RESEARCH ARTICLE

Smartphone Voice Calls Provide Early Biomarkers of Parkinsonism in Rapid Eye Movement Sleep Behavior Disorder

Vojtěch Illner, MSc,[1] Michal Novotný, PhD,[1] Tomáš Kouba, MSc,[1] Tereza Tykalová, PhD,[1] Michal Šimek, MSc,[1] Pavel Sovka, PhD,[1] Jan Švihlík, PhD,[1,2] Evžen Růžička, DSc,[3] Karel Šonka, DSc,[3] Petr Dušek, PhD,[3] and Jan Rusz, PhD[1,3,4*]

[1]*Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic*
[2]*Department of Computing and Control Engineering, Faculty of Chemical Engineering, University of Chemistry and Technology, Prague, Czech Republic*
[3]*Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University, Prague, Czech Republic*
[4]*Department of Neurology and ARTORG Center, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland*

ABSTRACT: Background: Speech dysfunction represents one of the initial motor manifestations to develop in Parkinson's disease (PD) and is measurable through smartphone.
Objective: The aim was to develop a fully automated and noise-resistant smartphone-based system that can unobtrusively screen for prodromal parkinsonian speech disorder in subjects with isolated rapid eye movement sleep behavior disorder (iRBD) in a real-world scenario.
Methods: This cross-sectional study assessed regular, everyday voice call data from individuals with iRBD compared to early PD patients and healthy controls via a developed smartphone application. The participants also performed an active, regular reading of a short passage on their smartphone. Smartphone data were continuously collected for up to 3 months after the standard in-person assessments at the clinic.
Results: A total of 3525 calls that led to 5990 minutes of preprocessed speech were extracted from 72 participants, comprising 21 iRBD patients, 26 PD patients, and 25 controls. With a high area under the curve of 0.85 between iRBD patients and controls, the combination of passive and active smartphone data provided a comparable or even more sensitive evaluation than laboratory examination using a high-quality microphone. The most sensitive features to induce prodromal neurodegeneration in iRBD included imprecise vowel articulation during phone calls ($P = 0.03$) and monopitch in reading ($P = 0.05$). Eighteen minutes of speech corresponding to approximately nine calls was sufficient to obtain the best sensitivity for the screening.
Conclusion: We consider the developed tool widely applicable to deep longitudinal digital phenotyping data with future applications in neuroprotective trials, deep brain stimulation optimization, neuropsychiatry, speech therapy, population screening, and beyond. © 2024 The Author(s). *Movement Disorders* published by Wiley Periodicals LLC on behalf of International Parkinson and Movement Disorder Society.

Key Words: prodromal synucleinopathy biomarker; Parkinson's disease; speech; wearables; machine learning

Early recognition of Parkinson's disease (PD) has crucial implications for the future development of neuroprotective therapy, as prodromal stages of the disease offer the best opportunity to intervene.[1-3] Therefore, establishing a suitable biomarker effective in prodromal stages would be a game-changing milestone that would

*__Correspondence to:__ Dr. Jan Rusz, Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, 160 00, Prague 6, Czech Republic; E-mail: ruszjan@fel.cvut.cz

Relevant conflicts of interest/financial disclosures: All authors report no conflict of interest concerning the research related to the manuscript.

Funding agencies: This study was supported by the Czech Ministry of Health (grant numbers NU20-08-00445 and MH CZ–DRO-VFN64165), the National Institute for Neurological Research (Programme EXCELES, ID project number LX22NPO5107)—funded by the European Union—Next Generation EU, and the Czech Technical University in Prague (grant number SGS23/170/OHK3/3T/13).

Received: 4 April 2024; **Revised:** 3 June 2024; **Accepted:** 21 June 2024

Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.29921

impact the diagnosis and future treatments of PD.[4] Isolated rapid eye movement sleep behavior disorder (iRBD) is now considered an essential prodromal symptom of synucleinopathies; most patients develop an overt neurodegenerative disease, particularly PD or dementia with Lewy bodies, within a decade.[5-8] Such a long prodromal window provides a unique opportunity to study disease development and design suitable biomarkers.

With the advent of digital health, there is the potential to remotely and noninvasively detect and track early signs of PD using tools such as smartphones.[9-13] However, many existing tests, such as finger tapping or walking a predefined distance, require an active, and instructed involvement,[14] while an ideal digital biomarker should be passively measured without additional effort from the subject or investigator. In this context, speech analysis offers intriguing potential advantages as a large part of the population communicates through smartphones daily. Thus, extracting speech patterns from smartphone calls in real-world settings presents a unique opportunity to provide a passive biomarker, allowing for continuous monitoring of the effectiveness of experimental treatments in a natural environment and facilitating large-scale screening.

Because speech represents the most complex quantitative marker of motor function that is highly sensitive to damage to neural structures,[15] it is unsurprising that speech dysfunction has been found to be one of the first signs to develop in PD.[16] Specifically, dysprosody and imprecise vowel articulation have been detected in iRBD subjects with impaired olfactory function but still largely functional nigrostriatal dopaminergic transmission,[17,18] that is, in Braak stage 2 before the substantia nigra is affected by synucleinopathy.[19] Unfortunately, these findings are based on actively performed speech recordings obtained using a professional condenser microphone in laboratory settings, which considerably limits the broader applicability of speech assessment.[20] Several challenges must be overcome, including typically the low quality of smartphone microphones, background noise in everyday environments, and unstable direction and distance of the smartphone from the lips due to various holding positions, to allow passive smartphone speech monitoring.[21,22]

We developed a fully automated and noise-resistant smartphone-based system that can unobtrusively monitor speech in a real-world scenario. We aimed to (1) test the reliability of passively obtained acoustic speech features via everyday smartphone calls to detect prodromal parkinsonism in subjects with iRBD, (2) compare the sensitivity of passive voice monitoring with active speech tasks performed using smartphones at home and professional microphones in laboratory settings, and (3) estimate the necessary sample duration to reach the optimal sensitivity for the detection of prodromal parkinsonism through smartphone-captured speech in a real-world setting.

Subjects and Methods

Study Design and Participants

From 2021 to 2023, we enrolled native Czech iRBD, early PD, and healthy control subjects at the Department of Neurology, First Faculty of Medicine, Charles University. Patients with iRBD were diagnosed according to the diagnostic criteria of the third edition of the *International Classification of Sleep Disorders*, including video polysomnography.[23] The exclusion criteria were as follows: (1) iRBD onset before age 50 years, (2) overt parkinsonism or dementia at baseline, and (3) iRBD onset within 12 months of introduction of antidepressant treatment. PD patients were diagnosed based on the Movement Disorder Society clinical diagnostic criteria for PD.[24] The exclusion criteria were as follows: (1) disease duration from diagnosis ≥5 years, (2) current involvement in any speech therapy, and (3) not on a stable dose of medication over the previous 4 weeks prior to the start of the study. Exclusion criteria for healthy controls were a history of parasomnias or other sleep disorders in adulthood, or the diagnosis of iRBD on video polysomnography. The exclusion criteria for all groups included (1) a history of communication disorders unrelated to parkinsonism (ie, problems in speech comprehension or expression) or other neurological disorders potentially affecting speech and (2) unwillingness to achieve at least 10 minutes of phone calls in a month.

