Jan Rusz

Detecting speech disorders in early Parkinson’s disease by acoustic analysis

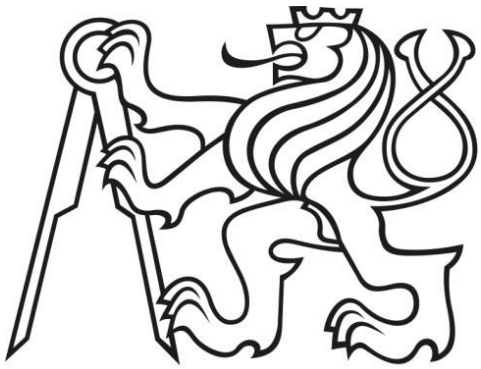
Czech Technical University in Prague

Faculty of Electrical Engineering

Department of Circuit Theory

Habilitation Thesis

Detecting speech disorders in early Parkinson's disease by acoustic analysis



Prague, October 2018

Jan Rusz

Copyright

Abstract

This interdisciplinary habilitation thesis is focused on the design of the feasible algorithms and analytical methods based on digital signal processing and advanced statistical analysis that are sensitive to capture pathological speech changes from very early stages of Parkinson's disease. Using objective acoustic analysis, we revealed distinctive speech impairment in patients with prodromal Parkinson's disease, newly diagnosed Parkinson's disease and atypical parkinsonian syndromes. Our findings suggest that automated vocal analysis may contribute to screening and diagnostic procedures to identify subjects at high risk of developing Parkinson's disease and related neurodegenerative disorders.

Anotace

Cílem této multidisciplinární habilitační práce je návrh vhodných algoritmů a analytických metod pro analýzu řeči založených na digitálním zpracování signálu a pokročilé statistické analýze, které budou dostatečně sensitivní a umožní zachycení patologických změn v řeči od velmi brzkých stádiích Pakinsonovy nemoci. S využitím objektivních metod akustické analýzy byla odhalena specifická forma řečové poruchy u pacientů s prodromální Pakinsonovou nemocí, nově diagnostikovanou Pakinsonovou nemocí a atypickými parkinsonskými syndromy. Tyto nálezy naznačují možnost využití automatické analýzy hlasu pro screeningové a diagnostické testy, které by umožnily identifikovat osoby ohrožené rozvojem Pakinsonovou nemocí a dalších extrapyramidových onemocnění.

Keywords

Speech and voice disorders; Acoustic analyses; Digital signal processing; Machine learning; Dysarthria; Parkinson's disease; Atypical parkinsonian syndromes; Progressive supranuclear palsy; Multiple system atrophy; REM sleep behavior disorder.

Klíčová slova

Poruchy hlasu a řeči; Akustické analýzy; Digitální zpracování signálu; Strojové učení; Dysartrie; Parkinsonova choroba; Atypické parkinsonské syndromy; Progresivní supranukleární obrna; Mnohočetná systémová atrofie; Porucha chování v REM spánku.

Acknowledgement

The topics presented in this habilitation thesis are results of an interdisciplinary work of the researchers from Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague and Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University. I would like to express my special gratitude to all my colleagues and friends, especially to Cecilia Bonnet, Roman Čmejla, Petr Dušek, Jan Hlavnička, Jiří Klempíř, Michal Novotný, Evžen Růžička, Hana Růžičková, Karel Šonka, and Tereza Tykalová, who directly and significantly contributed to papers collected in this habilitation thesis.

My acknowledgment also belongs to my colleagues Elmar Noth and Juan Rafael Orozco-Arroyave from the University of Erlangen in Germany, who further validated and extended my findings. Last but not least, I would like to thank all colleagues and their teams, who are willing to further collaborate on the topic proposed within this habilitation thesis, especially to Maria Teresa Pellecchia from University of Salerno in Italy, Ron Postuma from McGill University in Canada, Erik St. Louis from Mayo Clinic in US, Wolfgang Oertel from University of Marburg in Germany, Birgit Hogl from Medical University of Innsbruck in Austria, Yves Dauvilliers from Gui-de-Chauliac Hospital in France, Luigi Ferini-Strambi from San Rafaele Hospital in Italy, and Khalid Daoudi from Inria in France.

The studies included in this habilitation thesis have been financially supported mainly by the Czech Science Foundation, grant no. 102/12/2230, Ministry of Health of the Czech Republic, grant no. 15-28038A, and Czech Technical University in Prague under grants no. SGS12/185/OHK4/3T/13 and SGS15/199/OHK3/3T/13.

CONTENTS

General motivation regarding objective acoustic speech analyses

The ability to speak is very important for our daily life. The population of urban areas requires communication for their employment and common social life. Thus communication disorders have a strongly negative impact on such affected persons. Concerning people with communication disorders, the cost of care, as well as the following degradation of employment opportunities, has a major impact on the national economy. These circumstances indicate that communication disorders are one of the major medical challenges in the 21st century and that there is an urgent need for a reliable and cost-effective tools for quantifying and identification of specific speech deviations.

Speech is the most complex human motor skill, created by the coordinated actions of about 100 muscles. Speech production requires the integrity and integration of numerous activities such as speech planning and programming, cognitive-linguistics processes, and neuromuscular execution. Therefore, it is not surprising that the complexity of the act such as speech can be extremely sensitive to central nervous system diseases. Speech changes can be the only pathological manifestation in the early evolution of neurological disorder and sometimes represent the only significant neurological impairment. Therefore, identification of specific deviant speech characteristics can provide important clues about the underlying pathophysiology and localization of neurological diseases. Speech may also serve as a valuable marker of treatment efficacy, disease progression or even disease severity.

Acoustic analyses can help us conquer these challenges. Vocal, acoustic analyses are based on the digital signal processing of acoustic speech signal obtained from the microphone and represent a relatively novel assessment approach for speech disorders that holds promise in identifying a reliable, cheap, valid, and easy to administer biomarker of neurological disease. Indeed, this approach has received support across a rapidly growing number of studies.

Multidisciplinary research regarding speech disorders in neurodegenerative disorders requires the involvement of several scientific fields such as neuroscience, digital signal speech processing, and linguistics. Likely in this regard, the fact that recognition of speech changes can contribute to disease diagnosis and management is not widely recognized and taken advantage by practitioners in medicine or speech-language pathology. Our interdisciplinary approach unifies knowledge from several scientific perspectives and provides a fully automated solutions to obtain quantitative and transparent markers of neurodegeneration.

1. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder with pathological deposits of α-synuclein gradually spreading through peripheral and central nervous system. According to theory by Braak, α-synuclein aggregates are advancing along uniform predisposed pathways from the olfactory bulb and gut nerve plexus, to the brainstem and in the final stage also to the cerebral cortex [1]. Aggregated α-synuclein is toxic to affected cells ultimately leading to the loss of neuronal populations, notably of dopaminergic neurons in the substantia nigra, which is the direct cause of principal motor manifestations including bradykinesia, rigidity, and resting tremor [2]. The incidence of PD is estimated to be 1.8% in persons older than 65 years of age [3]. Neuroprotective treatment of neurodegenerative diseases such as PD is a strategic priority due to the increasing economic burden of prolonged life expectancy [4]. However, there is currently no treatment to halt or slow the progression of PD. Pharmacotherapy and neurosurgical interventions that are currently available only offer alleviation of certain symptoms. At the time of PD diagnosis, up to 50% of the neurons in the substantia nigra may be already irrecoverably damaged, and up to 80% of striatal dopamine has been depleted [5]. A major reason for the failure to develop disease-modifying therapy may be that the disease progresses for many years before the appearance of cardinal motor signs and then it is simply too late for intervention. Therefore, the early recognition of PD in prodromal stages has crucial implications for the future development of neuroprotective therapy [6,7].

Unfortunately, no sufficiently accurate biomarkers of preclinical PD are available that would allow early detection of PD to prevent the disease progression with potential neuroprotective therapies. In addition, no progression biomarker is available allowing to measure the effectiveness of experimental treatments on slowing the progression of PD. There are also no reliable means to identify people at high-risk for developing PD in the population. Thus, establishing a suitable biomarker would be a game-changing milestone that would impact diagnosis and future treatments of PD. Indeed, identification of potential biomarkers of prodromal PD, including early non-motor signs and markers of preclinical motor involvement, neuroimaging markers, and tissue biomarkers is becoming one of the most important topics in current PD-related research [8]. As motor abnormalities represent principal manifestations of PD, it is not surprising that impairments related to motor control appear to be a strong predictor of clinically manifest parkinsonism [7-9].

As the most complex human motor skill, involving more than 100 muscles, speech is a sensitive marker of damage to neural structures engaged in motor system control [10]. In fact, disorders of speech are among the most common clinical signs associated with PD. The vast majority of PD patients develop distinctive speech and voice abnormalities, collectively termed hypokinetic dysarthria, characterized mainly by the decreased quality of voice, hypokinetic articulation, hypophonia, monopitch, monoloudness and deficits in timing [11]. There is no doubt that speech disorder represents one of the earliest motor signs of PD. In the murine model of PD, ultrasonic vocalization deficits are among the first prodromal markers of motor dysfunction [12]. In humans, longitudinal voice changes in subjects at high risk for developing PD were estimated as the first motor signs which develop up to 10 years before the diagnosis, well before the appearance of rigidity, gait abnormalities and limb bradykinesia [9].

Thus, vocal assessment appears as an intriguing potential biomarker of PD and related neurodegenerations as it is inexpensive, non-invasive, simple to administer and can be performed remotely from the patients' home. Furthermore, acoustic analyses of speech in PD can be fully automated. Since the recording and processing of human speech is an area with an extensive background of knowledge, monitoring speech changes represents an excellent candidate as a preclinical diagnostic and progressive biomarker of PD.

2. State-of-the-art

2.1. Subjects at high risk for developing PD and other synucleinopathies

Rapid eye movement sleep behavior disorder (RBD) is a parasomnia caused by a lesion in the locus coeruleus complex located in the pons characterized by dream-enactment behaviors associated with REM sleep without muscle atonia [13]. Idiopathic RBD is a prodromal marker of neurodegenerative synucleinopathies, particularly PD and dementia with Lewy bodies (DLB) and less frequently multiple system atrophy (MSA). Importantly, due to a underlying α-synuclein pathology, a risk of developing neurodegenerative disease is extremely high (>80%) in subjects with RBD [14,15]. On the assumed pathway of α-synuclein spread, locus coeruleus is proximal to the substantia nigra, and symptoms of RBD precede parkinsonism. Therefore, research focused on RBD is essential for the development of neuroprotective therapy against synucleinopathy [7], as no other preclinical marker has comparable predictive value as RBD [16]. The high conversion rate of RBD to neurodegenerative disease provides a unique opportunity to study preclinical synucleinopathy, identify suitable preclinical biomarkers, and test disease-modifying therapies in the RBD group.

2.2. Atypical parkinsonian syndromes

Atypical parkinsonian syndromes (APS) such as progressive supranuclear palsy (PSP) and MSA differ from PD by more widespread neuronal involvement, resulting in additional clinical signs, more rapid disease progression and poor response to dopamine replacement therapy [17]. PSP and MSA are the most common APS, with an estimated prevalence of 30−40 per 100,000 among persons older than 65 years [17]. Characteristic clinical features of PSP include supranuclear gaze palsy, frequent falls, bradykinesia, axial rigidity, cognitive decline and communication disorders, reflecting widespread neurodegeneration involving the midbrain as well as the globus pallidus, striatum, hypothalamic nucleus, pons, superior cerebellar peduncle and cerebellar dentate nucleus [18]. Conversely, MSA manifests by various combinations of autonomic, cerebellar and parkinsonian features, corresponding to degeneration of the cerebellum, middle cerebellar peduncle, striatum, substantia nigra, inferior olivary nucleus and pons [19]. The underlying pathophysiology differs as PD and MSA are α-synucleinopathies while PSP is a tauopathy. However, the differentiation between PD and both PSP and MSA can be challenging as the initial signs are frequently nonspecific and overlap those of PD [18]. Accurate and early diagnosis is essential not only in assessing prognosis and making decisions regarding treatment but also for understanding the underlying pathophysiology and for the development of new therapies [20]. In the early stages of the disease, it is thus essential not

only to recognize between potential PD and healthy conditions but also between PD and different atypical parkinsonisms such as MSA and PSP.

2.3. Speech disorder in parkinsonism

Speech disorder is a common clinical manifestation occurring in 70–100% of patients with PD, PSP and MSA [21-24]. While the majority of PD patients develop a clear form of hypokinetic dysarthria [21,22], PSP and MSA patients typically evolve mixed dysarthria with various combinations of hypokinetic, spastic and ataxic components [23,24] due to the involvement of the basal ganglia, corticobulbar pathways, and the cerebellum. Only a few studies provided accurate objective descriptions of dysarthria in APS by acoustic analyses [25-27]. In general, these studies shown that the impairment of specific speech dimensions is more pronounced in APS than in PD [26-27]. Speech velocity, maximum phonation time, intonation variability and articulation precision were reduced, and pauses were prolonged in PSP in comparison to PD [26-27], while MSA patients manifested voice perturbations and slow and variable alternating motion rates [25]. Evidence regarding the occurrence of speech disorders in prodromal stages of PD supported by objective analysis is very rare. Only one previous studies reported cases with reduced intonation variability detectable several years before the onset of the first PD motor symptoms [28].

2.4. Available technologies for acoustic assessment of speech disorder in PD

Evaluation of dysarthria in PD is commonly performed by analysis of three types of vocal tasks. Those include sustained phonation, fast syllable repetition, and connected speech such as reading or monologue that can provide most of the information necessary for the objective description and interpretation of motor speech disorders [10]. Sustained phonation allows us to assess the regularity of vocal fold vibrations, fast syllable repetition measures the motor abilities of the speech articulators, and connected speech reflects a combination of speech-motor execution and cognitive-linguistic processing.

Majority of previous findings were based upon speech recordings obtained using a professional condenser microphone. Objective analyses of PD utterances have been traditionally performed using computer programs. There are several software packages allowing detection of various speech-related features such as pitch, loudness, jitter, shimmer, cepstral coefficients, formants, voiced/unvoiced segments, etc. The most popular of them include freely-available Praat [29], and commercially-available Multi-Dimensional Voice Program (MDVP; KayPENTAX, Lincoln Park, NJ). Unfortunately, these software packages frequently require user control of the analysis procedure. Nevertheless, novel approaches continuously show that more sophisticated analyses are possible. Indeed, since analysis and processing of speech disorders in PD has become an attractive scientific discipline in recent years, there is a number of vocal characteristics and advanced linear and non-linear methods available with proven efficiency in separating healthy controls from PD (for review, see Brabenec et al. [30]). However, the most widely used automated methods currently available are focused on the assessment of dysphonia via functional paradigm of sustained phonation [30], while methods allowing to yield distinctive PD-related speech patterns from connected

speech are rather scarce. Yet, PD speech performance varies across the specific task performed and demanding paradigms such as spontaneous speech are more likely to elicit speech deficits [31]. In particular, previous studies have investigated PD speech characteristics mostly in more advanced stages of disease on dopaminergic medication [30], making it difficult to generalize these observations to prodromal or very early PD. Dopaminergic medication may significantly affect the speech itself [32], making it hard to distinguish whether the observed symptoms are caused by PD symptoms or drug effects.

3. General aims of the thesis

The presented cumulative thesis comprises 10 peer-reviewed journal papers [A1-A10] with the following aims:

(i) To design the feasible algorithms and analytical methods that are sensitive and sufficiently accurate to capture pathological speech changes from very early stages of PD.

(ii) To determine specific dysarthria patterns and characterize the speech disorder in prodromal PD, early untreated PD, PSP, and MSA.

(iii) To estimate the reliability of speech assessment in differentiating between subjects at high risk for developing PD, patients with early PD, PSP patients, MSA patients, and heathy control speakers.

(iv) To improve the knowledge in the neurobiology of speech production by providing greater insights into the pathophysiology of speech disorder in PD and related neurodegenerative disorders.

4. General methods

Figure 1 shows the general methods used in the speech neuroscience research via a schematic overview. Although the particular methods used for speech evaluation are always dependent on the aim of individual study, the general approach might be described in the following four steps: (1) selection of available population sample and defining the inclusion/exclusion criteria; (2) recording of the larger speech protocol including mainly three types of speaking tasks of sustained phonations, syllable repetitions, and connected speech; (3) investigation of different patterns of speech disorder using suitable acoustic features based on digital signal processing methods; (4) design of suitable statistical approach to achieve estimated goal.

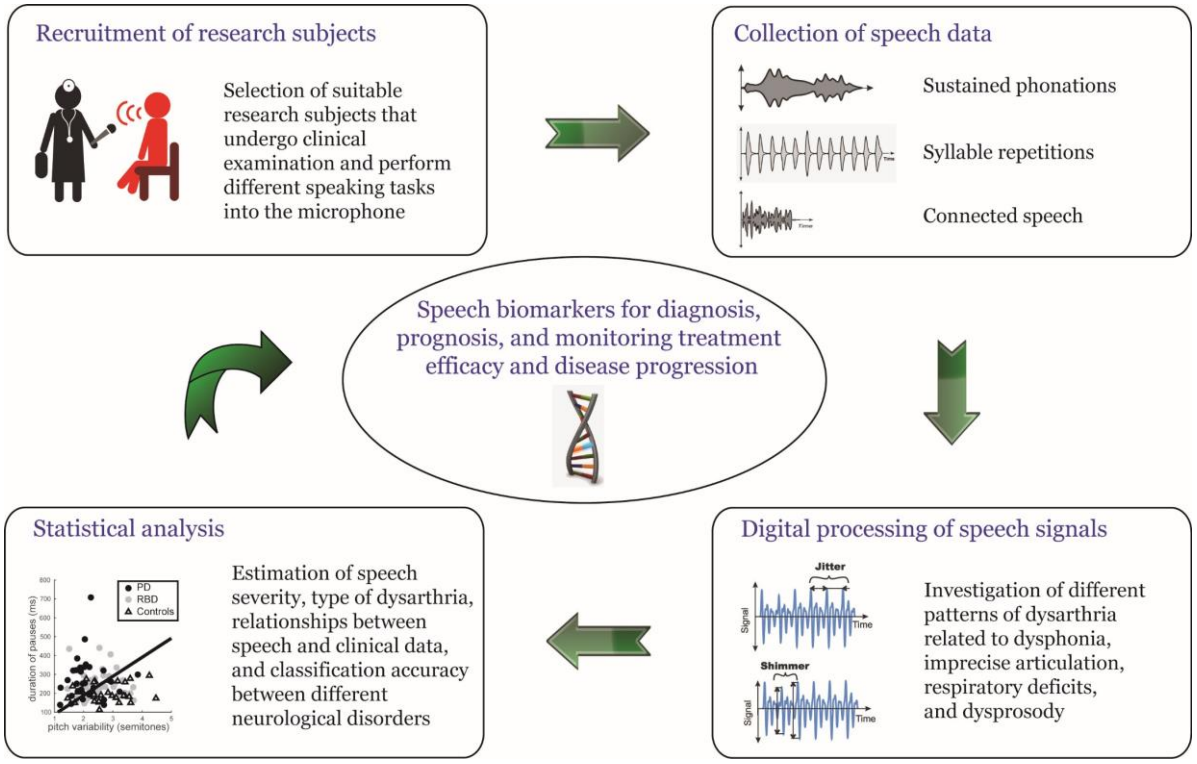


Figure 1: Schematic overview of general methods applied to evaluate motor speech disorders.

