino, E. C. (2024). Speech patterns during memory recall relates to early tau burden across adulthood. *Alzheimer's & Dementia, 20*(4), 2552–2563. https://doi.org/10.1002/alz.13731

66. Yuan, J., Cai, X., Bian, Y., Ye, Z., & Church, K. (2021). Pauses for detection of Alzheimer's disease. *Frontiers in Computer Science, 2*, 624488. https://doi.org/10.3389/fcomp.2020.624488

67. Ziegler, W., Schölderle, T., Staiger, A., & Vogel, M. (2018). BoDyS: Bogenhausener Dysarthrieskalen [Bogenhausen Dysarthria Scales]. Hogrefe.

Acknowledgements

We thank all participants for making their data available for research.

Author Contributions

Conceptualisation: G.K., T.T.,, M.T.B., S.B.E., S.W. Methodology: G.K., S.B.E., S.W. Resources: T.T., M.T.B., S.B.E., J.A.C., S.W. Data curation: G.K., T.T., Formal analysis and visualisation: G.K. Writing – original draft: G.K. Writing – review & editing: all authors. All authors read and approved the final manuscript.

Ethics declarations

The retrospective analysis of the PD sample was approved by the local ethics committee of the University Hospital Cologne (protocol codes 23-1461-retro; date of approval 6 December 2023 and 23-1355; date of approval 2 February 2024). The reference dataset was approved by the ethics committee of Heinrich Heine University Düsseldorf (study number: 6055R; date of approval: 14 September 2017). All participants provided written informed consent for their data to be analysed for research purposes. All procedures were performed in accordance with the Declaration of Helsinki and the relevant guidelines and regulations. Further details regarding the approval of the reference dataset can be found in the corresponding publication (Camilleri et al., 2024).

Declaration of conflicting interest

None of the authors have conflicts of interest to disclose in relationship with the current work. However, T.T. is an employee of ki:elements GmbH.

Data availability

Data of the PD sample is available upon request to Michael T. Barbe. Data of the reference sample is publicly available upon request. Researchers who wish to acquire access to the data are kindly asked to contact Julia A. Camilleri at spexdata@fz-juelich.de, as described in the related publication Camilleri et al., 2024.

Code availability

The code used in this study is publicly available at Github and archived with the DOI: xxx

Supplementary Files

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

- SupplementaryMaterialAssociationsofSpontaneousSpeechprosodywithCognitiveandMotorfunctioninParkinsonsdisease.docx