The clinical evaluation of each subject included the following: (1) medical history, history of drug and substance intake, and current drug usage; (2) quantitative testing of motor and nonmotor symptoms of PD using the Movement Disorders Society-Unified Parkinson's Disease Rating Scale, Part III (MDS-UPDRS)[25]; (3) cognitive testing using the Montreal Cognitive Assessment (MoCA)[26]; (4) autonomic testing using the Scales for Outcomes in Parkinson's Disease–Autonomic Dysfunction[27]; and (5) University of Pennsylvania Smell Identification Test.[28] Perceptual speech severity was estimated using the speech item score from the MDS-UPDRS, Part III. Symptom duration was estimated based on the self-reported first occurrence of dream enactment in iRBD and motor symptoms in PD.

Each participant provided written informed consent. The study was approved by the Ethics Committee of the General University Hospital in Prague, Czech Republic, in accordance with the ethical standards established in the 1964 Declaration of Helsinki.

Smartphone Speech Examination

All subjects received HONOR 9X Lite (Shenzen Zhixin New Information Technology, Shenzhen, China) phone, operating on the Android version 9 system. The phone was chosen as a mainstream product (ie, commercially available and a relatively inexpensive midrange smartphone) among the smartphones available in the Czech Republic in 2021. The recordings were sampled at 44.1 kHz with 16-bit quantization.

The smartphone was equipped with preinstalled application,[29] which ran in the background and recorded the subject's voice during calls, removing the content from the distant speaker using adaptive filtering (Fig. 1A). After each incoming or outgoing call, the user was prompted by a screen containing an option to delete the call or send it to confidential analysis. On agreement, the audio recording was kept on the device for 24 hours to allow participants to replay and eventually delete it. After 24 hours, the pseudonymized recording was sent to a secure server and validated by a speaker recognition framework. Comprehensive technical details of the application were previously described in the protocol.[29] In addition, the application contained an active part. Subjects were prompted to read a passage twice, selected randomly from six samples of ~80 words, displayed on the application screen every 14 days (mean duration: 35.1, standard deviation [SD]: 5.5 seconds). All data were collected from the smartphone during a period of up to 3 months after the in-person visit at the clinic. Acquisition and secure data transfer were carried out in accordance with the directive on the legislation on personal data protection of the European Union.

Laboratory Speech Examination

Speech recordings were performed in a quiet room with a low ambient noise level using a high-quality head-mounted condenser microphone (Beyerdynamic Opus 55, Heilbronn, Germany) placed ~5 cm from the subject's mouth. Speech signals were sampled at 48 kHz with 16-bit quantization. Each subject was recorded during a single session accompanied by a speech specialist who guided the standardized protocol. Participants were instructed to deliver a monologue about an arbitrary topic of at least 90 seconds (mean duration: 123.6, SD: 19.3 seconds) and perform a reading passage task twice of a standardized text of 80 words (mean duration: 35.7, SD: 5.1 seconds).

Smartphone Call Preprocessing

The incoming calls contained nonspeech periods with no relevant information due to the dialogue nature of a conversation on the phone. Therefore, the recordings were stripped of any nonspeech segments longer than 0.7 seconds. The threshold was set to preserve natural pauses as a significant aspect of speech production. Subsequently, to normalize the calls in terms of duration, the recordings were partitioned into time frames of equal length, each treated as an individual recording. The frame length was chosen as 20, 30, 45, and 60 seconds to evaluate the impact of different durations. If there was a remainder, it was considered only if longer than 50% of the corresponding segment length (eg, if a 20-second window was selected for a 32-second call, both 20- and 12-second segments were analyzed).

Acoustic Speech Features

We selected seven representative acoustic speech features (Fig. 1B) following three main criteria: (1) representing a unique aspect of speech (the features were found to be only weakly correlated [Pearson's $|r| < 0.48$]), aligning with the perceptual description of the primary patterns of hypokinetic dysarthria[30]; (2) enabling automated analysis of connected speech; and (3) proven sensitivity in iRBD or early PD in previous studies.[18,31] We limited the number of acoustic parameters included in the experiment to reduce the probability of a type I error and to reduce potential overfitting for the regression analysis.

Monopitch was assessed by an SD of pitch contour,[21] imprecise vowel articulation by a formant ratio index,[32] voice quality by cepstral peak prominence,[22] articulatory decay by an SD of mel-frequency cepstral coefficients (global MFCC),[33] monoloudness by an SD of intensity contour after removal of pauses,[34] prolonged pauses by median duration of pause intervals,[35] and articulation rate through net speech rate acquired via automatic speech recognizer followed by hyphenation.[36,37] All analyses were performed in MATLAB (MathWorks, Natick, MA) and Python.

Speech Sample Duration Estimation

A unified, sufficient speech sample from each participant as well as the most optimal call frame duration (eg, whether to analyze 10 or 30 minutes of cumulative calls per participant, in 20- or 30-second frames) was determined. A binary logistic regression followed by a leave-one-out cross-validation using a combination of all acoustic features was utilized to determine the classification accuracy. The speech sample was then chosen based on group classification accuracy, at the point when the average accuracy across call frames reached 95% of its maximum value in the cumulative analyzed interval. The largest sample duration across the three classifications (controls vs. iRBD, controls vs. PD, and iRBD vs. PD) was chosen for the statistical analysis. The effect of call frame length was assessed based on accuracy of the selected speech sample. In active speech assessment, the number of reading tasks required for analysis was determined analogously to sample duration determination via calls.