4.1. Research participants

From 2007 to 2017, a number of patients with idiopathic PD, RBD, probable PSP, probable MSA as well as healthy control speakers has been recruited and investigated for the individual studies [A1-A10]. All de-novo PD patients were diagnosed based on the Parkinson's disease Society Bank Criteria [33]. All RBD subjects were diagnosed with idiopathic RBD according to the International Classification of Sleep Disorders, third edition diagnostic criteria, including confirmation of REM sleep without atonia by polysomnography [34]. The diagnosis of PSP was established by the NINDS-PSP clinical diagnosis criteria [35], and MSA according to consensus diagnostic criteria for MSA [36]. All diagnoses were done by neurologists with experience in movement disorders. The healthy controls were without a history of neurological or communication disorders. All studies [A1-A10] were approved by the Ethics Committee of the General University Hospital, Prague, Czech Republic and all participants provided written, informed consent.

All PD patients were consecutively recruited at their first visit to the clinic and were examined before symptomatic treatment was started. No PD or RBD patient had a history of therapy with antiparkinsonian medication. None of the RBD participants subjectively complained of the motor or cognitive difficulties. In the PSP and MSA groups, medication consisted of various doses of levodopa alone or in combination with different dopamine agonists and/or amantadine. None of the patients received antipsychotic therapy. All PD and RBD patients were scored according to the motor score of the Movement Disorder Society–Unified Parkinson's Disease Rating Scale (MDS-UPDRS III, ranging from 0 to 132, with 0 for no motor manifestation and 132 representing severe motor disturbance) [37] or the previous

version of the UPDRS III (ranging from 0 to 108) [38]. PSP and MSA patients were rated by the natural history and neuroprotection in Parkinson plus syndromes–Parkinson plus scale (NNIPPS) [39]. Speech item of the respective scale was used for the perceptual description of speech severity (ranging from 0 to 4, with 0 representing normal speech and 4 indicating unintelligible speech).

4.2. Speech recording

Speech recordings were performed in a quiet room with a low ambient noise level using a professional head-mounted condenser microphone. The signals were sampled at 48 kHz with 16-bit resolution. Recordings were obtained during one session with a speech specialist who conveyed instructions to the subjects. There were no time limits during the recordings. All participants were asked to repeat their performance at any time if they or the examiner were not fully satisfied with their initial attempt.

Every participant underwent speech examination including various speaking tasks as a part of the larger protocol not exceeding 20 minutes. Investigation protocol consisted of various kind of speaking tasks including mainly three types: (a) sustained phonation of the vowel as long and stable as possible repeated two times, (b) fast repetition of syllables per one breath repeated two times, (c) connected speech including reading of standardized passage composed of 80 words and a monologue lasting at least 2 minutes on given topics including family, work, childhood or interests. Most of the speaking tasks were recorded twice to provide greater stability of speech assessment.

4.3. Acoustic analysis

The speech characteristics were particularly examined using free-available program Praat [29] or original algorithms developed in Matlab© (Mathworks, Massachusetts, USA). We quantitatively analyzed key dimensions of hypokinetic, spastic and ataxic dysarthria using most common aspects including subtests on phonation, articulation, and prosody.

As an example, regarding *phonation*, we focused on the evaluation of harsh voice using features like jitter, shimmer, harmonics-to-noise ratios, and others to quantify effect of dysarthria on voice quality (Figure 2A). In addition, dysarthria can be associated with excess fundamental frequency (F0) variations, strained-strangled voice or pitch breaks (Figure 2B). Concerning *articulation*, the first and second formant frequencies (F1, F2), their centralizations, and vowel space area represent the most widely used measurements; reduced vowel space area typically occur during imprecise vowel articulation in PD patients (Figure 2C). The most common deviations in dysarthrias that appear during *prosody* include reduced loudness, alterations in pitch, mono-loudness, and reduced intonation (Figure 1D). PD speakers also widely manifest different abnormalities in speech timing such as slow/variable rate or inappropriate silences, or changes in alternating motion rates affecting the length of consonants and vowels (Figure 1E). These aspects can be evaluated using features extracted using speech-pause detectors.

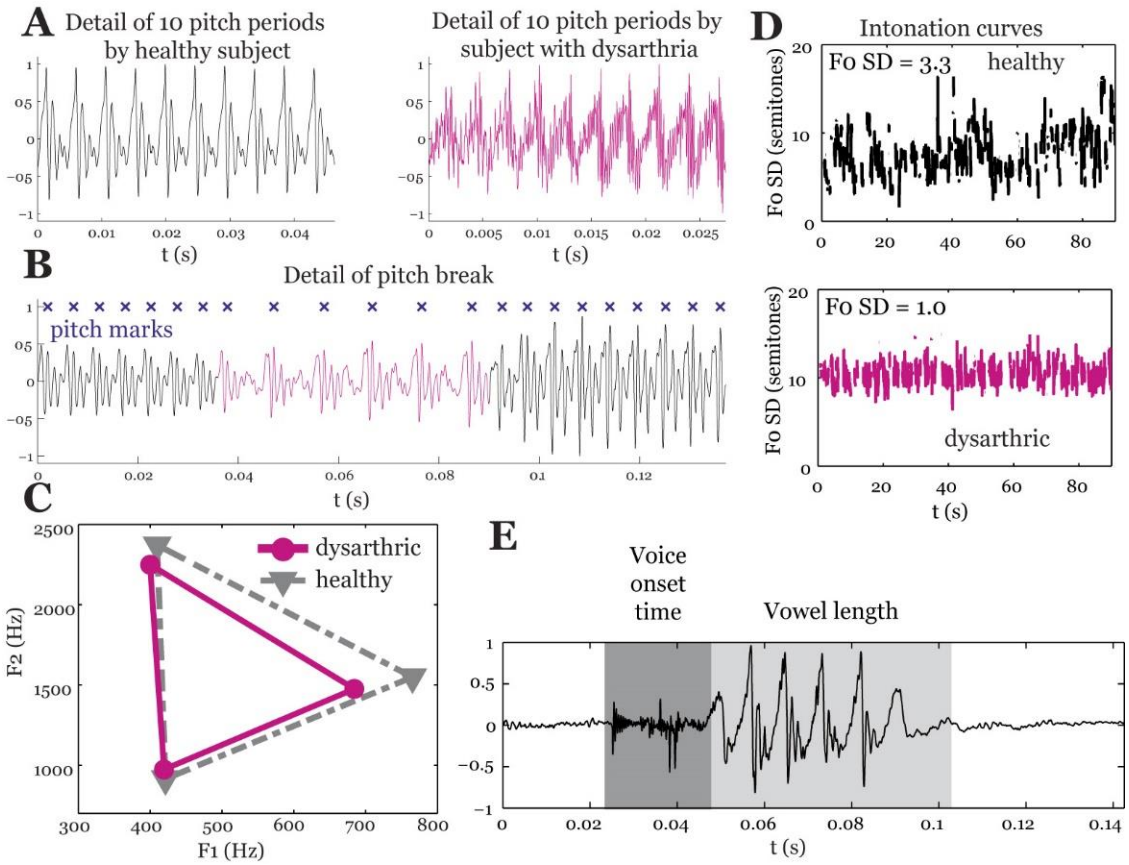


Figure 2. Example of acoustical signals and methods for evaluation of motor speech disorders: (A) harsh voice, (B) pitch break, (C) vowel space area, (D) intonation curves, (E) voice onset time and vowel length.

4.4. Statistical analysis

Individual parameters were typically first assessed using the Kolmogorov-Smirnov test for testing the normality of data distribution. Based on the character of results, alternative parametric or non-parametric tests were selected. For normally distributed data, we used the t-test or analysis of variance to assess group differences and Pearson analysis to search for correlations between variables. In case of non-normally distributed data, the Mann Whitney U test or Kruskal-Wallis were used to determine group differences and Spearman correlation to search for relationships. The minimal level of significance was always set as $p < 0.05$, mostly corrected with appropriate Bonferroni's adjustment. We also applied the sophisticated machine learning methods such as logistic regression, support vector machine, Wald task, k-means EM-algorithm and many others, as necessary. The validity of the obtained results (e.g., sensitivity/specificity) through classifiers were always based upon appropriate cross-validation scheme.

5. Discussion of main findings and potential impact of the thesis

5.1. Automated assessment of speech disorders in PD via acoustic analyses is possible

Until recently, the methods of objective monitoring of speech were limited to perceptual tests, with a limited accuracy depending on the examiner's skill. Most of methods enabling the objective evaluation of different patterns of dysarthria in PD are semi-automatic and require hand-labelling, or at a minimum, user control of the analysis procedure. Such hand-labelling is considerably time-consuming and requires an experienced investigator. Therefore, there is a general need for reliable, cost-effective and automatic methods allowing the precise and objective assessment of various speech patterns in neurodegenerative disorders. Increasing computational power has enabled a higher level of automation in speech assessment. However, most effort has been put into the automatic investigation of dysphonic features of dysarthria in PD through the sustained phonation task [30].

Therefore, we firstly developed the software package of traditional and non-traditional automated methods based on the digital signal processing of audio speech recordings allowing us to describe complexly multidimensional speech impairment in PD [A1]. PD speech measurement included fundamental frequency, perturbation measures, articulation rate, pause characteristics, intensity, diadochokinetic rate and regularity, formant frequencies, rhythm and several new methods for assessment of articulatory skills. Subsequently, we designed a novel, reliable automatic approach for the precise estimation of articulatory deficits in PD based on rapid, steady repetition of syllables [A5]. All phonemes of /pa/-/ta/-/ka/ syllables were first manually labelled. Based on these phoneme boundaries, 13 features describing six different articulatory aspects of speech including vowel quality, coordination of laryngeal and supralaryngeal activity, the precision of consonant articulation, tongue movement, occlusion weakening, and speech timing were analysed. Designed algorithm reached approximately 80% accuracy for a 5 ms threshold to detect correct phonemes boundaries. Finally, we developed a fully automated method that yields significant features related to respiratory deficits, dysphonia, imprecise articulation and dysrhythmia from the natural connected speech [A9]. A total of 200 randomly chosen recordings including both reading passages and monologues across PD, RBD, and healthy subjects was labeled blindly without awareness of segmentation output. The performance of the segmentation algorithm was evaluated for pause and respiration detection independently using these hand labels. The pause detection of our algorithm reached a very high accuracy of 86.2% and substantially outperformed conventional methods. Based on this segmentation approach, we designed 12 acoustic features allowing the assessment of all basic subsystems of connected speech including timing, articulation, phonation, and respiration.

Recent advances in semi-automated and automated methods of mathematic analysis of speech thus represent a significant breakthrough. They have become possible thanks to the advances in computer engineering and to a multifold increase in computational power that is approaching the target point of online analysis and differentiation of abnormal speech patterns. Shortly, when novel therapies appear to address the biological substrates of neurodegeneration, the fate of patients with PD and related neurological diseases may substantially change if

diagnostic markers are available, allowing to identify respective diseases in their prodromal phases. Automated speech analysis can thus become the most readily tool to recognize neurodegeneration in its earliest stage allowing for early initiation of effective therapy. Already in today's clinical practice, objective measures of speech can help to precisely evaluate the severity of speech impairment and moreover, they may serve as surrogate disease markers, helping to estimate the extent of overall patients' disability as well as to monitor the effects of therapy and rate of disease progression.

5.2. Multidimensional speech impairment is already detectable in the majority of PD patients at the time of diagnosis

The vast majority of previous evidence regarding speech disorders in PD was based on middle to the advanced stage of the disease and with different duration and doses of dopaminergic treatment [29]. However, the severity of speech disorder in PD increases with disease progression and certain speech deficits develop later in the course of the disease. In addition, dopaminergic medication can significantly influence the speech performance.

Therefore, we extensively explored the nature of speech and voice disorders in de-novo, untreated PD [A1-A5]. We showed that hypokinetic dysarthria is a multidimensional impairment affecting all different aspects of speech including phonation, articulation, and prosody with an unequal proportion and severity of speech patterns across individual PD patients [A1]. This research showed that 78% of early untreated PD patients indicated some form of speech impairment and highlighted that speech and voice disorders in PD are of a complex nature, suggesting importance to study all speech subsystem instead of focusing on a single one. Unlike the previous study based on a large sample of PD speakers documenting phonatory deficits as the most common sign [22], we revealed reduced melody as the most affected aspect of hypokinetic dysarthria. Further research confirmed that acoustic analyses are able to differentiate healthy speakers from de-novo PD patients with over 80% accuracy based upon three fundamental speaking tasks including sustained phonation, fast syllable repetition and running speech [A2]. Another study explored the suitability of imprecise vowel articulation as a possible early marker of PD [A3]. We performed a formant analysis of approximately 2800 vowels across different speech tasks including sustained phonation, short sentences, reading passage, and monologue. Imprecise vowel articulation was presented even in early stages of PD. Moreover, we found a significant effect of speaking task on vowel articulation performance in PD, suggesting that spontaneous speech is more likely to elicit acquired articulatory deficits in parkinsonian dysarthria. Since reduced stress is considered as one of the most deviant speech dimension in PD and its objective investigation was largely sparse, subsequent study investigated the vocal expression of stress in PD [A4]. We performed quantitative acoustic analysis of the sentences with unnaturally emphasized words including measurements of pitch, intensity, and duration as well as a newly designed measure called stress pattern index. We revealed that reduced stress is a distinctive pattern of early hypokinetic dysarthria. Our findings highlighted the importance and relevance of the introduction of speech therapy, as PD patients were able to consciously improve their speech performance during the investigated task. Finally, we designed a new reliable automatic approach for the precise estimation of articulatory deficits based on rapid, steady repetition of /pa/-/ta/-/ka/ syllables [A5]. Based on this approach,

we achieved the best classification accuracy of 88% in separating between PD and healthy subjects. We found prolonged voice onset time as the most powerful indicator of hypokinetic dysarthria.

These above-mentioned studies provided large evidence for speech and voice disorders in early-stage PD speakers prior to dopaminergic treatment. In general, results highlighted that specific speech changes due to PD may have the potential to contribute to existing assessment batteries for early detection of speech impairment and monitoring the disease progression and treatment efficacy.

5.3. Speech disorder allows discrimination between PD, PSP, and MSA

The previous research on PSP and MSA was mainly limited to estimation of type and severity of dysarthria, whereas little effort has been previously put into the investigation of complex speech impairment across specific dysarthria patterns [23,24]. A direct, objective comparison between individual speech patterns in PSP and MSA patients has not been performed, and distinctive speech markers that would be suitable for the differentiation of various forms of parkinsonism remained generally unknown.

Our results showed that the speech disorder reflects the underlying neuropathology of PD and APS [A6]. Dysarthria was uniformly present in all patients with PSP and MSA and generally consisted of a combination of hypokinetic, spastic and ataxic components, whereas PD patients manifested pure hypokinetic elements. In comparing PSP and MSA, in addition to hypophonic monotony of parkinsonian speech, dysarthria in our PSP patients was dominated by increased dysfluency, decreased slow rate, inappropriate silences, deficits in vowel articulation and harsh voice quality, whereas patients with MSA more frequently manifested pitch fluctuations, excess intensity variations, prolonged phonemes, vocal tremor and strained-strangled voice quality. Objective speech measurements were able to discriminate between APS and PD with 95% accuracy (93 % sensitivity and 100% specificity) and between PSP and MSA with 75% accuracy (74% sensitivity and 81% specificity). Another study attempted to clarify potential differences in consonant articulation deficits of both voiced and voiceless plosives between PD and APS [A8]. We found prolonged voiceless plosives as a common pattern in parkinsonism, confirming the critical role of basal ganglia circuit involvement in articulatory undershoot of stop consonants. Furthermore, we revealed that voiced plosives were shorter only in MSA, while nearly intact in PD and PSP. Since the extent of disruption of voiced plosives in MSA correlated with clinical severity of cerebellar involvement, the alteration in voiced consonant duration may represent a novel marker of cerebellar atrophy.

The relevance of these observations to the neurobiology of language and clinical practice is imminent. Careful evaluation of distinct speech characteristics may be diagnostically helpful in distinguishing between diseases with similar clinical manifestations but differing underlying brain pathophysiology. Moreover, a better understanding of processes underlying speech involvement is essential for optimization of treatment strategies as well as speech therapy improving the quality of patients' life.

5.4. Speech abnormalities indicate prodromal neurodegeneration in patients with high risk of developing PD

Our previous research showed that detailed voice analysis could differentiate newly diagnosed, untreated PD patients from controls with over 80% accuracy [A2]. Identification of different patterns of vocal disorder in the preclinical course of PD neurodegeneration has been severely restricted. The only study to track voice and face changes in prodromal parkinsonism did note that these appeared to be the first motor signs to develop [16]. However, this research used a very simple and crude analysis of voice and face changes using a subjective 4-point rating scale.