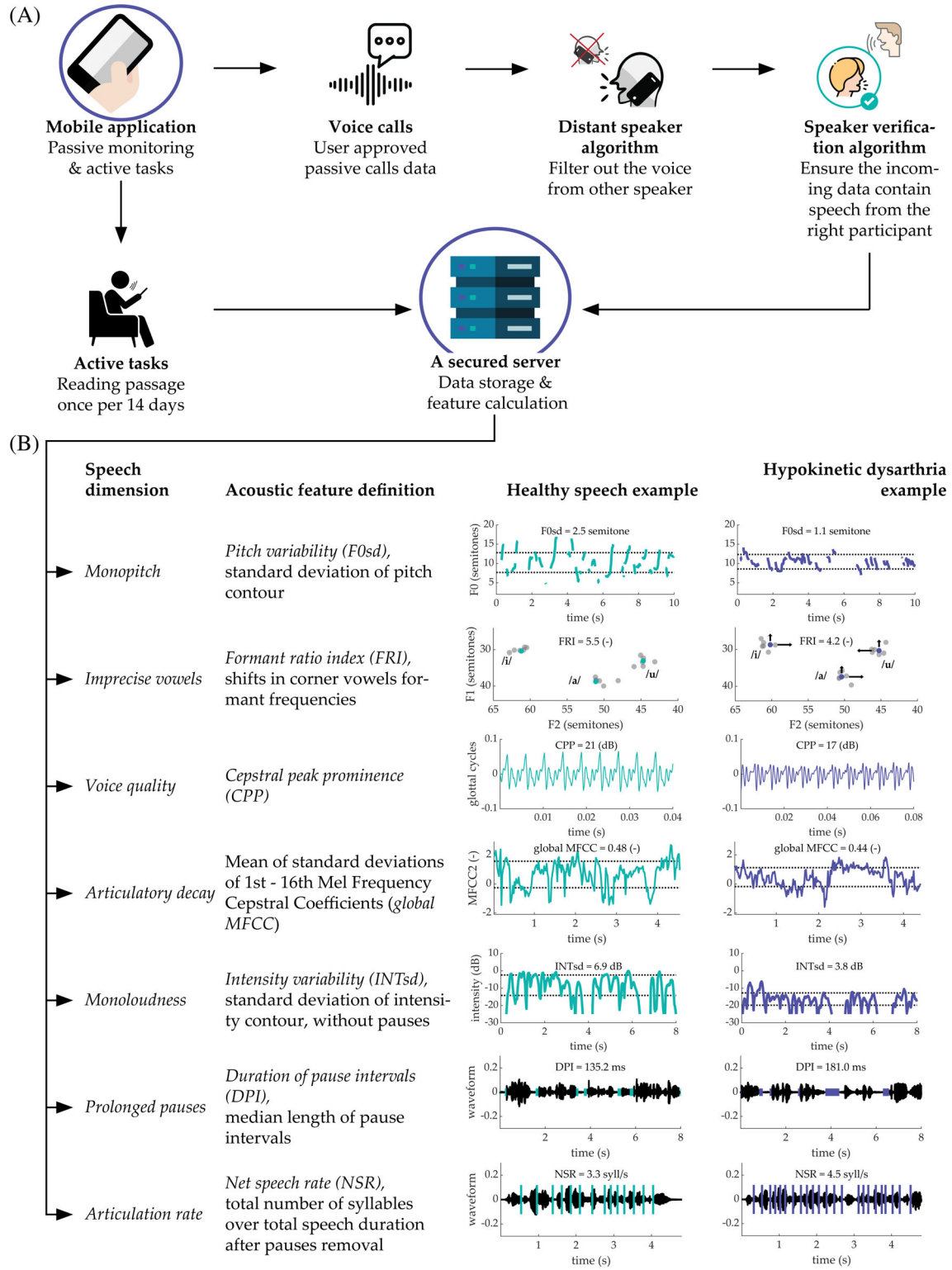

FIG. 1. The principal speech analysis scheme. (**A**) Illustrative diagram of the smartphone data acquisition system. (**B**) Illustrative table of speech dimensions described in the study, their definition, and example of healthy and dysarthric speakers. [Color figure can be viewed at wileyonlinelibrary.com]

Statistical Analysis

An ad hoc power analysis for a given large effect size (Cohen's d of 0.8), with the type I error probability (α) set at 0.05 and a power of 80%, based on a three-group analysis of variance with one covariate (group), determined a minimum sample size of 66 subjects (ie, 22 per group). A one-way analysis of variance with Bonferroni post hoc test was applied to analyze group

differences. The relationships between features were evaluated using Spearman's correlation coefficient. To assess the sensitivity between groups, a binary logistic regression model followed by leave-one-out cross-validation was utilized. The features used were determined based on an exhaustive search, providing the best outcome across spontaneous speech (calls and laboratory monologue) and reading tasks (smartphone and laboratory) and their combination, and we compared the receiver operating curve along with its area under the curve (AUC).

Results

Collected Data

Of 52 available iRBD subjects, 21 (40%) met the inclusion criteria and were willing to participate. The main reason for nonparticipation was that subjects (1) made only exceptional phone calls, (2) were unwilling to use smartphones, and (3) did not like the purpose of the project and/or the need to share personal voice calls. Additionally, we recruited 25 healthy controls and 26 early PD patients (Table 1).

During the 3 months of data collection, 3525 calls (mean: 49.0, SD: 61.1 per participant) were recorded and analyzed. Of these, 5990 minutes (mean: 83.2, SD: 119.7 per participant) of preprocessed speech was extracted for the analysis. On average, one call contained 2.26 minutes (SD: 1.96) of preprocessed speech useful for analysis. Considering active assessment, 950 (mean: 13.2, SD: 7.0 per participant) reading tasks were obtained.

Speech Sample Duration Estimation

For smartphone calls, the best accuracy reached the threshold (95% of maximum discriminative accuracy) at the cumulative call duration of 15 minutes for differentiating between PD and controls, 18 minutes between iRBD patients and controls, and 3 minutes between iRBD patients and PD (Fig. 2A). However, 8 minutes of sample duration provided stable classification between iRBD and controls, and the performance continuously increased with increasing sample duration. Regarding call frame durations, no specific option exhibited notable advantages in accuracy. Because the 20-second time frame provides enhanced flexibility given its brief duration, the analysis was carried out using 18 minutes of calls, preprocessed in 20-second frames.

For active reading tasks, the threshold was reached in two tasks (one trial) in differentiating between PD patients and controls, three tasks between iRBD patients and controls, and one task between iRBD and PD patients (Fig. 2B). As a result, an average of three reading tasks were considered for the analysis.

TABLE 1 *Clinical data of participants*

	Controls (n = 25)	iRBD (n = 21)	PD (n = 26)	*P*-value
Men	24 (96%)	20 (95%)	25 (96%)	0.99
Age (y)	67.1/7.3 (55–84)	68.3/8.6 (53–86)	58.5/8.6 (45–76)	<0.001[a,b]
Symptom duration (y)	–	10.3/6.7 (2–29)	5.5/2.1 (2–11)	–
MDS-UPDRS, Part III, total	6.5/2.7 (2–11)	9.8/2.5 (5–15)	25.9/9.8 (10–51)	<0.001[a,b]
MDS-UPDRS, Part III, speech item	0.3/0.5 (0–1)	0.4/0.5 (0–1)	1.0/0.3 (0–2)	<0.001[a,b]
MoCA	26.2/2.6 (22–30)	25.9/2.2 (21–30)	26.8/2.8 (18–30)	0.46
SCOPA-AUT	7.3/5.1 (1–24)	13.0/8.7 (3–39)	10.0/6.0 (1–24)	<0.05[c]
UPSIT	30.7/4.1 (21–39)	22.7/7.6 (13–36)	24.7/6.4 (12–35)	<0.001[a,b]
Antidepressant therapy	1 (4%)	2 (10%)	4 (15%)	0.4
Levodopa equivalent (mg/day)	0	0	632.9/311.7 (50–1440)	<0.001[a,b]
Clonazepam therapy	0 (0%)	10 (48%)	0 (0%)	<0.001[b,c]
RBD presence[d]	0 (0%)	21 (100%)	10 (38%)	<0.001[a–c]

Data are the mean/standard deviation (range) or the number (%).
[a]Significant difference between PD patients and controls.
[b]Significant difference between iRBD and PD patients.
[c]Significant difference between iRBD patients and controls.
[d]Presence of RBD was diagnosed by history and video polysomnography.
Abbreviations: iRBD, idiopathic rapid eye movement sleep behavior; PD, Parkinson's disease; MDS-UPDRS, Movement Disorder Society-Unified Parkinson's Disease Rating Scale; MoCA, Montreal Cognitive Assessment; SCOPA-AUT, Scales for Outcomes in Parkinson's Disease-Autonomic Dysfunction; UPSIT, University of Pennsylvania Smell Identification Test.

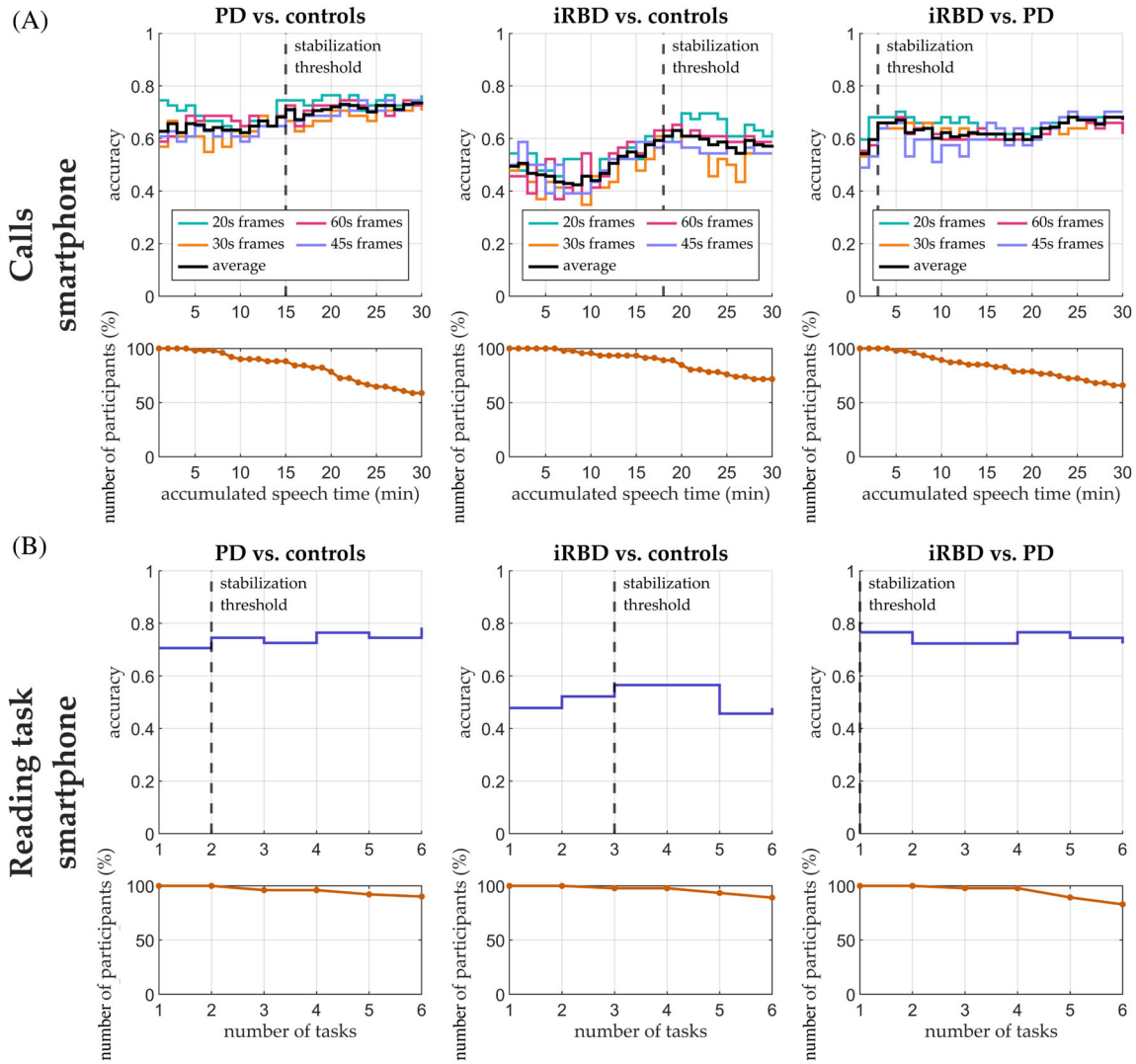

FIG. 2. Sample duration estimation. Accuracy of a binary logistic regression followed by a leave-one-out cross-validation using a combination of all the acoustic features for increasing sample duration of (**A**) speech from calls for different frame durations and their average value and (**B**) smartphone reading tasks. The dashed line (stabilization threshold) corresponds to the point when the average accuracy across call frames reached 95% of its maximum value in the cumulative analyzed interval. Below each plot is a percentage of participants able to reach such a sample duration. iRBD, isolated rapid eye movement sleep behavior disorder; PD, Parkinson's disease. [Color figure can be viewed at wileyonlinelibrary.com]

Speech Differences

Statistically significant differences between iRBD patients and controls were observed in monopitch during the laboratory reading task ($P = 0.049$) and imprecise vowels in calls ($P = 0.03$) (Fig. 3). Furthermore, compared to controls, PD patients exhibited impairment in monopitch during the laboratory reading task ($P < 0.001$), imprecise vowels in calls ($P = 0.01$), articulatory decay in laboratory monologue ($P < 0.001$), prolonged pauses in laboratory monologue ($P < 0.001$), and increased articulation rate in calls ($P = 0.01$) and both reading tasks ($P < 0.001$). Voice quality and monoloudness did not reach significance between the groups. No significant relationships were found between individual acoustic features and MDS-UPDRS, Part III, and MoCA scores in PD patients, iRBD patients, or controls. No relevant differences were detected between PD with RBD and PD without RBD (Fig. S1) as well as between iRBD treated with and iRBD not treated with clonazepam (Fig. S2).

Correlations among Data from Different Sources

Between calls and laboratory monologue, a significant correlation coefficient was achieved only for imprecise vowels ($r = 0.67$, $P < 0.001$) (Fig. 3). Between reading tasks, the significant correlations were generally more frequent, with a high correlation coefficient demonstrated for monopitch ($r = 0.70$, $P < 0.001$), voice quality ($r = 0.66$, $P < 0.001$), and articulation rate ($r = 0.70$, $P < 0.001$).