Based upon a quantitative acoustic assessment of 15 speech dimensions, we were the first to describe the presence of speech impairment in RBD [A7]. Speech disorder was already observed in 14 from 16 investigated RBD subjects and was detectable with a sensitivity of 96% and specificity of 79% when compared to healthy speech. Interestingly, when we applied this consolidated methodology to de-novo PD, we were able to achieve prediction of hypokinetic dysarthria with 99% sensitivity and 88% specificity. Since RBD is considered as a prodromal stage of parkinsonism, we may assume that observed speech abnormalities represent prodromal markers of neurodegeneration. In the subsequent study, automated analysis of connected speech revealed similar speech timing deficits in independent cohorts of 50 RBD and 30 de-novo PD patients [A9]. The main speech abnormalities found in RBD were prolonged duration of pauses, longer length of stop consonants and decreased rate of switching between follow-up speech segments. The values of these speech timing measures in RBD intermediated between those of de-novo PD patients and healthy controls, indicating a certain independence of speech disorder on parkinsonian phenotype. In general, speech disorders were more prominent in RBD subjects with higher motor scores on the Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [A7,A9], suggesting that speech impairment partially parallels increasing motor disability due to the underlying neurodegenerative process.

Very recent subsequent research showed that the detection of speech abnormalities in persons at high risk of developing PD and related disorders is possible even via a smartphone [A10]. Acoustic measures reflecting the reduced melody, increased duration of pauses, and slower rate of speech timing extracted from the spontaneous speech were sufficiently sensitive to significantly separate PD, RBD, and healthy control groups (area under curve of 0.85 between PD and controls and 0.69 between RBD and controls) and showed very strong correlation and reliability between the professional microphone and smartphone. Since the ability of a smartphone to capture prodromal speech impairment has never been previously investigated, these novel findings may provide considerable advantages for clinical practice as well as in future research. The smartphone allows easily and inexpensively increase the number of longitudinal vocal samples, which is critical not only for researchers performing traditional laboratory-based analyses but also for pharmaceutical companies developing drugs to treat disorders affecting motor speech performance such as in PD. Quick, inexpensive, and non-invasive vocal assessment by smartphone may help in the recruitment of appropriate cases into large studies of innovative therapies for prodromal PD and in the future may also bolster early presymptomatic diagnosis of PD and enable rapid access to neuroprotective therapy once it is available.

6. Conclusion

Figure 3 shows the summary of main findings via a schematic overview.

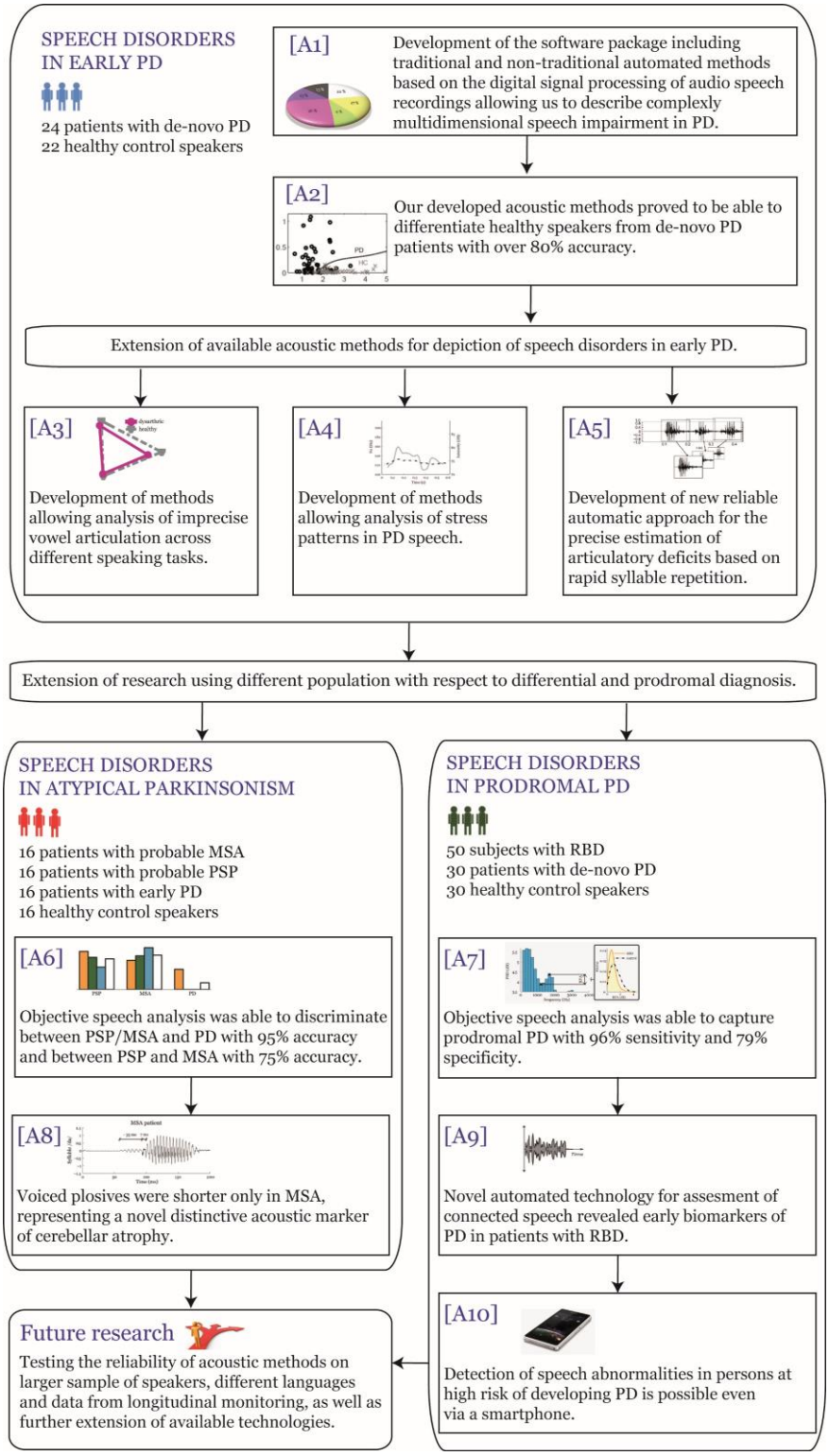


Figure 3: Schematic overview of the evolution of habilitation thesis over time and interconnection of individual papers [A1-A10] to the unified topic.

6.1. General contribution of the thesis

The main contribution of the research, hereby disclosed as the habilitation thesis, can be summarized into four points according to the general aims:

(i) We developed novel automated methods and methodological approaches allowing us to assess key aspects of (not only) hypokinetic dysarthria across a wide range of disease severities and fundamental speaking tasks such as sustained phonation, fast syllable repetition, and connected speech. We also showed that several features seem to be sufficiently independent on the quality of the microphone and can be potentially used in future monitoring through a smartphone.

(ii) We showed that newly diagnosed PD patient might already manifest multidimensional speech impairment affecting all speech subsystems including phonation, articulation, and prosody. Specifically, we found that early PD patients manifest reduced intonation, imprecise vowel and consonant articulation, harsh voice, monoloudness, inappropriate silences, decreased speech rate, and overall reduced stress patterns. While PD speakers manifested pure hypokinetic dysarthria, ataxic components such as pitch fluctuations, excess intensity variations, prolonged phonemes and vocal tremor were more affected in MSA. PSP subjects demonstrated severe hypokinetic and spastic components of dysarthria with distinctive patterns such as slow rate and dysfluency. In patients with RBD that represent a prodromal stage of parkinsonism, we already found slightly affected speech performance that intermediated between those of de-novo PD patients and healthy controls. The main speech abnormalities in RBD were similar to those in de-novo PD and included mainly reduced intonation, decreased rate of speech timing, and inappropriate silences.

(iii) Using our developed speech-based methods, we were able to separate newly diagnosed, untreated PD patients from healthy controls with up to 99% sensitivity and 88% specificity. Subsequently, speech measurements differentiated between APS and PD with 93 % sensitivity and 100% specificity and between PSP and MSA with 74% sensitivity and 81% specificity. Finally, we achieved 96% sensitivity and 79% specificity in discriminating between RBD subjects and healthy controls.

(iv) Based on a large cohort of RBD speakers and objective acoustic analyses, we provided strong evidence that speech impairment evolves from prodromal stages of PD and partially parallels increasing motor disability due to the underlying neurodegenerative process. Based upon investigation of APS, our results confirmed the distinctive critical role of basal ganglia and cerebellar control circuit involvement in articulatory undershoot of voiced and voiceless plosives.

From a practical point of view, among potential use in early and differential diagnosis, accurate evaluation and monitoring of speech abnormalities in PD may be helpful in the assessment of treatment efficiency, providing feedback to patients during speech therapy or assist clinicians in making different management decisions.

From an educational point of view, this research offered successful involvement of several doctoral and undergraduate students within their Ph.D. and diploma theses. The data collected within the habilitation thesis were used for practical educational purposes during two subjects including Biological Signals and Experimental Data Analysis. Also, data were used to propose challenging topics for students' semestral projects and contests related to Biosignal Challenge.

This interdisciplinary research also strengthened established collaboration between experts from various disciplines (signal processing, neuroscience, linguistics) and helped to establish new international partnerships.

6.2. Future work

There are several important goals related to the investigation of speech disorder in early PD that remained to be solved in future research:

(i) Fully automated algorithms for a description of several important speech dimensions associated with hypokinetic dysarthria such as dysfluency, imprecise vowel articulation or reduced stress have yet to be developed.

(ii) Majority of previous algorithms were tested only using speech recordings collected in common clinical environment with a low ambient noise level. In future, it is necessary to perform a test of the robustness of available speech features against noise and decide about the possibility of their inclusion in smartphone-based monitoring. In addition, the robustness of already existing methods for analysis of speech disorder could be further improved.

(iii) In clinically confirmed PD, it is likely that specific speech abnormalities are markers of distinct disease phenotypes. It remains to be proven whether the distinct variants of speech disorders in PD are consistent enough and to verify their predictive value for the development of individual disease phenotypes, with future potential clinical implications for prediction of the disease progression, response to medication, or late-stage complications.

(iv) The speech abnormalities seen in idiopathic RBD were not identical with those observed in PD, suggesting RBD to be a specific PD phenotype. It is therefore essential to assess various speech abnormalities in RBD and compare them to those of PD with and without the presence of RBD as well as to explore the relationships between speech aspects and other clinical and MRI markers.

(v) The predictive value of speech abnormalities in prodromal PD need to be established in further prospective follow-up. It is necessary to answer the question such as "How long before the formal diagnosis of PD can speech changes be identified with good sensitivity/specificity".

(vi) The previous findings were based mainly upon investigation of speech in the Czech language. Therefore, in collaboration with foreign experts, it is necessary to reproduce

these findings in an independent population, preferably with speakers of different languages.

(vii) It is necessary to expand the current research team and continuously involve new Ph.D. as well as undergraduate students in the solution of ongoing projects within this research field.

References

[1] Braak H, et al. (2004). Stages in the development of Parkinson's disease-related pathology. *Cell Tissue Res* **318**:121-134.

[2] Poewe W, et al. (2017). Parkinson disease. *Nat Rev Dis Primers* **23**:17013.

[3] de Rijk CM, et al. (2000) Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology* **54**(Suppl 5): S21–23.

[4] Findley LJ (2007) The economic impact of Parkinson's disease. *Park Relat Disord* **13**: S8–S12.

[5] Rodriguez-Oroz MC (2009) Initial clinical manifestations of Parkinson's disease: Features and pathophysiological mechanisms. *Lancet Neurol* **8**:1128–1139.

[6] Schenck CH, et al. (2013) Rapid eye movement sleep behavior disorder: devising controlled active treatment studies for symptomatic and neuroprotective therapy--a consensus statement from the International REM Sleep Behavior Disorder Study Group. *Sleep Med* **14**:795–806.

[7] Postuma RB, et al. (2015) Parkinson risk in idiopathic REM sleep behavior disorder: preparing for neuroprotective trials. *Neurology* **84**:1104–1113.

[8] Postuma RB, et al. (2016) Advances in markers for prodromal Parkinson disease. *Nat Rev Neurol* **12**:622-634.

[9] Postuma R, et al. (2012) How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder. *Brain* **135**:1860-1870.

[10] Duffy JR (2013) Motor Speech Disorders: Substrates, Differential Diagnosis, and Management, 3er ed. Elsevier Health Sciences.

[11] Ho AK, et al. (1999) Speech impairment in a large sample of patients with Parkinson's disease. *Behav Neurol* **11**:131-137.

[12] Grant LM, et al. (2014) Vocalization deficits in mice overexpressing alpha-synuclein, a model of pre-manifest Parkinson's disease. *Behav Neurosci* **128**:110-121.

[13] Boeve B (2010) REM sleep behaviour disorder. *Ann NY Acad Sci* **1184**:15–54.

[14] Iranzo A, et al. (2013) Neurodegenerative disease status and post-mortem pathology in idiopathic rapid-eye-movement sleep behaviour disorder: an observational cohort study. *Lancet Neurol* **12**:443-453.

[15] Schenck CH, et al. (2013) Delayed emergence of a parkinsonian disorder or dementia in 81% of older males initially diagnosed with idiopathic REM sleep behaviour disorder (RBD): a 16-year update on previously reported series. *Sleep Med* **14**:744-748.

[16] Postuma R, et al. (2012) Identifying prodromal Parkinson's disease: pre-motor disorders in Parkinson's disease. *Mov Disord* **27**:617-626.

[17] Schrag A, et al. (1999) Prevalence of progressive supranuclear palsy and multiple system atrophy: a cross-sectional study. *Lancet* **354**:1771–1775.
[18] Nath U, et al. (2003). Clinical features and natural history of progressive supranuclear palsy - A clinical cohort study. Neurology, 60:910-916.
[19] Gilman S, et al. (2008) Second consensus statement on the diagnosis of multiple system atrophy. *Neurology* **71**: 670-676.
[20] Wenning GK, et al. (2011) Milestones in atypical and secondary Parkinsonisms. *Mov Disord* **26**:1083–1095.
[21] Ho AK, et al. (1998) Speech impairment in a large sample of patients with Parkinson's disease. *Behav Neurol* **11**:131–137.
[22] Logemann JA, ET AL. (1978) Frequency and cooccurrence of vocal tract dysfunction in the speech of a large sample of Parkinson patients. *J Speech Hear Disord* **43**:47–57.
[23] Kluin KJ, et al. (1993) Perceptual analysis of speech disorders in progressive supranuclear palsy. *Neurology* **43**:563–566.
[24] Kluin KJ, et al. (1996) Characteristics of the dysarthria in multiple system atrophy. *Arch Neurol* **53**:545–548.
[25] Kim Y, et al. (2010) Perceptual and acoustic features of dysarthria associated with multiple system atrophy. *J Med Speech Lang Pathol* **18**:66–70.
[26] Skodda S, et al. (2011) Acoustical analysis of speech in progressive supranuclear palsy. *J Voice* **25**:725–731.
[27] Saxena M, et al. (2014) Assessing speech dysfunction using BOLD and acoustic analysis in parkinsonism. *Park Relat Disord* **20**:855–861.
[28] Harel BT, et al. (2004) Acoustic characteristics of Parkinsonian speech: a potential biomarker of early disease progression and treatment. *J. Neurolinguistics* **17**:439–453.
[29] Boersma P (2001) PRAAT, a system for doing phonetics by computer. *Glot Int* **5**:341–345.
[30] Brabenec L, et al. (2017) Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. *J Neural Transm* **124**:303-334.
[31] Kempler D, Van Lancker D (2002) Effect of speech task on intelligibility in dysarthria: A case study of Parkinson's disease. *Brain Lang* **80**: 449–464.
[32] Pinto S, et al. (2004) Treatments for dysarthria in Parkinson's disease. *Lancet Neurol* **3**:547-556.
[33] Hughes AJ, et al. (1992) Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* **55**:181–184.
[34] American Academy of Sleep Medicine (2014) *International Classification of Sleep Disorders, Third Edition: Diagnostic and Coding Manual*," Westchester, Illinois: American Academy of Sleep Medicine.
[35] Litvan I, et al. (1996) Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele–Richardson–Olszewski syndrome): report of the NINDS-SPSS international workshop. *Neurology* **47**:1–9.
[35] Gilman S, et al. (2008) Second consensus statement on the diagnosis of multiple system atrophy. *Neurology* **71**:670–676.

[37] Goetz CG, et al. (2008) Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord* **23**:2129–2170.

[38] Stebbing G, and Goetz C (1998) Factor structure of the Unified Parkinson's Disease Rating Scale: Motor Examination Section. *Mov Disord* **13**: 633-636.

[39] Payan CA, et al. (2011) Disease severity and progression in progressive supranuclear palsy and multiple system atrophy: validation of the NNIPPS-Parkinson plus scale. *PLoS One* **6**:e22293.

List of author's publications

[A1] **Rusz J**, Cmejla R, Ruzickova H, Ruzicka E (2011) Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *J Acoust Soc Am* **129**:350-367.

[A2] **Rusz J**, Cmejla R, Ruzickova H, Klempir J, Majerova V, Picmausova J, Roth J, Ruzicka E (2011) Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test. *Mov Disord* **26**:1951-1952.

[A3] **Rusz J**, Cmejla R, Tykalova T, Ruzickova H, Klempir J, Majerova V, Picmausova J, Roth J, Ruzicka E (2013) Imprecise vowel articulation as potential early marker of Parkinson's disease: effect of speaking task. *J Acoust Soc Am* **134**:2171-2181.

[A4] Tykalova T, **Rusz J***, Cmejla R, Ruzickova H, Ruzicka E (2014) Acoustic Investigation of Stress Patterns in Parkinson's Disease. *J Voice* **28**:129.e1-129.e8.

[A5] Novotny M, **Rusz J***, Cmejla R, Ruzicka E (2014) Automatic evaluation of articulatory disorders in Parkinson's disease. *IEEE/ACM T Audio Speech Lang Process* **22**:1366-1378.

[A6] **Rusz J**, Bonnet C, Klempir J, Tykalova T, Baborova E, Novotny M, Rulseh A, Ruzicka E (2015) Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *J Neurol* **262**:992-1001.

[A7] **Rusz J**, Hlavnicka J, Tykalova T, Buskova J, Ulmanova O, Ruzicka E, Sonka K (2016) Quantitative assessment of motor speech abnormalities in idiopathic REM sleep behaviour disorder. *Sleep Med* **19**:141-147.

[A8] Tykalova T, **Rusz J***, Klempir J, Cmejla R, Ruzicka E (2017) Distinct patterns of consonant articulation among Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Brain Lang* **165**:1-9.

[A9] Hlavnicka J, Cmejla R, Tykalova T, Sonka K, Ruzicka E, **Rusz J*** (2017) Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. *Sci Rep* **7**:12.