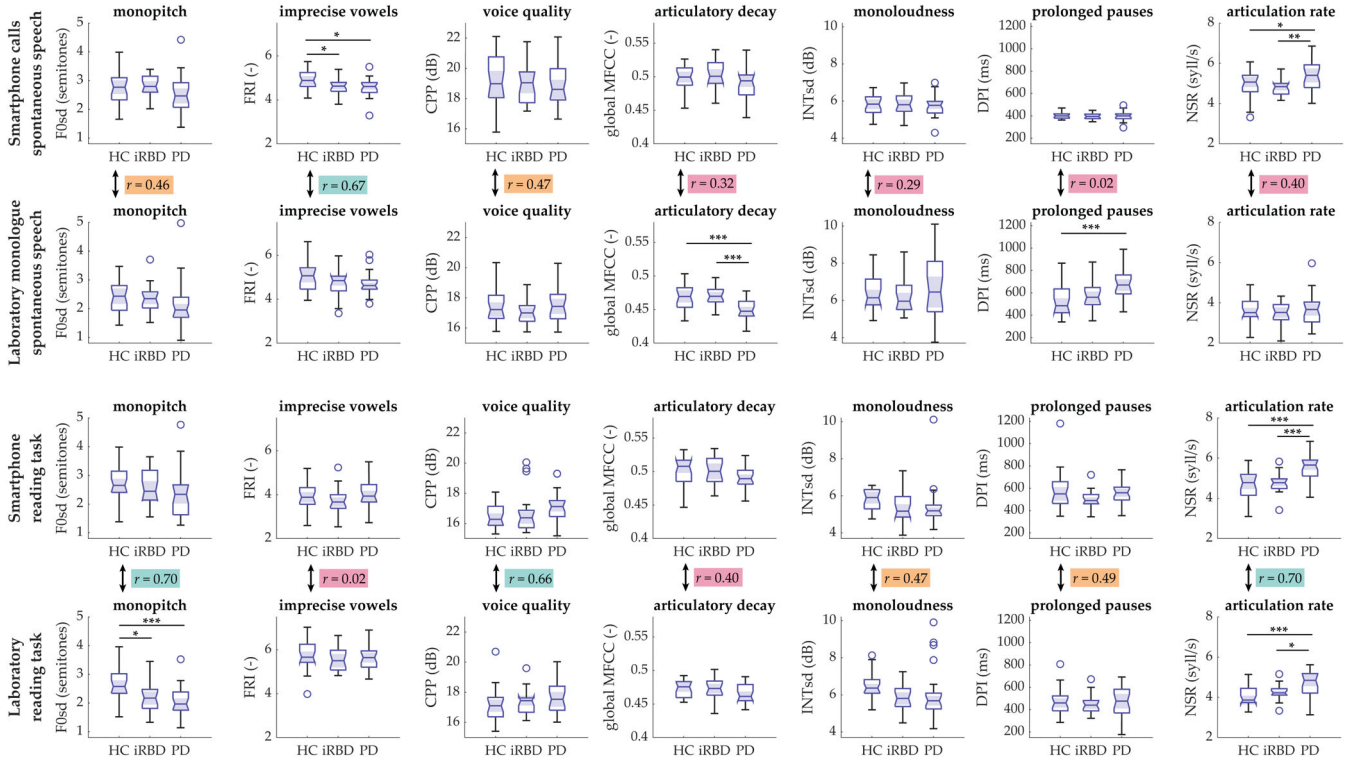

FIG. 3. Group differences between speech features in individual tasks and correlations of features between tasks. The calls measure is taken from 18 minutes of speech. The smartphone reading task is based on an average of three repetitions. Horizontal line represents median, the box lower and upper quartiles, the bars minimum and maximum values that are not outliers, and the circles outliers. ***, **, and * represent significant differences with $P < 0.001$, $P < 0.1$, and $P < 0.5$, respectively, after Bonferroni adjustment. CPP, cepstral peak prominence; DPI, duration of pause intervals; F0sd, standard deviation of pitch contour; FRI, formant ratio index; HC, healthy control; INTsd, standard deviation of intensity contour; iRBD, isolated rapid eye movement sleep behavior disorder; MFCC, mel-frequency cepstral coefficient; NSR, net speech rate; PD, Parkinson's disease; r, Spearman's correlation coefficient. [Color figure can be viewed at wileyonlinelibrary.com]

Sensitivity Analysis

Based on the exhaustive search, the optimal combination of features for spontaneous speech comprised monopitch, imprecise vowels, articulatory decay, prolonged pauses, and articulation rate, whereas that for reading tasks included monopitch, articulatory decay, monoloudness, and articulation rate. The highest AUC between iRBD patients and controls was 0.79 via calls compared to 0.66 via laboratory monologue (Fig. 4). Similar AUC values up to 0.87 were found between PD patients and controls for both smartphone and laboratory settings. For reading tasks, a superior AUC of up to 0.83 was obtained in laboratory settings compared to smartphones for distinguishing controls from both PD and iRBD patients. In general, the accuracy of prodromal speech disorders detection via smartphones improved to an AUC of up to 0.85 when combining both passive calls and active reading.

Discussion

This study is the first to evaluate speech characteristics collected in the wild in individuals with iRBD and early PD. It revealed that voice calls provide prodromal biomarkers of parkinsonism in iRBD patients, with sensitivity levels comparable to or even exceeding those of laboratory examination using high-quality equipment. The combination of passive and active smartphone data captured distinct yet complementary voice information, reaching a high AUC of 0.85 between iRBD and controls. Among the most prominent features of iRBD were monopitch in reading and imprecise vowel articulation in phone calls. These findings endorse the feasibility of employing a fully automated and noise-resistant smartphone-based system for passive speech monitoring in real-world scenarios for future clinical trials and detecting subjects at risk of later overt synucleinopathy development.

Speech Sample Duration

The effect of speech sample duration on biomarker performance has rarely been investigated. Although not systematically researched, a previous study suggested that 50 smartphone call sessions lasting between 15 and 75 seconds (corresponding to ~13–65 minutes) are sufficient to detect PD-related speech impairment.[38]