[A10] **Rusz J**, Hlavnicka J, Tykalova T, Novotny M, Dusek P, Sonka K, Ruzicka E (2018) Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease. *IEEE Trans Neural Syst Rehabil Eng* 26: 1495-1507.

*Supervision of the study/corresponding author.

Appendices

Appendix A1

Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease

Rusz J, Cmejla R, Ruzickova H, Ruzicka E (2011) Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *J Acoust Soc Am* **129**:350-367.

Journal name (category): **Journal of the Acoustical Society of America** (acoustics)
Quartile in category: **Q1**
Impact factor: **1.644**
Number of WoS citations (excluding autocitations): **74**

Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease

J. Rusz[a] and R. Cmejla
Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 116 27, Prague 6, Czech Republic

H. Ruzickova and E. Ruzicka
Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University in Prague, Katerinska 30, 120 00, Prague 2, Czech Republic

(Received 3 December 2009; revised 15 July 2010; accepted 7 October 2010)

An assessment of vocal impairment is presented for separating healthy people from persons with early untreated Parkinson's disease (PD). This study's main purpose was to (a) determine whether voice and speech disorder are present from early stages of PD before starting dopaminergic pharmacotherapy, (b) ascertain the specific characteristics of the PD-related vocal impairment, (c) identify PD-related acoustic signatures for the major part of traditional clinically used measurement methods with respect to their automatic assessment, and (d) design new automatic measurement methods of articulation. The varied speech data were collected from 46 Czech native speakers, 23 with PD. Subsequently, 19 representative measurements were pre-selected, and Wald sequential analysis was then applied to assess the efficiency of each measure and the extent of vocal impairment of each subject. It was found that measurement of the fundamental frequency variations applied to two selected tasks was the best method for separating healthy from PD subjects. On the basis of objective acoustic measures, statistical decision-making theory, and validation from practicing speech therapists, it has been demonstrated that 78% of early untreated PD subjects indicate some form of vocal impairment. The speech defects thus uncovered differ individually in various characteristics including phonation, articulation, and prosody.
 [DOI: 10.1121/1.3514381]

PACS number(s): 43.70.Dn [MAH] Pages: 350–367

I. INTRODUCTION

Parkinson's disease (PD) is a chronic neurodegenerative disorder characterized by the progressive loss of dopaminergic neurons in the substantia nigra.[1] PD is associated with dopamine deficiency and other affections of the brain neuromediator systems and accounts for a variety of motor and non-motor deficits.

PD is the second most common neurodegenerative disorder after Alzheimer's disease,[2] affecting over 1 million people in North America alone.[3] Previous studies suggest that PD usually affects people after the age of 50 yr; only approximately 10% of patients report symptoms before the age of 40 yr.[4] Moreover, PD is estimated to affect 1.6% of persons over the age of 65 yr.[5] Age is also the single most important factor for PD,with genetic predisposition second.[3] As a result, the statistics for the number of affected persons are expected to increase in proportion with the overall aging of the worldwide population as a whole.[6]

In addition to the most ostensible motor symptoms such as resting tremor, bradykinesia, muscular rigidity, and postural instability, many patients with PD develop non-motor deficits such as disorders of mood, behavior, and cognition and a distinctive alteration of speech characterized as hypokinetic dysarthria.[7,8]

Previous studies report that approximately 70%–90% of patients with PD show some form of vocal impairment,[9,10] and this deficiency may also be one of the earliest indicators of the disease.[11,12] Medical treatment, including neuro-pharmacological and neurosurgical methods, alleviates certain symptoms, but there is no causal cure now available, and early diagnosis of the disease has a vital role in improving the patients' live.[13,14] Research has shown that medical therapies alone are not as effective for treating speech symptoms as they are for motor functions,[15] and the effect of medical treatment on speech production tends to be individual.[16–18] Furthermore, only 3%–4% of PD patients receive speech therapy.[19] Behavioral speech therapy, including intensive voice treatment, appears to be the most effective type of speech intervention in the early and moderate stage of PD at present.[15,20] However, the requisite physical visit to the clinic for treatment is difficult and burdensome for many PD patients,[21] and the reduced ability to communicate is considered to be one of the most difficult aspects of the disease.[15]

Acoustical voice analyses and measurement methods might provide useful biomarkers for the diagnosis of PD in the early stage of the disease,[22] for possible remote monitoring of patients,[23] but above all, for providing important feedback in voice treatment for clinicians or patients themselves.

[a]Author to whom correspondence should be addressed. Electronic mail: ruszjan@fel.cvut.cz

For adult subjects, methods enabling the assessment of the speech impairment progress and the performance of the acoustic feedback tests can be essential for stimulating motivation and willingness for speech therapy. Acoustic measurements can also improve the individual treatment[16] and thus partially alleviate the inconvenience and cost of physical visits.[24] Moreover, voice measurement is non-invasive, cheap, and simple to administer.

The ability to speak can be subdivided into several dimensions, including respiration, phonation, articulation, and prosody.[25] The most salient features of PD are related to phonatory impairment, with the articulation being the second most affected speech subsystem,[10,26,27] although patients with PD can manifest abnormalities related to all of the dimension of speech, including monopitch, monoloudness, imprecise articulation, variability of speech rate, reduced stress, hoarseness, speech disfluencies, inappropriate silence, and others.[25,28] There are many voice and speech tests that have been devised to assess the extent of these symptoms including vocal recordings of sustained phonations, rapid syllable repetitions, and variable reading of sentences or freely spoken monologs. The speech signals are then commonly analyzed using several traditional measurement methods, which include sound pressure level, fundamental frequency, formant frequencies, speech rate, rhythm, and others.[28] A number of previous studies have used these methods to separate PD sufferers from a healthy control (HC) group, indicating that these standards could be useful measures in assessing the extent of vocal impairment, note, for example, Refs. 29–31.

In reality, however, the reliability and robustness of sound recording and measurement methods are impeded by several confounding issues including variables of physical condition and personal characteristics of the subject, such as, for example, gender and age. Thus the measurement methods performed on various vocal recordings should be chosen with an eye, as much as possible, to these confusing and in many cases even counteractive effects. Another relevant factor in determining the extent of PD vocal impairment is the dependence on the stage of the disease.[29] Although there are many studies using traditional measures performed on several vocal tasks for assessment of PD voice and speech disorders, there are no studies that can efficiently characterize the extent of vocal impairment and the suitability of these measures at the onset of PD, when the progression of symptoms of PD speech is not affected by medication.

Several speech recording and measurement methods may be needed to perform a reliable feedback test for the assessment of vocal impairment. For this reason, we introduce a brief PD-related characterization of voice and speech disorders, explaining the choice of traditionally used acoustic measures, and subsequently design specific measurement methods with a view toward their automatic assessment. There are many tests performed on simple sustained vowels for efficiently characterizing PD-specific dysphonia, including the traditional measures of fundamental frequency, variants of jitter and shimmer, and noise-to-harmonics (NHR) ratios. While articulation is the second most affected speech subsystem, there is a lack of available measures for its simple and efficient assessment. Therefore, we supplemented the traditional measures with new measures of articulation performed on rapid steady syllable repetition, which is the standard vocal test used to evaluate the articulatory skills.

Although statistically significant relationships between the extent of vocal impairment and measurement methods have been found in most of the traditional measures, statistical significance alone is not sufficient to determine the suitability of measurement methods for assessment of vocal impairment. Recently, many further innovative studies have appeared making use of acoustic measurement methods for voice disorder detection on the basis of machine learning tools—see, for example, Refs. 32–34. Consider the practical limitations of effort and cost-outcome associated with obtaining and verifying each of the methods, which are often dependent on a specific and unavailable speech sample, what is most needed is a reliable classifier that can determine the optimal set for classification from a varied number of independent available methods and speech samples.

The Wald task is a method from non-Bayesian statistical decision-making theory,[35] and it is given preference here because of its capability to separately assess each measure confronting the problem of making a decision in classifying subjects as PD, HC, or "not sure" in case of an indecisive situation. This latter case occurs when the rated observation does not provide enough information for a safe decision about assignment to the correct group. For complete assessment of vocal impairment, it is better to decide only in specific items where the rated observation clearly matches speech performance of the PD or HC group. With such a classification method, it is then possible to combine the user-selected traditional and novel measures. Nonetheless, there are still a number of measurements that can measure very similar aspects of a speech signal.

In order to gain an optimal amount of information for effective classification, in the present study we will first find and remove redundant and statistically insignificant measurements. Subsequently, the subset of available measurements will be used for the classifier based on the Wald task. On the basis of the classifier, we can not only discover the suitability of each measure for separating PD patients from HC but also the extent of vocal impairment in early untreated PD patients.

The organization of this paper is as follows. In the section "Methods," we describe the speech data and participants, introduce a brief review of classical acoustical PD speech analyses, detail the methods of speech measurements, and explain the statistics, pre-selection stage, and classification used in this study. In the section "Results," we present the results obtained. The section "Conclusion" contains a summary of our findings and provides a conclusion of the results for future work.

II. METHODS

The methodology of this study is broken down into eight stages: (a) the recruited participants; (b) the speech data; (c) a brief characterization of the PD speech; (d) calculation of traditional used clinical acoustic measures; (e) calculation of new non-standard acoustic measures; (f) the pre-selection of features and statistics; (g) the application of Wald's classifier to

pre-selected features; and (h) overall calculation of results and their validation by a speech therapist.

A. Participants

A grand total of 46 Czech native speakers were studied. Twenty-three individuals (19 men and 4 women) were diagnosed with an early stage of idiopathic PD [mean age, 61.74 yr [±standard deviation (SD), 12.60 yr]; duration of PD, 30.22 months (±SD, 22.21 months), Hoehn & Yahr stage 1–2, Unified Parkinson's Disease Rating Scale (UPDRS) III score 17.52 (±SD, 7.26)]. None of these PD subjects received symptomatic pharmacotherapy or speech treatment; all PD patients were examined in the drug-naive state, before the symptomatic treatment was started. In addition, 23 neurologically healthy speakers matched for age served as a control, including 16 men and 7 women [mean age, 58.08 yr (±SD, 12.91 yr)]. See Table I for subject details.

The Hoehn & Yahr scale is a commonly used system for describing the progression of symptoms of PD.[36] The scale comprises stages 1 through 5, where 1, unilateral involvement only usually with minimal or no functional disability; 2, bilateral or midline involvement without impairment of balance; 3, bilateral disease: mild to moderate disability with impaired postural reflexes, physically independent; 4, severely disabling disease, still able to walk or stand unassisted; and 5, confinement to bed or wheelchair unless aided.

The UPDRS part III score represents the motor rating known as UPDRS III, scaled from 0 to 108, with 0 representing a symptom-free state and 108 severe motor impairment.[37] The UPDRS III score encompasses areas such as tremor, rigidity, facial expression, speech, and others. Speech is ranked from 0 to 4, with 0 representing no signs of speech impairment and 4, complete unintelligibility.

B. Speech data and recording

The speech data were recorded in a sound-treated booth using an external condenser microphone placed at approximately 15 cm from the mouth and coupled to a Panasonic NV-GS 180 video camera. The voice signals were recorded directly to the computer, sampled at 48 kHz with 16-bit resolution; the purpose behind the use of video camera was the clinical examination of faciokinesia in PD patients, though no video material was used in the present study. The use of sound-treated booth (or at least a quiet room with a low ambient noise level) is recommended for its influence on assessment of intensity, articulation rate, and pause characteristics measurements, all of which are based on the energy of the signal and thus can be greatly influenced by noisy acoustic environments.

The vocal tasks used in this study ranged from producing isolated vowels to reading short sentences and producing a short, spontaneous monolog about a given subject. The duration of all of the vocal tasks used in this study was approximately 5 min [mean, 313.04 s (±SD, 36.40 s)]. See Table II for details of the vocal tasks.

The recording of each participant was obtained during a single session with a speech therapist. Recording began with a set of practice items to familiarize the speakers with instruction for the tasks and the recording procedure. No time limits were imposed during the recordings. Each participant was tested individually and received the production tasks in a fixed order. All participants were asked to repeat their production of an attempt that resulted in the erroneous production of any task, and they could repeat their production at any time if they or the speech therapist were not fully satisfied with their initial performance, though erroneous

TABLE I. List of participants with sex, age, and duration of disease prior to recording. Entries labeled "n/a" are for HC, for which duration of disease is not applicable.

Subject code	Sex	Age (yr)	Duration of PD prior to recording (months)
PD02	M	73	36
PD03	M	82	24
PD04	M	60	48
PD05	M	57	12
PD06	M	58	16
PD08	F	62	15
PD09	M	56	33
PD10	M	79	33
PD11	M	71	82
PD12	M	61	58
PD13	F	52	70
PD14	M	68	12
PD15	M	60	17
PD16	M	54	9
PD17	M	34	39
PD18	M	76	22
PD19	M	61	36
PD20	M	56	48
PD21	M	72	35
PD22	F	52	60
PD23	F	37	13
PD25	M	83	6
PD26	M	56	6
HC02	M	74	n/a
HC03	M	61	n/a
HC04	M	40	n/a
HC05	M	64	n/a
HC06	M	67	n/a
HC07	F	42	n/a
HC08	F	61	n/a
HC09	F	53	n/a
HC10	F	43	n/a
HC11	F	48	n/a
HC12	F	45	n/a
HC13	F	55	n/a
HC14	M	69	n/a
HC15	M	71	n/a
HC17	M	77	n/a
HC18	M	60	n/a
HC19	M	68	n/a
HC20	M	50	n/a
HC21	M	80	n/a
HC22	M	73	n/a
HC23	M	52	n/a
HC24	M	36	n/a
HC25	M	47	n/a

TABLE II. List of the vocal tasks.

Task code	Speech data
[TASK 1]	Sustained phonation of /i/ at a comfortable pitch and loudness as constant and long as possible, at least 5 s. [mean, 21.56 s (±SD, 7.98 s)]. This task was performed on one breath.
[TASK 2]	Rapid steady /pa/-/ta/-/ka/ syllables repetition as constant and long as possible, repeated at least 5 times [mean number of /pa/-/ta/-/ka/ 6.83 (±SD, 1.62)]. This task was performed on one breath.
[TASK 3]	Approximately 5-s sustained vowels of /a/, /i/, /u/ at a comfortable pitch and loudness [mean, 5.78 s (±SD, 0.57 s)]. The vowels were performed on one breath.
[TASK 4]	Reading the same standard phonetically non-balanced text of 136 word [mean, 57.52 s (±SD, 8.59 s)].
[TASK 5]	Monolog, at least approximately 90 s [mean, 109.96 s (±SD, 29.37 s), mean words, 232.50 (±SD, 86.24)]. The participants were generally instructed to speak about what they did current day or last week, their interests, their job, or their family.
[TASK 6]	Reading the same text containing 8 variable sentences of 71 words with varied stress patterns on 10 indicated words [mean, 39.78 s (±SD, 6.09 s].
[TASK 7]	Reading 10 sentences according specific emotions in a comfortable voice in response to an emotionally neutral sentence including excitement, sadness, confusion, fear, boredom, anger, bitterness, disappointment, wonder, and enjoyment [mean, 39.76 s (±SD, 6.11 s)].
[TASK 8]	Rhythmically read text containing 8 rhymes of 34 words following the example set by the examinator [mean, 24.22 s (±SD, 4.21 s)].

productions occurred rarely. The final vocal task productions were retained for acoustic analyses.

C. PD speech measurements

As discussed in the Introduction, abnormalities of the PD speech can be associated with several dimensions. Because it would far exceed the scope of this paper to discuss all speech measures, we briefly characterize only the traditional speech acoustics measures in PD related to this study, including phonation, articulation, and prosody. It is important to note that a deficit in respiration and quality of phonation may affect, among other things, the speaker's ability to produce normal phrasing and intensity. In addition, a decrease in respiratory pressure may cause deficits in phonation and articulation, i.e., decreased loudness and decreased ability to after loudness.[38]

Phonation is the vibration of the vocal folds to create sound.[39] In examining phonation in PD speakers, the most traditional measurements are performed during sustained vowel phonation and include measurement of F0 (the fundamental frequency or pitch of vocal oscillations), jitter (extent of variation of voice range), shimmer (the extent of variation of expiratory flow), and NHR ratios (the amplitude of noise relative to tonal components in the speech).[40] The other phonatory measure that has commonly been studied in PD is voice onset time (VOT), defined as the duration of time from articulatory release of a stop consonant to the onset of voicing for the following vowel.[41] VOT can be categorized as a phonatory measure because its changes in PD are generally attributed to disruptions of phonation.[42] Previous research has revealed PD-related dysphonia symptoms in all phonatory measures, including a higher mean value for F0 and increased variation of F0 in sustained vowel prolongation, and deficits in producing normal VOT.[29,42] It has been proposed that rigidity of laryngeal musculature causes a reduction in the opening of the vocal fold for PD patients in comparison to HC.[43]

Articulation is the modification of the position and shape of the speech organs (e.g., tongue) in the creation of sound.[39] In examining articulation in participants with PD, previous studies have reported that stop consonants were imprecise and were produced as fricatives.[43] This finding suggests that the articulatory deficits may have been partially the result of inadequate tongue elevation and constriction for stops and fricatives.[44] The most common method of evaluating articulatory skills is that of the diadochokinetic (DDK) task. Typically, the DDK task measures the subject's ability to repeat a consonant–vowel (C-V) combination with bilabial, alveolar, and velar places of articulation, quickly, at a constant level and a rhythmic manner. Subjects are asked to repeat a combination of the three-syllable item, for example, /pa/-/ta/-/ka/, as fast and long as possible.[45] A number of patients have demonstrated defects in the ability to make rapid articulator movements for DDK tasks.[46] Other measurements found differences in vocal tract resonances (i.e., formants), indicating increased variability of the first and second formant (F1 and F2, respectively) frequency. The centralization of these vowel formant frequencies is well captured by the vowel space area, and it can be a metric of tongue movement.[47] A few studies have reported smaller areas of vowel space for speakers with PD, but these differences were not significant.[47,48] Articulation measures also include measurement of the F2 slope (or F2 transition) from syllable repetition, representing the rate of tongue movement from a consonant into a vowel. The results found that F2 transition rates in PD patients were lower compared to HC.[42]

Prosody is the variation in loudness, pitch, and timing accompanying natural speech.[39] Prosodic measures are usually determined from running speech and include measurement of F0, intensity (relative loudness of speech), articulation rate, pause characteristics, and rhythm. A decreasing pitch range in PD has been noted during the reading task,[7,49] and various changes in speech rate and pause characteristics have also been found in people with PD in comparison to HC.[30,42,43] Prosodic intensity changes have also been examined, when PD patients produced significantly smaller intensity variation compared to normal speakers during the reading of a standard passage.[49] Overall, patients with PD demonstrate production defects in all of these measurements, including reduced frequency and intensity variations, and differences in speech rate and pause characteristics in reading tasks.

D. Traditional measurement methods

The present section of our study involves a selection of the major part of traditional clinically used measurement methods for PD-related voice disorders assessment.[28] These measurement methods are chosen and designed with attention paid to automatic feature extraction and to individual subject

TABLE III. Overview of measurement methods used as features applied to acoustic signals recorded from each subject.

Feature	Determined from	Speech subsystem	Description
1. Traditional			
F0 SD	[TASK 1] [TASK 4–7]	Phonation Prosody	Variations of fundamental frequency, vibration rate of vocal folds.
Jitter:local	[TASK 1]	Phonation	Average absolute difference between consecutive periods, divided by the average period.
Jitter:RAP	[TASK 1]	Phonation	Relative average perturbation, the average absolute difference between a period and the average of it and its two neighbors, divided by the average period.
Jitter:PPQ5	[TASK 1]	Phonation	Five-point period perturbation quotient, the average absolute difference between a period and the average of it and its four closest neighbors, divided by the average period.
Jitter:DDP	[TASK 1]	Phonation	Average absolute difference between consecutive differences between consecutive periods, divided by the average period.
Shimmer:local	[TASK 1]	Phonation	Average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude.
Shimmer:APQ3	[TASK 1]	Phonation	Three-point amplitude perturbation quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbors, divided by the average amplitude.
Shimmer:APQ5	[TASK 1]	Phonation	Five-point amplitude perturbation quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closest neighbors, divided by the average amplitude.
Shimmer:APQ11	[TASK 1]	Phonation	Eleven-point amplitude perturbation quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its ten closest neighbors, divided by the average amplitude.
Shimmer:DDA	[TASK 1]	Phonation	Average absolute difference between consecutive differences between the amplitudes of consecutive period.
NHR	[TASK 1]	Phonation	Noise-to-harmonics ratio, the amplitude of noise relative to tonal components.
HNR	[TASK 1]	Phonation	Harmonics-to-noise ratio, the amplitude of tonal relative to noise components.
Percent pause time	[TASK 4,5]	Prosody	The percent change from the unedited sample length to the edited sample length.
Articulation rate	[TASK 4]	Prosody	The number of syllables produced per second, after removing silence period exceeding 60 ms.
No. pauses	[TASK 4,5]	Prosody	The number of all pauses compared to total time duration, after removing silence period not lasting more than 60 ms.
Intensity SD	[TASK 4–6]	Prosody	Variations of average squared amplitude within a predefined time segment (“energy”) after removing silence period exceeding 60 ms.
DDK rate	[TASK 2]	Articulation	The number of /pa/-/ta/-/ka/ syllable vocalizations per second.
DDK regularity	[TASK 2]	Articulation	The degree of /pa/-/ta/-/ka/ syllable vocalizations rate variations in the period.
VOT	None	Phonation	Duration of time from articulatory release of a stop consonant to the onset of voicing for the following vowel.
Vowel area	[TASK 3]	Articulation	Quantitative measure which involves plotting the three corner vowels in F1/F2 plane.
Rhythm	[TASK 8]	Prosody	Measurement of ability to reproduce perceived rhythm through DTW.
2. Non-standard			
RIRV	[TASK 2]	Articulation	Relative intensity range variation, the variations of energy.
RRIS	[TASK 2]	Articulation	Robust relative intensity slope, the robust linear regression of energy.
SDCV	[TASK 2]	Articulation	Spectral distance change variation, the variations of spectral distance changes in signal spectrum.
RFPC	[TASK 2]	Articulation	Robust formant periodicity correlation, the first autocorrelation coefficient of F2 contour.

differences—see the first part of Table III for a list of the measures used as features in this part of the study.

1. The fundamental frequency

Standard methods include measures of the F0 mean, F0 range, and F0 SD. Although significant differences have been found between absolute and range values of F0 in PD patients compared to HC,[7,29] we do not use these as a measure, since they are affected by individual differences such as gender. In particular, the extent of F0 variation is related to the individual average voice pitch. Subjects with naturally high-pitched voices (traditionally women) will have much larger vibrato and microtremor than persons with lower-pitched voices (usually men),[50] thus causing a significant problem when these variations are measured on an absolute frequency scale in hertz. Observations suggest that the SD of the F0-distribution is approximately the same for men and women if it is expressed in semitones (logarithmic tonal scale). Specifically, a doubling of frequency, that is, 100–200 Hz or 200–400 Hz is represented by an equal semitone interval.[50] The observations also suggest that a logarithmic tonal scale will work better in capturing pitch variation due to speech impairment.[51]

The *fundamental frequency variation* (F0 SD) measurements were determined using several vocal tasks. First, for demonstrating the defects in phonation, we measured F0 SD on sustained vowel phonation [TASK 1]. In this measure, a higher value of F0 SD represents a dysphonic symptom of impaired control of stationary voice pitch. As we discussed earlier, people with PD often exhibit symptoms such as reduced melody variations during speech. Therefore, we performed F0 SD measurements using traditional voice

recordings, such as reading a text [TASK 4] and monolog [TASK 5]. Reduced melody variations in speech can also be related to the lowered ability of stress pronouncement and emotional intonation imitation and perception.[52] For this reason, we created two modified voice recordings. The vocal task of stress patterns [TASK 6] was designed to measure the subject's ability to produce unnatural increasing stress on labeled words. This ability of stress pronouncement can be then well captured by F0 SD measurements. The other newly set up vocal task [TASK 7] consists of 10 successive sentences pronounced with variable emotional context. The goal of this task is to evaluate how adults with PD express a particular emotion through prosodic features of their voice, in comparison with HC. The participants made the various intonations on the basis of the specific emotions, which should greatly improve variations on the final pitch.

For obtaining the F0 sequence, we used the application of the automatic algorithm of direct-time domain fundamental frequency estimation (DFE) and voiced/unvoiced (V/UV) classification of the speech signal.[53] The DFE algorithm consists of spectral shaping, detection of significant extremes based on adaptive thresholding, and actual frequency estimation under several truth criteria. These criteria are used to select the voiced part and eliminate estimation errors such as frequency halving and doubling. The first criterion is related to the level of the signal. No frequency estimations are performed for levels of signal lower than the threshold E_{th}. The actual level of energy is evaluated by an envelope detector; this criterion was set to approximately 0.5% level of the signal, and it was used as a noise gate. The second criterion is the expected frequency range of F0, with no frequency accepted outside of the specific range, which was set at 60–400 Hz. The third criterion is the M-order majority, whereby more than a one-half of M consecutive detected frequencies must lie in the same frequency band of chosen width. If the majority criterion M is satisfied, the actual signal is evaluated as voiced. Here, the majority criterion was set at five. As the last criterion, five-point median filtering was applied to the obtained F0 sequence to deal with incorrectly captured pitch periods outliers that may occur as a consequence of pitch doubling or pitch halving. The obtained pitch sequence was subsequently converted to the logarithmic semitone scale and its SD calculated. An optimal sampling frequency for DFE algorithm is 44 100 Hz—see Ref. 53 for more algorithm details. Among other things, this algorithm was applied to show that reliable automatic assessment of the F0 is possible. There is also the possibility of using novel robust pitch trackers,[54,55] that provide better F0 evaluation results.

Almost the same results can be obtained using the software PRAAT with the standard autocorrelation based procedure,[56] which was also used for validation of the obtained results. In comparison with DFE algorithm, though, the disadvantage of PRAAT lies in its need for checking the correct set up of the frequency range and other pitch settings as a consequence of pitch doubling and halving.

2. Variants of jitter and shimmer and NHR ratios

The most popular measurements of voice functions are the perturbation measures jitter and shimmer and their variants, and NHR ratios.[40,57] These measures were obtained using sustained vowel phonation [TASK 1].

Calculation of these measures is usually based on an autocorrelation method for determining the frequency and location of each cycle of vibration of the vocal folds (pitch marks).[58] The *jitter* and measures of period perturbation represent the variability of the speech fundamental frequency (pitch period) from one cycle to the next. The *shimmer* and measures of amplitude perturbation are derived from the sequence of maximum extent of the amplitude of the signal within each vocal cycle. Jitter and shimmer are used as measures to assess the micro-instability of vocal fold vibrations. From these perturbation measures, we used only measurements expressed as a percentage, as this method better reflects differences in gender. The NHR and *harmonics-to-noise* (HNR) ratios are derived from the signal-to-noise estimates from the autocorrelation of each cycle and are used for assessing voice hoarseness.

In this study, the measurements including jitter:local, jitter:RAP (relative average perturbation), jitter:PPQ5 (period perturbation quotient), jitter:DDP, shimmer:local, shimmer:APQ3 (amplitude perturbation quotient), shimmer:APQ5, shimmer:APQ11, shimmer:DDA, NHR, and HNR were calculated using algorithms supplied in the software package PRAAT.[56]

3. Articulation rate and pause characteristics

PD subjects reveal differences in articulation rate and pause characteristics during speech in comparison with HC.[30,42,43] In this study, *articulation rate, percent pause time, and number of pauses* were calculated for reading the text [TASK 4], while percent pause time and the number of pauses were also calculated for the monolog [TASK 5]. In order to perform an automatic assessment, we used only calculation of pause features in the monolog. The other speech material used in this study is not suitable for articulation rate and pause characteristics assessment because it consists of single sentences.

Percentage pause time calculation was based on the formula: $100 \times [(\text{total time} - \text{articulation time})/\text{total time}]$, where total time is the duration of the entire speech sample and articulation time is the length of time remaining after pause removal. The articulation rate was calculated after removal of pauses from each sample, where pauses were defined as silent periods lasting more than 60 ms that are not associated with stop closure. The articulation rate was calculated as the number of syllables produced per second after removing the pauses. Similarly, the number of pauses was then measured as the number of all pauses compared to total time duration, after removing the pauses not lasting more than 60 ms. Previous studies found significant differences after the removal of this time duration in PD patients compared to HC.[59]

In the present experiment, we designed a simple speech-pause detector based on signal intensity and zero-crossing rate (ZCR). We obtained the intensity and ZCR sequences of the entire speech signal and performed three thresholds, including, intensity mean value (IMV), intensity standard deviation values (ISDV), and zero-crossing rate mean value (ZCRMV). First, we compared the intensity of the current sample with the first threshold (IMV). If the sample has a

higher value than IMV, it is classified as speech. In the other case, we compare the actual intensity sample with the second threshold (|IMV − ISDV|). If it has a lower value, it is classified as silence. Once the sample is ranked in the <IMV − ISDV, IMV> interval, we compare the actual ZCR sample with the third threshold (ZCRMV) and, in case of a higher value, classified the sample as speech, while less value is classified as silence. The thresholds of the speech-pause detector were based on the experimental set up, and the results were validated by hand-marking. Algorithm evaluation using a speech-pause detector can be performed automatically.

4. Intensity of voice

PD speakers have been found to have an overall lower intensity level, deficits in intensity range, and intensity variations during speech production.[31] Similarly to the F0 measures, we do not use an absolute value of intensity level or an absolute range of intensity as measurements, based on a need for precise calibration for obtaining reliable estimates. As a result, we are restricted to relative measures of intensity variation with relative calibration to the reference of 0 dB. A precondition for successful measuring is then to maintain a constant distance from the microphone during the entire course of each recording.

The measurements of *intensity variations* (intensity SD) were determined using the reading text [TASK 4] and the monolog [TASK 5]. Similarly as in the F0 SD measurement, we also used measurement of stress patterns [TASK 6] with the aim of improving the intensity loudness variations.

The calculation of intensity variation was determined as a SD from the intensity sequences after removing all silence periods exceeding 60 ms to ensure that only clear speech was acquired. In this study, the window size of 1024 points (21.3 ms) was used to compute all energy contours. The intensity SD feature extraction can be performed automatically.

5. DDK rate and regularity

The DDK task is the measurement of the subject's ability to repeat rapidly and steadily a C-V combination and usually consists of two measures. The average *DDK rate* is the number of syllable vocalizations per second. The coefficient of *DDK regularity* measures the degree of rate variations in the period and assesses the ability to maintain a constant rate of C-V combinations. These two measurements were determined from repetition of the three-syllable items of /pa/-/ta/-/ka/ [TASK 2].

In order to devise a reliable algorithm for determining the DDK task measurements, we have to detect the local maxima (maximum energy during each syllable). First, we construct an integral envelope with the constant of integration set to 0.997. Subsequently, we normalize the integral envelop to the range [–1, 1]. Then we perform zero-phase digital filtering (averaging filter) by processing the input data in both the forward and reverse direction using a 1024-point size window. As a result, we arrive at smoothed sinusoidal signal that we again normalize to the range [–1, 1]. Finally, we estimate the local maxima which are computed from three continuous samples. Each sample value is compared to its neighboring value, and if it is larger than both of its neighbors, it is a local maximum. The feature extraction using this algorithm can be obtained automatically. The DDK rate is calculated as the number of estimate maximums per second and DDK regularity as the variance of the maximums.

6. VOT duration

The VOT is typically measured as the duration of time from the articulatory release of a stop consonant to the onset of voicing for the following vowel. VOT commonly refers to the temporal coordination between the oral articulation of a stop consonant and the laryngeal mechanism required to produce periodic vibration of the vocal folds. The measurement of VOT from the DDK task [TASK 2] may be a suitable measurement for detecting the extent of PD speech impairment, yet findings exist in the literature indicating that VOT changes in persons with PD are inconsistent.[42,43] Moreover, the PD speech impairment may be affected by hoarseness in the voice. Consequently, it is difficult to achieve a precise assessment of the VOT boundaries. Although the VOT is a traditionally used measurement method, we do not include it as a measure, because it is adversely affected by inconsistent results and no reliable algorithm exists for its reliable measurement.[60,61]

7. Formant frequencies F1 and F2

The main traditional measurement method using formant frequencies is the vocal tract *vowel area*. It is calculated by obtaining the mean values of the F1 and F2 frequencies during production of corner vowels and by subsequently plotting on an *xy* coordinate plane with F1 on the *x*-axis and F2 on the *y*-axis. This total area is calculated by measuring the entire triangle area. The vowel area was determined from phonation of three corner vowels including /i/, /u/, and /a/ [TASK 3].

We used the robust formant trackers of Mustafa and Bruce[62] for continuous speech with speaker variability for obtaining the formant sequences. The algorithm targets robust noise tracking and is based on a different approach than PRAAT, where the formant extraction relies on linear predictive coding (LPC) analysis. The algorithm works as follows: After a pre-emphasis and Hilbert transformation the signal is filtered by four formant filters. These are adaptive bandpass filters whose zeros and poles are updated based on the formant frequency estimates at the previous time stage, by means of which separation of formants into different channels can be achieved. A first-order LPC analysis performed on each of the four filter channels finally estimates the F1–F4 formants. Each formant filter consists of an all-zero filter cascaded with a single-pole dynamic tracking filter. The filter combinations are used to simplify normalization of the filter frequency response. The zeros and pole of each formant filter are updated for every sample; updating is based on the previous formant frequency estimates, allowing for dynamic suppression of interference from neighboring formants, while tracking an individual formant frequency as it varies over time. Finally, we obtain F1 and F2 formant sequences from the tracker and convert them to the logarithmic semitone scale and calculate their mean values. The entire total area is then calculated by the Euclidean distances between the F1 and F2 formant coordinates of the corner

vowels, and it is expressed in semitone squared. The formant tracker uses 8 kHz as an optimal sampling frequency.

The algorithm has a low signal delay and provides smooth and accurate estimates for the first four formant frequencies at moderate and high signal-to-noise ratios. Thorough testing of the algorithm has shown that it is robust over a wide range of signal-to-noise ratios for various types of background noises. The main advantage of the robust formant tracker is its full automatic assessment. The obtained results were also validated using PRAAT software, though its use can be regarded as optional.

8. Rhythm

A lowered ability to reproduce perceived speech rhythm may be one of the deficits in PD speech. We performed a speech measurement in which the participants were asked to repeat eight rhymes in the same rhythm prolongation as they heard in the reference speech sample (*template*) recorded by a speech therapist [TASK 8]. The purpose of the measure is thus for efficient comparison of the similarity between the subject and the template.

As a solution suitable for the measurement of rhythm, we used a technique known as dynamic time warping (DTW), a well-known method that has been used in speech recognition for aligning time series.[63] DTW uses the principle of dynamic programming (principle of optimality) in order to find the distance along the optimal warp path to determine the similarity between two speech waveforms.

To implement these insights algorithmically, the speech recordings were first down-sampled to 16 kHz. As features in DTW, we used a spectral representation of the speech data by calculation of the short-time Fourier transformation (STFT). We apply a Hamming window with a default size of 32 ms (512 points) and with a default overlap of 24 ms (384 points). In order to align utterance with the template, we created a similarity matrix, in which each point gives the Euclidean distance between short-time spectral analyses of the speech recordings. Subsequently, we used dynamic programming to find the lowest-cost path between the starts and ends of the sequences through the similarity matrix. Finally, this general cost of path distance normalized by the total sum cost of the matrix is used as the classifier for the relative *rhythm* similarity measurements between the individual's speech recording and the speech therapist's template. The measurement of rhythm can be performed automatically. An implementation process overview is shown in Fig. 1.