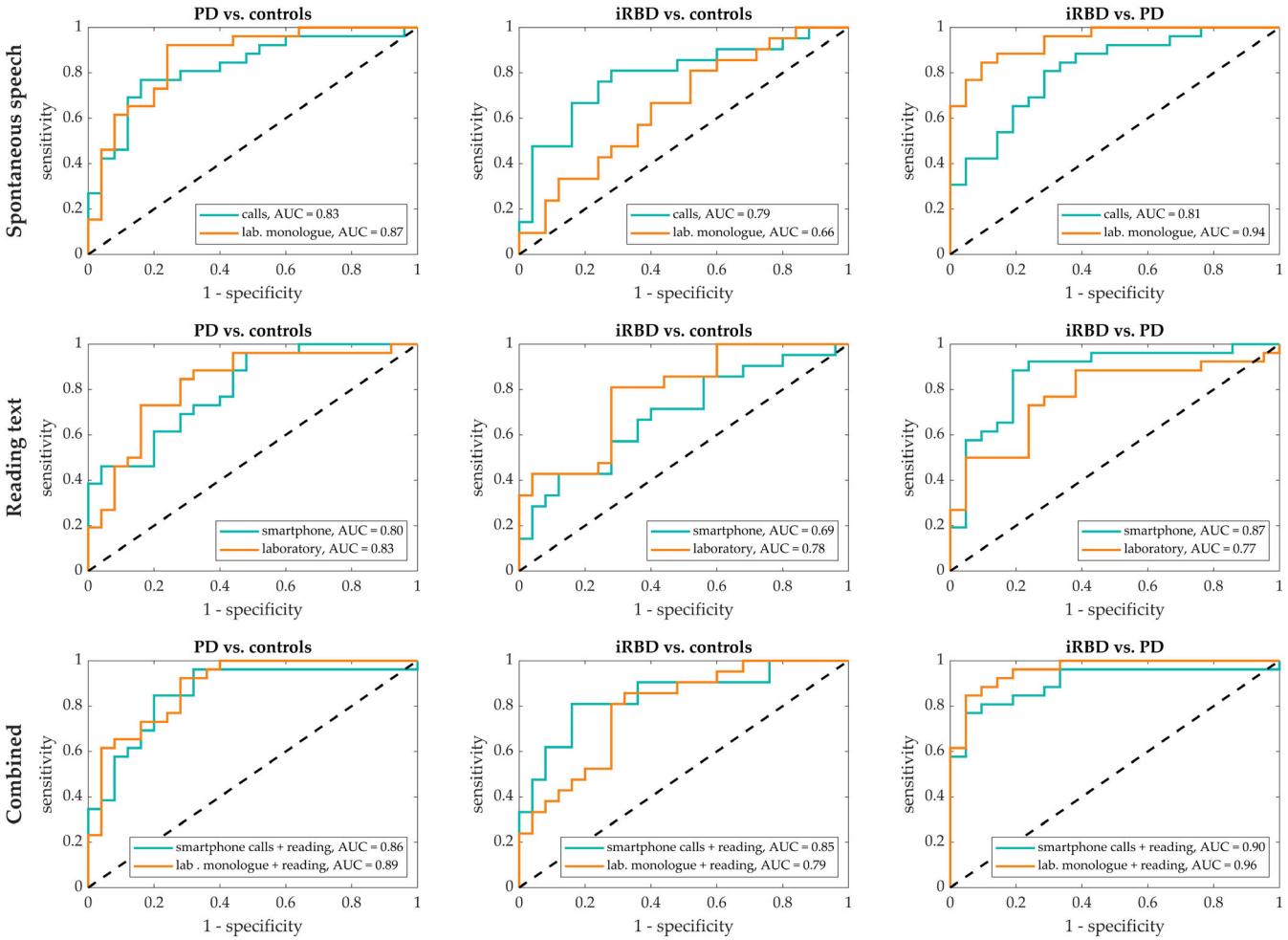

FIG. 4. Sensitivity analysis. Receiver operating characteristic curves of a binary logistic classification of optimal features based on an exhaustive search. The optimal features for spontaneous speech were monopitch, imprecise vowels, articulatory decay, prolonged pauses, and articulation rate, whereas those for reading task were monopitch, articulatory decay, monoloudness, and articulation rate. The calls measure is taken from 18 minutes of speech. The smartphone reading task is based on an average of three repetitions. AUC, area under curve; iRBD, isolated rapid eye movement sleep behavior disorder; PD, Parkinson's disease. [Color figure can be viewed at wileyonlinelibrary.com]

The duration is greater than that in the current study, where we found 18 minutes of speech (corresponding to approximately nine calls) sufficient to capture prodromal voice changes in the wild using smartphones. Interestingly, it takes a lesser duration to capture voice impairment in PD patients, suggesting that the higher the severity of dysarthria, the less data are needed. However, a stable but lower accuracy for distinguishing between iRBD patients and controls was achieved already for 8 minutes of calls.

Considering active smartphone data collection, three reading tasks are sufficient to fully capture prodromal voice characteristics in iRBD, demonstrating that guided tasks require a smaller sample size for effective analysis.[39] This is principally in agreement with a previous study showing that at least 120 words are necessary to obtain stable results during reading in controlled settings.[40] In general, the need for a high-quality microphone and controlled environment can be replaced by a longer sample duration.

Speech Biomarkers

In agreement with previous studies,[17,18,31,41] speech disorder in iRBD was mainly characterized by monopitch and imprecise vowel articulation. The novel observation is that vowel articulation was particularly affected during spontaneous speech, whereas intonation was reduced only during reading. This behavior can likely be explained by different compensatory mechanisms involved.[41] The low intonation pattern that is admissible during a reading of a prepared neutral passage is likely compensated for during dialogue to make the speech more compelling for the second side. On the contrary, deficits in internal cueing specific for PD might lead to higher intelligibility in prepared

utterances compared to spontaneous speech.[42] The quality of vowel articulation is highly related to intelligibility.[43] Because vowel articulation is a demanding process of articulatory coordination and intelligibility is preserved in the early stages of the disease, it may be difficult or unnecessary for patients to compensate during spontaneous speech. A previous study showed slower articulation rate and prolonged pauses during monologue in iRBD as predictors of development into parkinsonism.[44] In the present iRBD cohort, we observed only trends toward these changes, possibly due to lower sample size or better cognitive performance (average MoCA 25.9 vs. 24.8 previously). Considering our early PD cohort, the observed trend for worse voice quality, articulatory decay, and prolonged pauses is consistent with previous literature.[31,45] Interestingly, spontaneous speech assessment during phone calls led to the finest sensitivity in increased speech rate in PD, which is presumably a precursor of oral festination.[46] Contributing to palilalia, this is one of the most debilitating symptoms with no available therapies,[47] leading to social isolation and degradation of interpersonal interactions. Because laboratory speech material is typically short and not representative of everyday situations, it might not be sufficient for advanced analyses. Therefore, spontaneous speech evaluation through phone calls in natural settings may provide a novel approach to identify markers for predicting the onset of symptoms such as oral festination, potentially paving the way for more effective personalized therapies.

Effect of Smartphone Assessment on Individual Speech Biomarkers

The characteristics of the microphone, environmental noise, position of the microphone, and hardware filtering can all influence the robustness of speech assessment.[48] Many relationships were still surprisingly strong considering that smartphones and laboratory microphone recordings were not done in parallel but at different times and situations. In accordance with previous research,[21] the acoustic measurement of fundamental frequency variability reflecting monopitch demonstrated high resistance against device effect. This is likely due to the nature of the fundamental frequency, which represents a major trend in the frequency domain of a speech signal, and thus can be detected accurately despite the influence of detrimental factors. Imprecise vowels, reflecting the position of resonant frequencies (so-called "formants"),[32] represent another frequency measure that demonstrated good robustness in the analysis of smartphone data. The voice quality measure was unsurprisingly robust only in a controlled environment without substantial noise.[22] Articulatory decay calculated from MFCCs represents, in principle, an amplitude measure. It demonstrated little resistance against the device effect, as the coefficients tend to be impacted by microphone position and type[49] and, therefore, is unsuitable for phone screening. Another amplitude measure, monoloudness, was robust only in reading text, probably because of varying conditions in calls. Due to the dialogue nature of calls, pauses cannot be directly compared to those from uninterrupted monologue. In reading, pauses were moderately correlated between the smartphone and the high-quality microphone, which could be due to the insufficient accuracy of speech-pause detection.[35] Articulation rate, calculated as the number of syllables per time, reached high reliability between both devices, indicating the high robustness of the automatic speech-to-text transcription independent of the microphone quality.[36]

Strengths and Limitations

In some participants, we did not obtain enough speech data from calls, likely due to factors such as reluctance to share all speech calls or difficulties associated with smartphone operation. However, most of the participants achieved at least 18 minutes of call speech, that is, duration sufficient to detect prodromal voice impairment. In future scenarios, smartphone skills are likely to be widespread across the population. Additionally, the software can be directly implemented on smartphones to process recordings immediately after completion, computing selected features as anonymized numerical values, thus eliminating the need for audio transfer ensuring maximum privacy by storing only the values of speech features. Furthermore, speech testing should be followed by a more detailed neurological and neuroimaging analysis if the screen is abnormal. Legally and ethically, it is essential to create frameworks that allow for large-scale passive monitoring while safeguarding privacy and security.[50]

A similar diagnostic accuracy of an AUC of 0.86 was observed between PD patients and controls compared to an AUC of 0.85 between iRBD patients and controls. This could be associated with the fact that all PD patients were on stable dopaminergic therapy, which has been shown to ameliorate several speech manifestations in the early stages of the disease.[51] Furthermore, we were unable to recruit enough older PD volunteers with less than 5 years of disease duration, resulting in the PD group being on average 10 years younger than the iRBD and control groups. The inclusion of an older control group likely also negatively affected the reported accuracy of PD diagnostics.