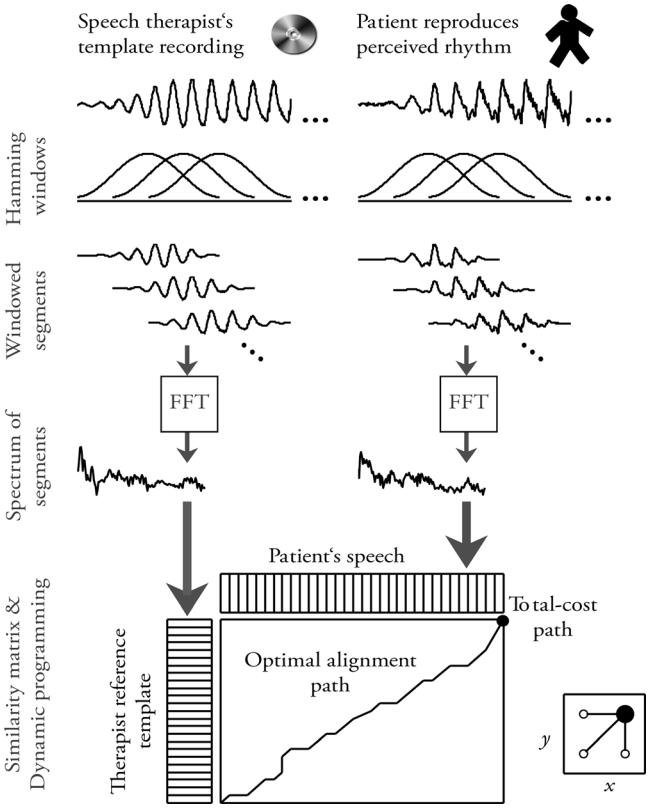


FIG. 1. The alignment process using DTW for measurement of ability to reproduce perceived rhythm. The resulting alignment path may be visualized as a low valley of Euclidean distance scores between speech spectrum segments of patients and speech therapist template recording, meandering through the hilly landscape of the matrix, beginning at $(0,0)$ and ending at the final point (X,Y).

E. New non-standard measurement methods of articulation

Articulation is one of the most strongly affected PD speech subsystems. The use of the DDK task allows for the performance of an efficient and a quick articulatory test. With such a measurement, we can efficiently assess the defects in PD articulation. As a consequence of rapid steady syllable repetition, problems can develop in the syllable rate and variations, but simultaneously significant defects can be present in respiratory pressure level, accuracy, and clarity of articulation. Thus, new measurement methods determined from the DDK task are introduced [TASK 2], which complement the standard DDK task measurements. Although the intensity measurements presented here below are more likely interconnected with problems in respiration, we introduced them as articulation measurements because the intensity defects can be developed simply as a consequence of rapid articulation. The DDK task is performed without pausing for breath. The feature extraction using these new proposed methods can be obtained automatically. See the second part of Table III for a list of the measures used as features in this part of study.

1. Relative intensity range variation (RIRV)

As we discuss in the Sec. II D 4, one of the observations related to the intensity deficits in PD dysarthria is reduced loudness. The other problem that can occur as a consequence of PD production deficits during fast articulation is occlusive weakening.[64] We notice that these PD-related differences may be captured when performing relative intensity contour during DDK task articulation in comparison to the HC. As a result, we performed RIRV measure calculation as a SD of the intensity curve with relative calibration to a reference of 0 dB. The first difference between measurements of RIRV and intensity SD is that the DDK task is performed without pausing for breath, and thus there is no need for removal of pauses from speech signal. The second difference is the fact

that here the occlusive weakening causes only lower variations in the relative intensity contour.

2. Robust relative intensity slope (RRIS)

One other problem occurring as a result of defects in respiratory function caused by PD-related dysarthria may be the inability to maintain the intensity level. For this reason, we perform a measure that we dub RRIS, a robust measurement of the intensity decline in performing the DDK task.

To implement it algorithmically, we need to perform a linear regression to calculate the slope of the relative intensity contour. Although the use of standard linear regression based on least squares estimation can be suitable for fitting the slope of the intensity contour, it can behave badly when the error distribution is outside of the normal range, particularly when the errors are heavy-tailed. Our approach is to employ a fitting criterion that is not as vulnerable to unusual data, such as that of least squares. Therefore, we perform a robust regression based on usage of iteratively reweighted least squares (IRLS) with a bisquare weighting function.[65] The IRLS algorithm uses weighted least squares, the influence of outliers being reduced by giving that observation a smaller weight. The weights chosen in a single iteration are related to the magnitudes of the residuals in the previous iteration, with a large residual earning a small weight. The weights are related to *M*-estimates, the measures of location that are not as sensitive as the mean to outlier values. See Ref. 65 for a detailed algorithm description. The final RRIS value is computed as total intensity decline divided by the total time duration of the DDK task.

The advantage of the robust fit achieved using the ILRS approach is that it is less influenced by the outliers than the least squared fit. Therefore, the robust intensity slope will be more suitable in practice when performing an automatic assessment.

3. Spectral distance change variation (SDCV)

The PD voice disorder is also affected by impaired clarity of articulation. The deficiencies of articulation clarity can be better demonstrated in the signal speech spectrum. In order to capture these deficits, we used the Bayesian autoregressive change-point detector.[66]

We consider the signal model for the Bayesian detector to consist of two parts, which are described by two different autoregressive models: the "left" autoregressive (AR) model with M_1 parameters a_k and the different "right" AR model with M_2 parameters b_k

$$d(n) = \begin{cases} \sum_{k=1}^{M_1} a_k \cdot d(n-k) + e(n), & n \le m \\ \sum_{k=1}^{M_2} b_k \cdot d(n-k) + e(n), & n > m \end{cases} \quad n = 1, \ldots, N, \quad (1)$$

where m is the change-point position, $e(n)$ is the excitation process with SD σ. The Eq. (1) can be written compactly in matrix form as $\mathbf{d} = \mathbf{G}_A \cdot \mathbf{b}_A + \mathbf{e}$, where, the matrix $\mathbf{G}_A$ has the Jordan form and depends on the unknown index of change-point m.

We likewise evaluate the value of the change between models. Using an analytical solution of the Bayesian theorem we obtain the formula for posterior probability, which is a function of the analyzed data, the signal length, and order autoregressive models only[67]

$$\tilde{p}(m|\mathbf{d}, \mathsf{M}) = \frac{\left(D - \mathbf{g}_A \mathbf{\Phi}_A \mathbf{g}_A^{\mathrm{T}}\right)^{-(N-M/2)}}{\sqrt{\Delta_A}}, \quad (2)$$

where the matrix $\mathbf{\Phi}_A = \left(\mathbf{G}_A^{\mathrm{T}} \mathbf{G}_A\right)^{-1}$ is the inverse correlation matrix, $D = \mathbf{d}^{\mathrm{T}}\mathbf{d}$ is the signal energy, $\mathbf{g}_A = \mathbf{d}^{\mathrm{T}}\mathbf{G}_A$ is the correlation vector, and $\Delta_A = \det\left(\mathbf{G}_A^{\mathrm{T}}\mathbf{G}_A\right)$.

The signal sample with the largest change in the signal (change-point) is determined by the maximum of the posterior probability, which is calculated from the Eq. (2). However, if there are more changes in the signal then the formula could not be used directly. The assumption of a single change is very restrictive in practice, since more abrupt spectral changes are invariably present in human speech. However, this drawback can be overcome by calculating the probability in a sliding window with fixed length and normalized using Bayesian evidence[66]

$$\tilde{p}(m|\mathbf{d}, \mathsf{M}) = \frac{\left(D - \mathbf{g}_A \mathbf{\Phi}_A \mathbf{g}_A^{\mathrm{T}}\right)^{-(N-M/2)}}{\sqrt{\Delta_A}} \times \frac{\sqrt{\Delta}}{\left(D - \mathbf{g}\mathbf{\Phi}\,\mathbf{g}^{\mathrm{T}}\right)^{-(N-M/2)}}. \quad (3)$$

The second term represents data dependent Bayesian evidence, where $\mathbf{\Phi}$, D, $\mathbf{g}$, and Δ are defined similarly to the previously established parameters but with respect to the entire signal segment without any division into left and right parts. Posterior probability [Eq. (2)] was derived from the Bayesian formula under the condition that a given data segment $\mathbf{d}$ is constant. Thus the Bayesian evidence in the denominator of the Bayesian formula was constant. But if the posterior probability is repeatedly used for new signal samples, then the data are not constant, and thus Bayesian evidence must be evoked to normalize.

The probability of the signal changes is then calculated from Eq. (3) for the sample signal which is situated in the middle of the rectangular window. In other words, the output of the algorithm is the degree of unlikeness between the signal in the left and right half of the window through which we pass all signals sample by sample. The normalized recursive autoregressive Bayesian change-point detector of sixth order with a windows length of 512 samples was used for detection.

We introduce a new measure called SDCV, a robust measure sensitive to observed changes in articulation clarity. The SDCV is calculated as a SD from the detector output, where the higher values of the output signal are proportional to the greater spectral distance of two adjacent segments and represent a greater clarity of articulation. A possibility likewise exists of using alternative detectors—see Ref. 68 for a comprehensive description.

4. Robust formant periodicity correlation (RFPC)

The F2 slope is a traditional measurement representing the rate of tongue movement from a consonant into a

vowel.[42] For assessing the accuracy of articulation, it is measured by comparing the F2 value at the onset of voicing to the F2 value in fixed time, for example, 50 ms into the vowel. As in the case of VOT measurement, the robust F2 slope assessment requires precise C-V boundaries detection.

In order to avoid designing a complex algorithm, we perform a new robust measurement called RFPC, the measure used to assess the similarity of tongue movement and thus accuracy of articulation. We use a robust formant tracker to obtain the F2 sequence from the DDK task.[62] The obtained formant values in low-energy segments are processed using a moving average filter, which ensures smooth tracking in unvoiced segments. The final obtained sequence represents the similarity of F2 slopes during entire vocal task. Then we simply estimate the first autocorrelation coefficient using the short-time autocorrelation, where a higher autocorrelation coefficient value means better articulation accuracy of the tongue.

F. Statistics and pre-selection stage

In practice, we need to find the relevance of individual measures that can subsequently be used to assess the extent of voice impairment. To obtain statistically significant differences between the groups, we compare the individual measures by using the non-parametric two-sided Wilcoxon rank sum test against the null hypothesis of equal medians, at a significance probability of 0.05.

Also, many measurements can be highly correlated with other measurements for the reason that they measure very similar aspects of the signal. For example, all types of shimmer features measure the extent of variation in speech amplitude cycle to cycle. Therefore, calculation of the Pearson product-moment correlation coefficient was used to test for significant correlations.

To set the best classification performance, we discard all measures with statistically non-significant relationships between the PD and HC groups. Subsequently, from all highly correlated measures with a correlation coefficient of greater than 0.95 (95% confidence interval), only one measurement will be kept which correlates with the greatest number of similar measurements and gains the most statistically significant differences between the PD and HC groups.

G. Classification stage

In this final stage, we apply the classification based on the Wald task to assess the relevance of the individual measures, as well as the extent of vocal impairment. The Wald task presents only a tiny part of scientific area known as Wald sequential analysis.[35,69]

This task is classification method based on the Neyman–Pearson task,[70] where the object is characterized by the feature x which assumes the value from the set X. There are two possible states including the normal one, $k = 1$, and the dangerous (undesirable) one, $k = 2$. Thus, the set of states K is $\{1, 2\}$. The probability distribution of the feature x depends on the state k to which the object belongs. The probability distributions are known and defined by a set of conditional probabilities $p_{X|K}(x|k),\ x \in X, k \in K$. For purpose of recognition, the set X is divided into two subsets $X1$ and $X2$. If the observation is $x \in X1$, the object is determined to be the normal state, and the dangerous state is thus for an observation $x \in X2$. In real conditions, some values of the feature x can occur both in the normal and dangerous states. The result of the decision is then characterized by two numbers where the first is the probability of an event that the normal state will be recognized as a dangerous state (false positive or false alarm), and the second one is the probability of the event that the dangerous state will be recognized as a normal state (false negative or overlooked danger). The conditional probability of the false positive state is given by

$$\omega(1) = \sum_{x \in X2} p_{X|K}(x|1), \tag{4}$$

and the conditional probability of the false negative state is then

$$\omega(2) = \sum_{x \in X1} p_{X|K}(x|2). \tag{5}$$

In the Neyman–Pearson task, the classification strategy is chosen from all strategies satisfying the above condition for which, first, the conditional probability of the false negative is not larger than a predefined value ε. Second, the conditional probability of the false positive is the smallest.

However, there is a lack of symmetry with respect to the states of the recognized object, which is apparent where the Neyman–Pearson task is recalled. To provide a thorough elimination of this lack of symmetry, the Wald task is not formulated as the set X of the two subsets $X1$ and $X2$ corresponding to a decision for the benefits of the first and second state but as a classification in three subsets $X0$, $X1$, and $X2$ with the following meaning:

if $x \in X1$, then $k = 1$,
if $x \in X2$, then $k = 2$,
if $x \in X0$, then it is decided that the observation x does not provide enough information for a safe decision about the state k.

The classification strategy is characterized by

$\omega(1)$ is the conditional probability of a wrong decision about the state $k = 1$,
$\omega(2)$ is the conditional probability of a wrong decision about the state $k = 2$,
$\chi(1)$ is the conditional probability of an indecisive situation under the condition that the object is in the state $k = 1$,

$$\chi(1) = \sum_{x \in X0} p_{X|K}(x|1), \tag{6}$$

$\chi(2)$ is the conditional probability of an indecisive situation under the condition that the object is in the state $k = 2$,

$$\chi(2) = \sum_{x \in X0} p_{X|K}(x|2). \tag{7}$$

For such strategies the requirements $\omega(1) \leq \varepsilon$ and $\omega(2) \leq \varepsilon$ are not contradictory for an arbitrary non-negative

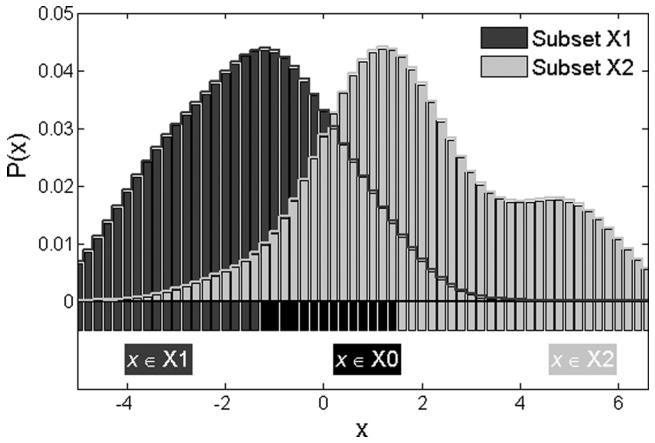


FIG. 2. The top part of the figure shows a selected example of probability densities. The bottom of the figure shows the result of Wald task classification. As a result, the dark-gray shaded bars are the regions in which the feature x assumes the value from the subset $X1$ and the light shaded bars are the areas predicted for the subset $X2$. The black bars represent the indecisive situation, for which observation $x \in X0$ does not provide sufficient information for a safe decision regarding one of the subsets $X1$ or $X2$.

value ε because the strategy $X0 = X, X1 = 0, X2 = 0$ belongs to the class of allowed strategies too. Each strategy meeting the requirements $\omega(1) \leq \varepsilon$ and $\omega(2) \leq \varepsilon$ is, moreover, characterized by how often the strategy is reluctant to decide, i.e., by the number $\max(\chi(1), \chi(2))$.

The Wald task seeks among the strategies satisfying the requirements $\omega(1) \leq \varepsilon$ and $\omega(2) \leq \varepsilon$ for a strategy which minimizes the value $\max(\chi(1), \chi(2))$. The solution of this task is based on the calculation of the likelihood ratio

$$\gamma(x) = \frac{p_{X|K}(x|1)}{p_{X|K}(x|2)}. \tag{8}$$

We used the Gaussian kernel density method with automatic data-driven bandwidth to estimate probability distributions from each measurement for PD and HC groups.[71] This pair of distributions represents the feature for classification.

For gaining the best classification accuracy, both predefined values ε were set to a 95% significance level. Finally, the linear programming technique was used to solve the Wald task. The comprehensive description of the solution of the Wald task through linear programming can be found in Ref. 35. Figure 2 shows the result of the Wald task classification applied to an example of a generated distributions pair.

H. Overall calculation and validation

For overall calculation, each of the measures selected by the pre-selection stage represents one feature for Wald task classification. In the case that the subject's speech performance in the set measure matches the disordered speech performance of the PD group (is classified as PD), the subject is rated by "1" positive point. In the other case, where the subject's speech performance matched the intact speech performance of the HC group (is classified as HC), the subject is rated by "−1" negative point. In case of an indecisive situation, where the subject's speech performance does not have sufficient predictive quality for secure assignment to one of the PD or HC groups, i.e., is not fully intact or impaired but matches the extent of speech performance of the wider norm, the subject is rated by "0" point. This "−1, 0, 1" three state-scale was designed with respect to physiological background, where we want to determine clearly if the tested subject reached the PD-specific (or intact) vocal performance in the selected task instead of giving the various weight to each classification. We want to be "sure" that any speech performance which has the possibility of belonging to the wider norm of healthy people (not obviously PD-specific or intact) will not be marked. This approach also gives the same weight to all measurements; we do not consider that the combinations of certain measurements may be more useful for overall classification performance as most classifiers do. For obtaining the final results, we calculate the sum of points for each subject. The higher quantity of positive points predicts the greater vocal impairment. The number of negative points corresponds to the performance of healthy speech production. The overall number of classified points for each measurement represents the suitability of the measurement in separating patients with PD from HC participants, and it is calculated as sum of all assigned values.

In order to validate our classification results, comparisons with speech therapist evaluations were performed. The speech therapist performed an independent examination based on various voice and speech recordings composed from a number of items including measuring of phonation and phonetics and then assessed each participant using a seven-point rating scale. The rating scale values represent the complete speech performance of each subject; a value equal to 1 point signifies intact speech performance, and a value equal to 7 represents progressing vocal impairment. Finally, the Pearson correlation coefficient was performed to ascertain the relationships between the score obtained from the speech therapist and the acoustic evaluation methods.

III. RESULTS

A. Voice and speech characteristics

The means, SDs, correlations between the measurement methods, statistical significances, and summaries of retained measures for the Wald classifier are listed for all measures in this study (see Table IV for more details). The results are presented below according to speech characteristics (i.e., phonation, articulation, and prosody).

Statistical significances between the PD and HC group were found in all measurements of phonation except pitch variations (F0 SD). This can be caused by the fact that people with early stages of PD need not show impaired control of stationary voice pitch during sustained phonation. On the other hand, more signal noise addition captured by NHR measures can indicate incomplete vocal fold closure and incorrect vocal fold oscillations. The noise in speech can be also generated by turbulent airflow through the vocal fold. Significant findings in measurements, including all types of shimmers and jitters features, NHR, and HNR, can be manifested clinically as hoarseness, hypophony, and tremolo.

From traditional articulatory measurements including DDK rate, DDK regularity, and vowel space area, only the

TABLE IV. List of results of all measures with mean values, SD values, correlations between the measurements methods, statistical significances, and summaries of retained measures. See main text for detailed description of the algorithm used to calculate these results.

Measurement	Subjects				Redundant to measurement?	Difference between groups	Retained for Wald task classification?
	PD		HC				
	Mean	SD	Mean	SD			
Phonation							
[TASK 1] Sustained phonation							
01. F0 SD (semitones)	0.46	0.49	0.35	0.23	No	$p = 0.29$	No
02. Jitter:local (%)	1.53	1.37	0.65	0.78	Jitter(RAP,PPQ5,DDP)	**$p < 0.01$**	No
03. Jitter:RAP (%)	0.88	0.81	0.38	0.52	Jitter(local,PPQ5,DDP)	**$p < 0.01$**	No
04. Jitter:PPQ5 (%)	0.83	0.75	0.32	0.32	Jitter(local,RAP,DDP)	**$p < 0.01$**	**Yes**
05. Jitter:DDP (%)	2.65	2.42	1.14	1.56	Jitter(local,RAP,PPQ5)	**$p < 0.01$**	No
06. Shimmer:local (%)	7.51	4.97	2.72	2.27	Shimmer(APQ[3,5,11],DDA)	**$p < 0.001$**	**Yes**
07. Shimmer:APQ3 (%)	3.69	2.57	1.39	1.36	Shimmer(local,APQ5,DDA)	**$p < 0.001$**	No
08. Shimmer:APQ5 (%)	4.37	3.07	1.45	1.13	Shimmer(local,APQ[3,11],DDA)	**$p < 0.001$**	No
09. Shimmer:APQ11 (%)	6.32	3.85	2.20	1.64	Shimmer(local,APQ5)	**$p < 0.001$**	No
10. Shimmer:DDA (%)	11.07	7.71	4.17	4.07	Shimmer(local,APQ[3,5])	**$p < 0.001$**	No
11. NHR (-)	0.16	0.27	0.02	0.04	No	**$p < 0.01$**	**Yes**
12. HNR (dB)	16.01	7.36	24.02	5.61	No	**$p < 0.001$**	**Yes**
Articulation							
[TASK 2] DDK task							
13. DDK rate (syll/s)	6.22	0.63	7.16	0.73	No	**$p < 0.001$**	**Yes**
14. DDK regularity (-)	0.54	0.58	0.67	0.36	No	$p = 0.49$	No
15. RIRV (dB)	7.54	1.52	10.99	1.96	No	**$p < 0.001$**	**Yes**
16. RRIS (dB/s)	2.75	1.51	1.16	1.12	No	**$p < 0.001$**	**Yes**
17. RFPC (-)	0.46	0.17	0.60	0.09	No	**$p < 0.01$**	**Yes**
18. SDCV (-)	0.14	0.03	0.18	0.03	No	**$p < 0.001$**	**Yes**
[TASK 3] Sustained vowels							
19. Vowel area (semitones^2)	94.19	29.24	95.10	25.84	No	$p = 0.66$	No
Prosody							
[TASK 4] Reading text							
20. F0 SD (semitones)	1.71	0.66	2.48	0.56	No	**$p < 0.001$**	**Yes**
21. Intensity SD (dB)	5.93	1.05	7.55	1.62	No	**$p < 0.001$**	**Yes**
22. Percent pause time (%)	0.30	0.02	0.29	0.02	No	$p = 0.30$	No
23. Articulation rate (syll/s)	6.09	0.78	6.09	0.84	No	$p = 0.58$	No
24. No. pauses (pauses/s)	3.29	0.67	3.98	0.51	No	**$p < 0.01$**	**Yes**
[TASK 5] Monolog							
25. F0 SD (semitones)	1.53	0.32	2.44	0.65	No	**$p < 0.001$**	**Yes**
26. Intensity SD (dB)	7.05	1.41	8.75	1.51	No	**$p < 0.001$**	**Yes**
27. Percent pause time (%)	0.32	0.03	0.31	0.03	No	$p = 0.14$	No
28. No. pauses (pauses/s)	3.04	0.83	3.86	0.69	No	**$p < 0.01$**	**Yes**
[TASK 6] Stress patterns							
29. F0 SD (semitones)	2.06	0.81	2.78	0.62	No	**$p < 0.01$**	**Yes**
30. Intensity SD (dB)	6.40	1.07	7.84	1.97	No	**$p < 0.01$**	**Yes**
[TASK 7] Emotional sentences							
31. F0 SD (semitones)	2.59	0.74	3.82	0.56	No	**$p < 0.001$**	**Yes**
[TASK 8] Rhythmic text							
32. Rhythm (-)	2.65	0.55	2.27	0.28	No	**$p < 0.01$**	**Yes**

DDK rate contains significant differences between both groups. The DDK regularity did not show any significant differences. Although the PD total vowel area was found to be slightly reduced in comparison with HC, there was no statistically significant difference. In Fig. 3, we can see the plot of the total vowel triangle area of the PD and HC groups. The patients with PD show abnormalities in each new non-standard articulation measurement. Figure 4 shows the result of calculating the RIRV, RRIS, SDCV, and RFPC values for a selected speech signal. As can be seen, the PD speech signal show lower similarity in repeated syllable production, which can indicate reduced movement of orofacial muscles. In many patients with PD are developed intensity defects in instances of rapid articulation. The reduced intensity variations that can be caused by occlusive weakening are demonstrated by the RIRV measure. As an example, voiceless occlusives, which are normally associated with a silent gap, tend to exhibit energy during the silent gab. This energy can be caused by turbulent noise generated at the site of oral constriction because of an incomplete occlusion or voicing

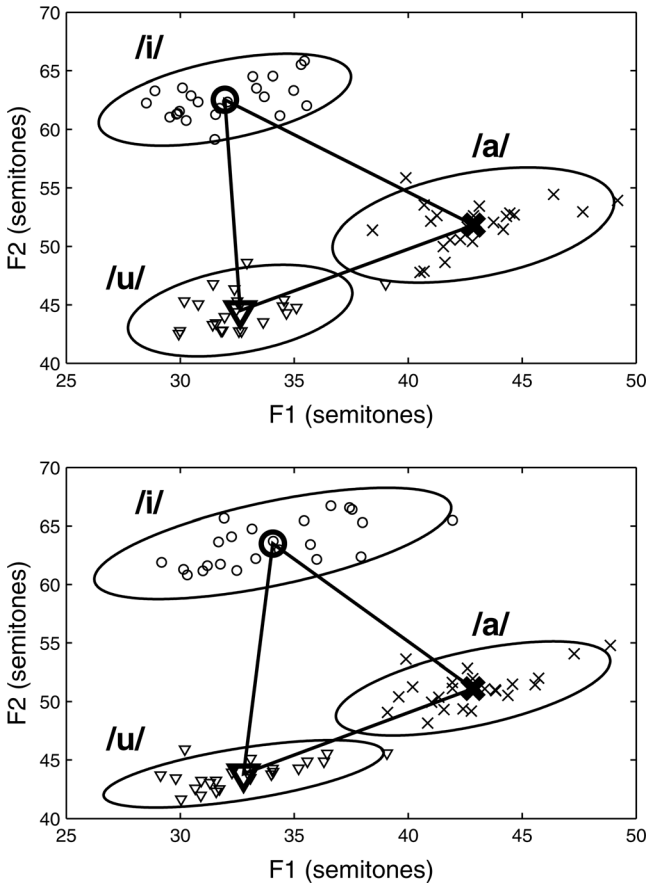


FIG. 3. The vowel triangle space areas for subjects with PD (up) and HC subjects (down).

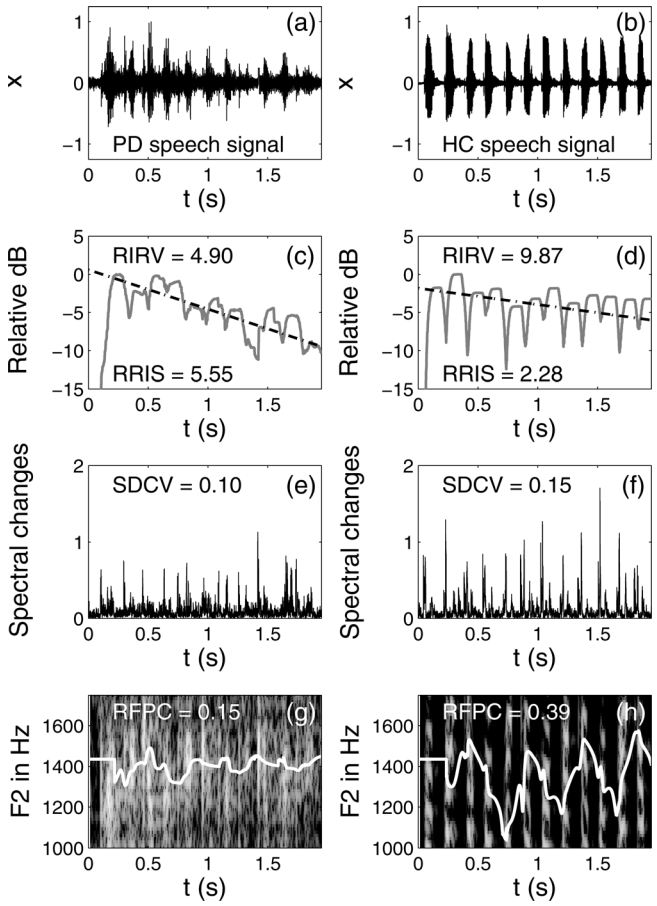


FIG. 4. Details of new articulation measures performed on rapid steady syllable repetition. (a and b) Speech signals of rapid steady /pa/-/ta/-/ka/ syllables repetition; (c and d) light gray lines represent obtained intensity sequences, dashdot lines represent the RRIS. The RIRV is computed as SD from the obtained intensity; (e and f) SDCV; (g and h) RFPC. The left panel is for a person with PD, the right panel is for a HC subject.

energy which occurs as a result of poor coordination between laryngeal and supralaryngeal gestures. The results of RRIS show that PD patients have a lower ability to maintain the intensity level, which can be caused by weakness in the production of stable airflow from the lungs. The remaining two measures involve the spectral speech changes. The higher number peaks in SDCV represent a greater clarity of articulation. The rate and similarity of tongue movement are well represented by the RFPC measure. The higher periodicity in the obtained F2 sequence represents better articulation accuracy of tongue.

Ten of 13 measures of prosody contained significant statistical results. The patients with PD show lower melody intonation in all F0 SD measurements and also decreased intensity variations in all intensity SD measurements. This situation can be caused by changed laryngeal tension, decreased breath support, and decreased range of motions. The persons with PD have not shown any significant differences in the articulation rate compared to HC. From pause characteristics, only the measurements of number of pauses show significant differences between groups. This can be indicated by breathiness and starting time of the tongue movement. The patients with PD also show a lower ability to reproduce perceived rhythm perception.

B. Data pre-selection

Statistically significant relationships between the HC and PD groups have been found in 26 of a total of 32 measures. The rest of the measures were statistically insignificant and were discarded. These include the measures of F0 SD in sustained phonation, DDK regularity, vowel area, percent pause time, and articulation rate.

The perturbation measures, including all kind of jitter and shimmer features, are highly correlated with correlation coefficients greater than 95%. The measurements of jitter:APQ5 and shimmer:local were retained for their optimal performance in separating HC from PD patients. Correlation filtering removes the following measures: jitter:local, jitter:RAP, jitter:DDP, shimmer:APQ3, shimmer:APQ5, shimmer:APQ11, and shimmer:DDA.

The rightmost column in Table IV represents the retained measurements for Wald's classifier after correlation and removal of statistically insignificant measurements.

C. Feature selection and classification

After pre-processing by removing statistically insignificant and highly correlated measures, Fig. 5 shows distributions estimated by using the Gaussian kernel density method for all of the 19 representative measures that have passed the significance and pre-selection test. The articulation and prosody measures show more distinction between the modes of

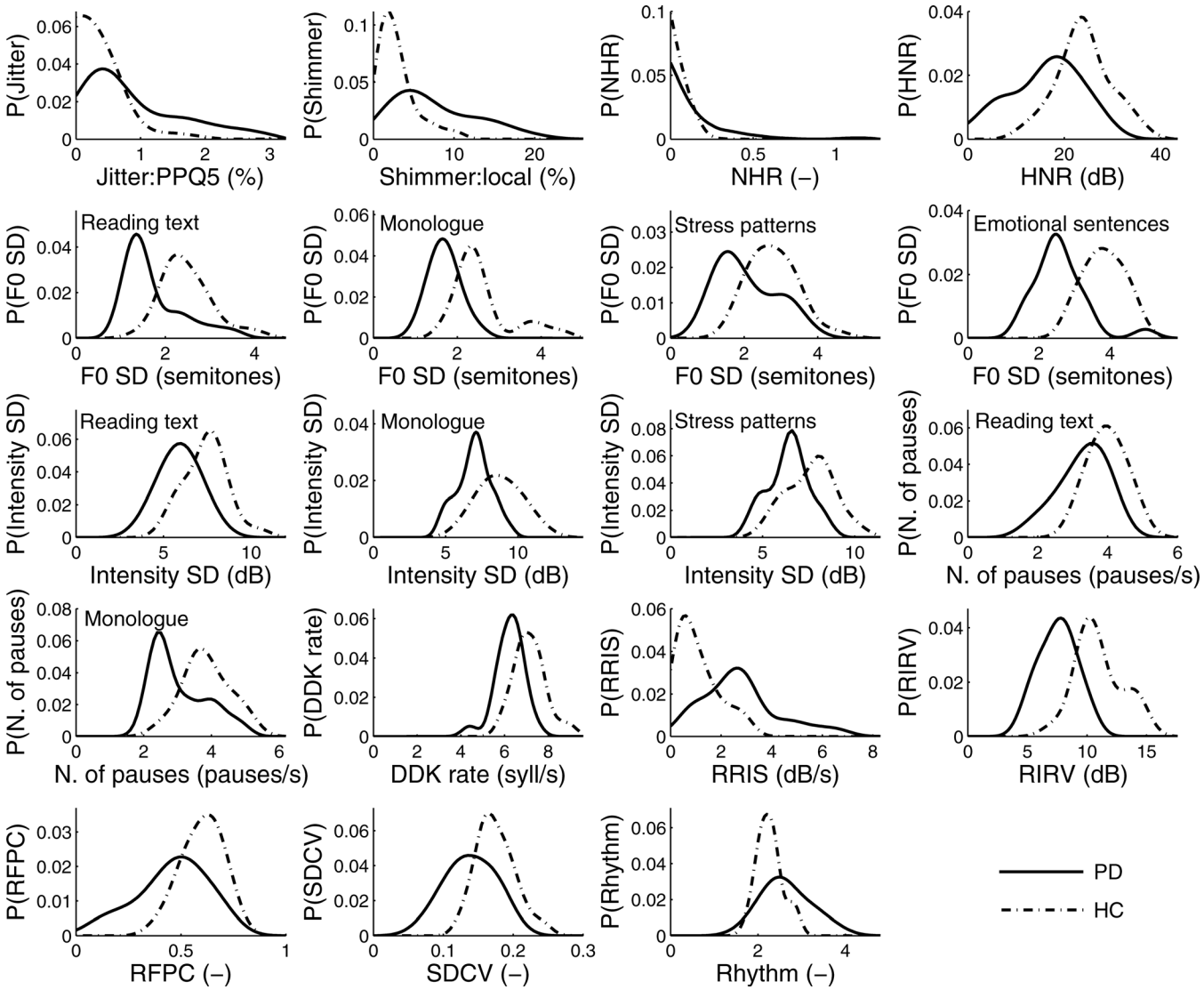


FIG. 5. Selected probability densities of all representative measures (features) in preparation for Wald task classification. The vertical axes are the probability densities P("measure") of the normalized features values of each measure. The dashdot lines are for HC speakers, the solid lines for PD subjects.

the values for both groups, whereas the modes of phonation measures are not as well separated. The visual inspection of the layout of each pair of these measures indicates that the optimal decision separating HC from PD may not be a simple intersection between two distributions. Thus, the Wald task provides a greater opportunity for not deciding to classify the subject as HC or PD, instead of forcing classification into one of the groups. This strategy essentially increases the overall effectiveness of classification in finding the signatures of specific voice and speech impairment.

Table V details the results of classification. From a total sum of 874 points, 234 were classified according to their group while only 11 were classified to the inverse group (26.77% vs 1.26%), which signifies that they achieved the speech performance of the inverses group in the selected task. The 629 remaining points (71.97%) were classified as an indecisive situation. Two of the 23 patients with PD reached performance as healthy people (8.7%), and none of 23 HC was classified as PD. The Wald task classifier was confirmed reliably to find the signs of vocal impairment or healthy voices according to the subject's speech performance. The F0 SD measurements in monolog and emotional sentences carry the greatest amount of information for separating HC from PD patients with classification performance of 60.87% (28 decisions). The RIRV in DDK task was the third best assessment method and gained 50% performance (23 decisions). The lowest scores in determining both of the groups were found in the measurement of jitter and F0 SD determined from stress patterns (10.87%, five decisions). The accuracy of the remaining representative measures ranged between 15% and 37% (7–17 decisions), which is why we have not listed detailed results. The correlation between the speech therapist and the classification result was 79.32% ($r = 0.7932$, $p = 0.4948 \times 10^{-11}$) and thus complemented the correctness of the acoustic measures and validated classification performance.

Finally, we need to decide the number of the signs that could characterize some form of vocal impairment in patients with PD. Considering all of the 32 performed measures (including the measures removed in the pre-selection stage) and demanding the significance probability of 5% from a correctly classified subject with PD, we find that all

TABLE V. List of Wald task classification values for all pre-selected measures and values of speech therapist evaluation. See main text for detailed description of the algorithm used to calculate these results. The total sum of the overall performance value represents the speech performance of each subject: More positive points are associated with higher progression of PD vocal impairment; the negative points show healthy speech performance. The total sum of ratings represents the suitability of the measure in differentiating PD patients from HC. The speech dimension was designated as affected when it reached at least two points of voice impairment assessment. The speech therapist evaluation represents rating scores of full vocal assessment: 1 point = intact speech performance, 7 points = progressed speech impairment.