Conclusion

This study has revealed that phone calls provide a novel passive biomarker of prodromal and early

parkinsonism and has established a pipeline for the capture of speech biomarkers in real-world settings. Enhancing sensitivity through a combination with active speech tasks amplifies its potential. To roll out the developed technique in a multicenter, clinical, therapeutic trial, the sensitivity of speech features should be validated by a longitudinal design, tested across various smartphone models and different languages, and ideally implemented directly to smartphones to eliminate the need for transferring speech content. If proven successful, our tool might be broadly applied in neuroprotective trials, neurodegeneration screening, deep brain stimulation optimization, neuropsychiatry, speech therapy, population screening, and beyond. ■

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. Schenck CH, Boeve BF, Mahowald MW. Delayed emergence of a parkinsonian disorder or dementia in 81% of older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder. Sleep Med 2013;14(8):744–748.
2. Postuma RB, Gagnon JF, Bertrand JA, Génier Marchand D, Montplaisir JY. Parkinson risk in idiopathic REM sleep behavior disorder: preparing for neuroprotective trials. Neurology 2015;84(11):1104–1113. https://doi.org/10.1212/WNL.0000000000001364
3. Videnovic A, Ju YES, Arnulf I, et al. Clinical trials in REM sleep behavioural disorder: challenges and opportunities. J Neurol Neurosurg Psychiatry 2020;91(7):740–749. https://doi.org/10.1136/jnnp-2020-322875
4. Postuma RB, Berg D. Advances in markers of prodromal Parkinson disease. Nat Rev Neurol 2016;12(11):622–634.
5. Miglis MG, Adler CH, Antelmi E, et al. Biomarkers of conversion to α-synucleinopathy in isolated rapid-eye-movement sleep behaviour disorder. Lancet Neurol 2021;20(8):671–684.
6. Dauvilliers Y, Schenck CH, Postuma RB, et al. REM sleep behaviour disorder. Nat Rev Dis Primers 2018;4(1):1–16. https://doi.org/10.1038/s41572-018-0016-5
7. Högl B, Stefani A, Videnovic A. Idiopathic REM sleep behaviour disorder and neurodegeneration — an update. Nat Rev Neurol 2018;14(1):40–55.
8. Fereshtehnejad SM, Yao C, Pelletier A, Montplaisir JY, Gagnon JF, Postuma RB. Evolution of prodromal Parkinson's disease and dementia with Lewy bodies: a prospective study. Brain 2019;142(7):2051–2067. https://doi.org/10.1093/brain/awz155
9. Adams JL, Kangarloo T, Tracey B, et al. Using a smartwatch and smartphone to assess early Parkinson's disease in the WATCH-PD study. npj Parkinsons Dis 2023;9(1):1–10. https://doi.org/10.1038/s41531-023-00497-x
10. Bot BM, Suver C, Neto EC, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016;3:160011. https://doi.org/10.1038/sdata.2016.11
11. Lakovakis D, Hadjidimitriou S, Charisis V, et al. Motor impairment estimates via touchscreen typing dynamics toward Parkinson's disease detection from data harvested in-the-wild. Front ICT 2018;5:28. https://doi.org/10.3389/fict.2018.00028
12. Papadopoulos A, Kyritsis K, Klingelhoefer L, Bostanjopoulou S, Chaudhuri KR, Delopoulos A. Detecting parkinsonian Tremor from IMU data collected in-the-wild using deep multiple-instance learning. IEEE J Biomed Health Inform 2020;24(9):2559–2569. https://doi.org/10.1109/JBHI.2019.2961748
13. Burton A. Smartphones versus Parkinson's disease: i-PROGNOSIS. Lancet Neurol 2020;19(5):385–386.
14. Arora S, Baig F, Lo C, et al. Smartphone motor testing to distinguish idiopathic REM sleep behavior disorder, controls, and PD. Neurology 2018;91(16):1528–1538.
15. Duffy J. Motor Speech Disorders: Substrates, Differential Diagnosis, and Management. 4th ed. St. Louis: Elsevier; 2019.
16. Postuma RB, Lang AE, Gagnon JF, Pelletier A, Montplaisir JY. How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder. Brain 2012;135(6):1860–1870. https://doi.org/10.1093/brain/aws093
17. Rusz J, Janzen A, Tykalová T, et al. Dysprosody in isolated REM sleep behavior disorder with impaired olfaction but intact Nigrostriatal pathway. Mov Disord 2022;37(3):619–623. https://doi.org/10.1002/mds.28873
18. Skrabal D, Rusz J, Novotny M, et al. Articulatory undershoot of vowels in isolated REM sleep behavior disorder and early Parkinson's disease. npj Parkinsons Dis. 2022;8(1):1–7. https://doi.org/10.1038/s41531-022-00407-7
19. Braak H, Del TK, Rüb U, de Vos RAI, Jansen Steur ENH, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging 2003;24(2):197–211. https://doi.org/10.1016/S0197-4580(02)00065-9
20. Brabenec L, Mekyska J, Galaz Z, Rektorova I. Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. J Neural Transm 2017;124(3):303–334. https://doi.org/10.1007/s00702-017-1676-0
21. Illner V, Sovka P, Rusz J. Validation of freely-available pitch detection algorithms across various noise levels in assessing speech captured by smartphone in Parkinson's disease. Biomed Signal Process Control 2020;58:1–13.
22. Šimek M, Rusz J. Validation of cepstral peak prominence in assessing early voice changes of Parkinson's disease. J Acoust Soc Am 2021;150(6):4522–4533. https://doi.org/10.1121/10.0009063
23. American Academy of Sleep Medicine. International Classification of Sleep Disorders, Third Edition: Diagnostic and Coding Manual. Darien, IL: American Academy of Sleep Medicine; 2014:1387–1394.
24. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 2015;30(12):1591–1601. https://doi.org/10.1002/mds.26424
25. Goetz CG, Fahn S, Martinez-Martin P, et al. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): process, format, and clinimetric testing plan. Mov Disord 2007;22(1):41–47. https://doi.org/10.1002/mds.21198
26. Kopecek M, Stepankova H, Lukavsky J, Ripova D, Nikolai T, Bezdicek O. Montreal cognitive assessment (MoCA): normative data for old and very old Czech adults. Appl Neuropsychol Adult 2017;24(1):23–29. https://doi.org/10.1080/23279095.2015.1065261
27. Visser M, Marinus J, Stiggelbout AM, Van Hilten JJ. Assessment of autonomic dysfunction in Parkinson's disease: the SCOPA-AUT. Mov Disord 2004;19(11):1306–1312. https://doi.org/10.1002/mds.20153
28. Doty RL, Shaman P, Kimmelman CP, Dann MS. University of pennsylvania smell identification test: a rapid quantitative olfactory function test for the clinic. Laryngoscope 1984;94(2):176–178. https://doi.org/10.1288/00005537-198402000-00004
29. Kouba T, Illner V, Rusz J. Study protocol for using a smartphone application to investigate speech biomarkers of Parkinson's disease and other synucleinopathies. BMJ Open 2022;12(6):1–11. https://doi.org/10.1136/bmjopen-2021-059871
30. Darley FL, Aronson AE, Brown JR. Differential diagnostic patterns of dysarthria. J Speech Hear Res 1969;12(2):246–269. https://doi.org/10.1044/jshr.1202.246
31. Rusz J, Hlavnička J, Novotný M, et al. Speech biomarkers in rapid eye movement sleep behavior disorder and Parkinson disease. Ann Neurol 2021;90(1):62–75. https://doi.org/10.1002/ana.26085

32. Illner V, Tykalova T, Skrabal D, Klempir J, Rusz J. Automated vowel articulation analysis in connected speech among progressive neurological diseases, dysarthria types, and dysarthria severities. J Speech Lang Hear Res 2023;66(8):2600–2621. https://doi.org/10.1044/2023_JSLHR-22-00526

33. Illner V, Krýže P, Švihlík J, et al. Which aspects of motor speech disorder are captured by Mel frequency Cepstral coefficients? Evidence from the change in STN-DBS conditions in Parkinson's disease. INTERSPEECH 2023. Dublin: ISCA; 2023:5027–5031. https://doi.org/10.21437/Interspeech.2023-1744

34. Rusz J, Cmejla R, Ruzickova H, Ruzicka E. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. J Acoust Soc Am 2011;129(1):350–367. https://doi.org/10.1121/1.3514381

35. Hlavnička J, Čmejla R, Tykalová T, Šonka K, Růžička E, Rusz J. Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. Sci Rep 2017;7(1):1–13.

36. Radford A, Kim JW, Xu T, Brockman G, McLeavey C, Sutskever I. Robust speech recognition via large-scale weak supervision. Proceedings of the 40th International Conference on Machine Learning. Proceedings of Machine Learning Research; 2023:1182. http://arxiv.org/abs/2212.04356.

37. Illner V, Tykalová T, Novotný M, Klempíř J, Dušek P, Rusz J. Toward automated articulation rate analysis via connected speech in Dysarthrias. J Speech Lang Hear Res 2022;65(4):1386–1401. https://doi.org/10.1044/2021_JSLHR-21-00549

38. Laganas C, Iakovakis D, Hadjidimitriou SK, et al. Parkinson's disease detection based on running speech data from phone calls. IEEE Trans Biomed Eng 2022;65(9):1573–1584.

39. Rusz J, Tykalova T, Ramig LO, Tripoliti E. Guidelines for speech recording and acoustic analyses in Dysarthrias of movement disorders. Mov Disord 2021;36(4):803–814. https://doi.org/10.1002/mds.28465

40. Krýže P, Tykalová T, Růžička E, Rusz J. Effect of reading passage length on quantitative acoustic speech assessment in Czech-speaking individuals with Parkinson's disease treated with subthalamic nucleus deep brain stimulation. J Acoust Soc Am 2021;149(5):3366–3374. https://doi.org/10.1121/10.0005050

41. Thies T, Mücke D, Geerts N, et al. Compensatory articulatory mechanisms preserve intelligibility in prodromal Parkinson's disease. Parkinsonism Relat Disord 2023;112:105487. https://doi.org/10.1016/j.parkreldis.2023.105487

42. Sapir S. Multiple factors are involved in the dysarthria associated with Parkinson's disease: a review with implications for clinical practice and research. J Speech Lang Hear Res 2014;57(4):1330–1343. https://doi.org/10.1044/2014_JSLHR-S-13-0039

43. Weismer G, Jeng JY, Laures JS, Kent RD, Kent JF. Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. Folia Phoniatr Logop 2001;53(1):1–18. https://www.karger.com/Article/FullText/52649

44. Šubert M, Novotný M, Tykalová T, et al. Spoken language alterations can predict Phenoconversion in isolated rapid eye movement sleep behavior disorder: a multicentric study. Ann Neurol 2023;95:530–543. https://doi.org/10.1002/ana.26835

45. Jeancolas L, Mangone G, Petrovska-Delacrétaz D, et al. Voice characteristics from isolated rapid eye movement sleep behavior disorder to early Parkinson's disease. Parkinsonism Relat Disord 2022;95:86–91. https://doi.org/10.1016/j.parkreldis.2022.01.003

46. Moreau C, Ozsancak C, Blatt JL, Derambure P, Destee A, Defebvre L. Oral festination in Parkinson's disease: biomechanical analysis and correlation with festination and freezing of gait. Mov Disord 2007;22(10):1503–1506. https://doi.org/10.1002/mds.21549

47. Nancy H. Management of Palilalia with a pacing board. J Speech Hear Disord 1979;44(3):350–353. https://doi.org/10.1044/jshd.4403.350

48. Kardous CA, Shaw PB. Evaluation of smartphone sound measurement applications. J Acoust Soc Am 2014;135(4):186–192. https://doi.org/10.1121/1.4865269

49. Rusz J, Novotny M, Hlavnicka J, Tykalova T, Ruzicka E. High-accuracy voice-based classification between patients with Parkinson's disease and other neurological diseases may Be an easy task with inappropriate experimental design. IEEE Trans Neural Syst Rehabil Eng 2017;25(8):1319–1321. https://doi.org/10.1109/TNSRE.2016.2621885

50. Martinez-Martin N, Insel TR, Dagum P, Greely HT, Cho MK. Data mining for health: staking out the ethical territory of digital phenotyping. npj Digit Med 2018;1(1):1–5. https://doi.org/10.1038/s41746-018-0075-8

51. Rusz J, Tykalova T, Novotny M, et al. Defining speech subtypes in De novo Parkinson disease. Neurology 2021;97(21):e2124–e2135. https://doi.org/10.1212/WNL.0000000000012878

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

SGML and CITI Use Only
DO NOT PRINT

Author Roles

(1) Research project: A. Conception, B. Organization, C. Execution; (2) Statistical analysis: A. Design, B. Execution, C. Review and critique; (3) Manuscript: A. Writing of the first draft, B. Review and critique.

V.I.: 1A, 1B, 1C, 2A, 2B, 3A
M.N.: 1B, 1C, 2C, 3B
T.K.: 1B, 1C, 2C, 3B
T.T.: 1B, 1C, 1B, 2C, 3B
M.Š.: 1C, 2C, 3B
J.Š.: 1C, 2C, 3B
E.R.: 1C, 2C, 3B
K.Š.: 1C, 2C, 3B
P.D.: 1B, 1C, 2C, 3B
J.R.: 1A, 1B, 1C, 2A, 2C, 3A

Full financial disclosures of all authors for the preceding 12 months

V.I. received funding from the Czech Ministry of Health and Czech Technical University in Prague. M.N. received funding from the Czech Ministry of Education, Czech Ministry of Health, and Czech Technical University in Prague. T.K. received funding from the Czech Ministry of Education and Czech Technical University in Prague. T.T. received funding from the Czech Ministry of Education, Czech Ministry of Health, and Czech Technical University in Prague. M.Š. received funding from the Czech Ministry of Education, Czech Ministry of Health, and Czech Technical University in Prague. P.S. received funding from the Czech Ministry of Health and Czech Technical University in Prague. J.Š. received funding from the Czech Technical University in Prague and the University of Chemistry and Technology. E.R. received funding from the Czech Ministry of Education, Czech Ministry of Health, and Charles University and General University Hospital. K.Š. received funding from the Czech Ministry of Education, Czech Ministry of Health, and Charles University and General University Hospital. P.D. received funding from the Czech Ministry of Education, Czech Ministry of Health, and Charles University and General University Hospital. J.R. received funding from the Czech Ministry of Education, Czech Ministry of Health, Czech Technical University in Prague, and Charles University and General University Hospital.