	PHONATION				ARTICULATION					PROSODY												
Subject code	Jitter:PPQ5	Shimmer:local	NHR	HNR	RRIS	RIRV	SDCV	RFPC	DDK rate	F0 SD reading text	F0 SD monolog	F0 SD stress patterns	F0 SD emotional sentences	Intensity SD reading text	Intensity SD monolog	Intensity SD stress patterns	No. pauses reading text	No. pauses monolog	Rhythm	Affected speech subsystems	∑ Overall performance	Speech therapist evaluation
PD02	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	0	1	**PH+AR+PR**	**15**	**5**
PD03	0	1	1	1	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0	**PH+PR**	**7**	**4**
PD04	0	0	1	1	0	1	1	0	1	1	0	0	0	1	1	1	0	0	0	**PH+AR+PR**	**9**	**3**
PD05	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	None	**1**	**2**
PD06	0	1	1	1	0	1	0	1	0	1	0	0	1	0	0	0	0	0	0	**PH+AR+PR**	**7**	**4**
PD08	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1	**AR**	**4**	**3**
PD09	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	**PR**	**3**	**3**
PD10	0	1	0	1	0	0	0	0	1	0	1	0	1	0	0	0	1	1	0	**PH+PR**	**7**	**6**
PD11	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	**AR**	**3**	**3**
PD12	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	**PR**	**2**	**5**
PD13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−1	None	**−1**	**2**
PD14	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	1	**AR+PR**	**5**	**5**
PD15	0	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1	**PR**	**6**	**5**
PD16	0	0	0	0	0	1	0	0	1	0	1	0	1	0	1	0	0	−1	0	**AR+PR**	**4**	**2**
PD17	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	1	1	**PR**	**6**	**4**
PD18	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	**AR**	**3**	**3**
PD19	1	1	1	1	1	0	0	0	0	1	1	0	1	0	0	0	1	0	0	**PH+PR**	**9**	**4**
PD20	0	0	0	0	0	1	0	0	0	−1	1	0	1	1	1	1	1	0	0	**PR**	**6**	**4**
PD21	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	**AR**	**3**	**5**
PD22	0	0	−1	0	0	0	0	−1	0	0	0	0	1	0	0	0	0	0	−1	None	**−2**	**3**
PD23	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	None	**1**	**2**
PD25	0	0	0	0	0	0	0	0	1	0	−1	0	0	0	0	0	0	0	0	None	**0**	**3**
PD26	0	0	−1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	**PR**	**2**	**4**
HC02	0	−1	−1	−1	0	0	0	0	0	0	−1	0	−1	0	0	0	0	0	−1	None	**−6**	**2**
HC03	0	0	0	0	0	0	0	0	−1	−1	0	0	0	0	−1	0	0	0	0	None	**−3**	**2**
HC04	0	0	0	0	0	0	0	−1	−1	0	−1	0	−1	−1	−1	0	−1	0	0	None	**−7**	**1**
HC05	0	0	0	0	0	0	0	0	1	0	−1	0	−1	0	0	0	0	0	0	None	**−1**	**2**
HC06	−1	−1	−1	−1	0	−1	0	0	0	0	−1	0	−1	−1	−1	−1	0	0	−1	None	**−11**	**1**
HC07	−1	0	−1	−1	−1	0	−1	0	0	0	−1	0	−1	0	1	0	0	−1	0	None	**−7**	**1**
HC08	0	0	0	0	−1	0	0	0	−1	−1	0	0	0	0	−1	0	0	0	0	None	**−4**	**1**
HC09	0	−1	0	0	0	−1	−1	−1	−1	0	−1	−1	0	0	0	−1	−1	0	−1	None	**−10**	**1**
HC10	0	0	0	0	0	−1	0	0	0	−1	−1	0	−1	0	0	0	−1	0	0	None	**−5**	**2**
HC11	0	0	0	0	0	−1	−1	−1	−1	0	0	0	0	0	0	0	0	0	0	None	**−4**	**1**
HC12	−1	−1	−1	−1	−1	−1	0	0	−1	0	−1	0	0	0	0	0	−1	−1	0	None	**−10**	**1**
HC13	0	0	0	0	0	−1	0	0	−1	0	−1	0	−1	−1	−1	0	0	0	0	None	**−6**	**1**
HC14	0	−1	0	0	0	1	0	0	0	0	−1	0	0	0	0	0	0	0	0	None	**−1**	**2**
HC15	0	0	−1	0	−1	−1	0	0	0	0	−1	0	−1	0	0	0	0	0	0	None	**−5**	**2**
HC17	0	0	0	0	0	−1	0	0	0	0	0	−1	−1	−1	0	−1	0	0	−1	None	**−6**	**3**
HC18	0	0	0	0	−1	0	0	0	−1	0	0	0	−1	0	0	0	−1	0	−1	None	**−5**	**2**
HC19	0	−1	0	0	0	−1	0	0	−1	−1	−1	0	−1	−1	0	0	0	0	0	None	**−7**	**2**
HC20	0	0	0	0	0	−1	0	0	0	0	−1	0	0	0	0	0	0	0	0	None	**−2**	**2**
HC21	0	−1	−1	−1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	None	**−4**	**2**
HC22	0	−1	−1	0	−1	0	0	0	0	0	−1	−1	−1	−1	0	0	0	0	−1	None	**−8**	**2**
HC23	0	0	0	0	0	0	0	−1	0	0	0	0	−1	0	0	0	0	0	0	None	**−2**	**1**
HC24	0	−1	0	0	0	−1	0	0	0	−1	0	−1	−1	0	0	0	0	0	0	None	**−5**	**2**
HC25	0	0	0	0	0	0	0	0	0	−1	0	0	−1	0	0	0	−1	−1	0	None	**−4**	**1**
∑ |Ratings|	**5**	**14**	**14**	**11**	**12**	**23**	**7**	**9**	**17**	**11**	**28**	**5**	**28**	**13**	**10**	**7**	**12**	**6**	**13**			

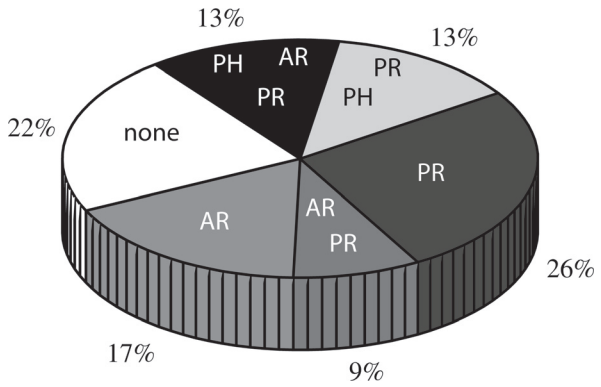


FIG. 6. Details of affected speech dimensions in PD: PH, phonation; AR, articulation; PR, prosody.

PD patients with a final score higher than or equal to 2 exhibit some form of speech impairment.

When each speech characteristic including phonation, articulation, and prosody is taken separately, we can see that the patterns of speech performance are spread through all speech dimensions only in the HC group (see Table IV). As can be seen in Fig. 6, the vocal impairment in early stage of the PD in the view of all speech dimensions is rather individual. From 23 people with PD, 18 are affected (78.26%). Considering that 2 signs are enough to determine the speech impairment, we found phonatory deficits in 6 cases (26.09%), lower ability of articulation in 9 cases (39.13%), and 14 cases of certain problems with prosody (60.87%) in PD patients. Deficits in all speech characteristics were found only in three people with PD (13.04%). Six PD subjects show deficits only in prosody (26.09%), four PD patients only in articulation (17.39%), none in phonation, and five in some combination of two speech characteristics (21.74%). There is also a need to take into account that the speech measurements can be partially interconnected in all speech dimensions. Hence, the speech impairment in early stages of PD might be considered as the total of speech defects in various speech characteristics.

Underscored by statistical decision-making theory, validated by a speech therapist, we propose that at least 78% of PD subjects in the early stage of their disease indicate symptoms of vocal impairment, prior to intervention through medical or speech therapy treatment.

IV. CONCLUSION

Our main finding is that 78% of early untreated PD subjects show some form of vocal impairment. This study concentrates on three speech subsystems including phonation, articulation, and prosody. It is important to note that the PD disorders of the individual subsystems not only influence each other but also frequently overlap. Disturbances of respiration and phonation consequently reflect, in particular, disruptions in speech prosody and partially articulation. Although a number of researchers have found that the most salient features of PD speech were related to phonatory impairment, with articulation being the second most affected subsystem,[10,26,27] in the case of early untreated PD, prosody of speech appears to be the most often damaged speech subsystem of the hypokinetic dysarthria. The specific PD voice and speech defects were found to differ individually in various characteristics (see Fig. 6). These results also show that persons with early untreated PD need not to have such a demonstrably impaired voice as to differ from the speech production of the wider norm of healthy people.

We also find that from the 19 representative measures, the variations of fundamental frequency in monolog and emotional sentences contain very useful information in separating HC from PD. The other representative measurements achieve sufficient accuracy, with the sole exception of jitter. In addition, the knowledge of incomplete vocal fold closure, lack of lung pressure, and lower articulation accuracy as a consequence of difficult articulation of fast syllable repetition in the DDK task lead to the design of new articulation measures that are gaining significance for increasing performance.

Taking into account that the number of participants in this study is 46, half of whom are patients with PD; we can therefore consider that the rate of participants is low. Despite this circumstance, we can expect that the probability distributions estimated by the Gaussian kernel density method in combination with the Wald task classification can ensure sufficient accuracy of the results. This is because we do not assume that there will be essential changes in the shape of distribution curves in the case of increasing the number of subjects.

We believe the automatic measurement methods and new measures of articulation will be useful in assessment of vocal impairment and will have a potential for positive feedback in speech treatment. The classifier based on the Wald task may also gain value in vocal impairment assessment and could be helpful in additionally enlarging the number of participants and efficiency measures. We also believe that the selected measures should be useful in most countries worldwide because of their independence from language.

It is necessary to stress that our patient sample is unique and cannot be compared with that of other authors who have published results from patients undergoing pharmaceutical treatment.

Future research could further test these findings in practice, and the speech measurement methods could involve the improvement of the individual voice in terms of treatment and motivation for therapy.

ACKNOWLEDGMENT

We are obliged to doctors Jan Roth, Jiri Klempir, Veronika Majerova, and Jana Picmausova for provision of clinical data. The study was partly supported by Czech Ministry of Education, projects MSM 6840770012, MSM 0021620849, MSM 0021620806, MSM 2C06031, Czech Science Foundation, project GACR 102/08/H008, Czech Ministry of Health, project IGA NR9220, Grant Agency of the Czech Technical University in Prague, project SGS 10/180/OHK3/2T/13, and by Creative Connection Ltd.

[1]O. Hornykiewicz, "Biochemical aspects of Parkinson's disease," Neurol. Suppl. **51**(2), S2–S9 (1998).
[2]M. C. de Rijk, L. J. Launer, K. Berger, M. M. B. Breteler, J.-F. Dartigues, M. Baldereschi, L. Fratiglioni, A. Lobo, J. Martinez-Lage, C. Trenkwalder, and A. Hofman, "Prelevance of Parkinson's disease in Europe: A collaborative study of population-based cohorts," Neurology **54**, 21–23 (2000).
[3]A. Lang and A. Lozano, "Parkinson's disease: First of two parts," New Engl. J. Med. **339**, 1044–1053 (1998).

[4]M. Hoehn, "The natural history of Parkinson's disease in the pre-levodopa and post-levodopa eras," Neurol. Clin. **10**, 331–339 (1992).
[5]M. C. de Rijk, C. Tzourio, M. M. Breteler, J. F. Dartiques, L. Amaducci, S. Lopez-Pousa, J. M. Manubens-Bertran, A. Alpérovitch, and W. A. Rocca, "Prevalence of parkinsonism and Parkinson's disease in Europe: The EUROPARKINSON Collaborative Study. European Community Concerted Action on the Epidemiology of Parkinson's disease," J. Neurol. Neurosurg. Psychiatry **62**, 10–15 (1997).
[6]S. K. Van Den Eeden, C. M. Tanner, A. L. Bernstein, R. D. Fross, A. Leimpeter, D. A. Bloch, and L. M. Nelson, "Incidence of Parkinson's disease: Variation by age, gender, and race/ethnicity," Am. J. Epidemiol. **157**, 1015–1022, (2003).
[7]G. J. Canter, "Speech characteristic of patients with Parkinson's disease. I. Intensity, pitch, and duration," J. Speech Hear. Disord. **28**, 221–229 (1963).
[8]F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria," J. Speech. Hear. Res. **12**, 426–496 (1969).
[9]A. K. Ho, R. Iansek, C. Marigliani, J. Bradshaw, and S. Gates, "Speech impairment in large sample of patients with Parkinson's disease," Behav. Neurol. **11**, 131–137 (1998).
[10]J. A. Logemann, H. B. Fisher, B. Boshes, and E. R. Blonsky, "Frequency and coocurrence of vocal tract dysfunction in the speech of a large sample of Parkinson patients," J. Speech Hear. Disord. **43**, 47–57 (1978).
[11]J. R. Duffy, *Motor Speech Disorders: Substrates, Differential Diagnosis and Management*, 2nd ed. (Mosby, New York, NY, 2005), pp. 1–592.
[12]B. Harel, M. Cannizzaro, and P. J. Snyder, "Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study," Brain Cogn. **56**, 24–29 (2004).
[13]N. Singh, V. Pillay, and Y. E. Choonara, "Advances in the treatment of Parkinson's disease," Prog. Neurobiol. **81**, 29–44 (2007).
[14]J. King, L. Ramig, J. H. Lemke, and Y. Horii, "Parkinson's disease: Longitudinal changes in acoustic parameters of phonation," J. Med. Speech Lang. Pathol. **2**, 29–42 (1994).
[15]M. Trail, C. Fox, L. O. Ramig, S. Sapir, J. Howard, and E. C. Lai, "Speech treatment for Parkinson's disease," Neurorehabilitation **20**, 205–221 (2005).
[16]A. M. Goberman and M. Blomgren, "Parkinsonian speech disfluencies: Effects of L-dopa-related fluctuation," J. Fluency Disord. **28**, 55–70 (2003).
[17]M. Putzer, W. J. Barry, and J. R. Moringlane, "Effect of bilateral stimulation of the subthalamic nucleus on different speech subsystems in patients with Parkinson's disease," Clin. Linguist. Phonetics **22**, 957–973 (2008).
[18]A. K. Ho, J. L. Bradshaw, and R. Iansek, "For better or worse: The effect of levodopa on speech in Parkinson's disease," Mov. Disord. **23**, 574–580 (2007).
[19]M. Oxtoby, *Parkinson's Disease People and Their Social Needs* (Parkinson's Disease Society, London, United Kingdom, 1982), pp. 1–75.
[20]S. Sapir, L. O. Ramig, P. Hoyt, S. Countryman, C. O'Brien, and M. Hoehn, "Speech loudness and quality 12 months after intensive voice treatment (LSVT) for Parkinson's disease: A comparison with an alternative speech treatment," Folia Phoniatr. Logop. **54**, 296–303 (2002).
[21]D. M. Huse, K. Schulman, L. Orsini, J. Castelli-Haley, S. Kennedy, and G. Lenhart, "Burden of illness in Parkinson's disease," Mov. Disord. **20**, 1449–1454 (2005).
[22]B. T. Harel, M. S. Cannizaro, H. Cohen, N. Reilly, and P. J. Snyder, "Acoustic characteristic of Parkinsonian speech: A potential biomarker of early disease progression and treatment," J. Neurolinguistics **17**, 439–453 (2004).
[23]M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurement for telemonitoring of Parkinson's disease," IEEE Trans. Biomed. Eng. **56**, 1015–1022 (2009).
[24]C. Ruggiero, R. Sacile, and M. Giacomini, "Home telecare," J. Telemed. Telecare **5**, 11–17 (1999).
[25]F. L. Darley, A. E. Aronson, J. R. Brown, *Motor Speech Disorders* (Saunders, Philadelphia, PA, 1975), pp. 1–305.
[26]C. Ludlow, N. Connor, and C. Bassich, "Speech timing in Parkinson's and Huntingston's disease," Brain Lang. **32**, 195–214 (1987).
[27]P. Zwirner, and G. Barnes, "Vocal tract steadiness: A measure of phonatory stability and upper airway motor control during phonation in dysarthria," J. Speech Hear. Res. **35**, 761–768 (1992).
[28]A. M. Goberman and C. Coelho, "Acoustic analysis of Parkinsonian speech I: Speech characteristics and L-dopa therapy," Neurorehabilitation **17**, 237–246 (2002).
[29]R. J. Holmes, J. M. Oates, D. J. Phyland, and A. J. Hughes, "Voice characteristics in the progression of Parkinson's disease," Int. J. Lang. Commun. Disord. **35**, 407–418 (2000).
[30]S. Skodda, and U. Schlegel, "Speech rate and rhythm in Parkinson's disease," Mov. Disord. **23**, 985–992 (2008).
[31]P. J. Watson and B. Munson, "Parkinson's disease and the effect of lexical factors on vowel articulation," J. Acoust. Soc. Am. **124**, 291–295 (2008).
[32]J. I. Godino-Llorente and P. Gomez-Vilda, "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors," IEEE Trans. Biomed. Eng. **51**, 380–384 (2004).
[33]P. Henriquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, J. I. Godino-Llorente, and F. Diaz-de-Maria, "Characterization of healthy and pathological voice through measures based on nonlinear dynamics," IEEE Trans. Audio, Speech, Lang. Process. **17**, 1186–1195 (2009).
[34]D. A. Rahn, M. Chou, J. J. Jiang, and Y. Zhang, "Phonatory impairment in Parkinson's disease: Evidence from nonlinear dynamics analysis and perturbation analysis," J. Voice **21**, 64–71 (2007).
[35]M. I. Schlesinger and V. Hlavac, "Ten lectures on statistical and structural pattern recognition," in *Volume 24 of Computational Imaging and Vision* (Kluwer Academic Press, Dordrecht, 2002), pp. 1–544.
[36]M. Hoehn and M. Yahr, "Parkinsonism: Onset, progression and mortality," Neurology **17**, 427–442 (1967).
[37]G. Stebbing and C. Goetz, "Factor structure of the Unified Parkinson's Disease Rating Scale: Motor Examination Section," Mov. Disord. **13**, 633–636 (1998).
[38]L. Ramig, S. Countryman, L. Thompson, and Y. Horii, "Comparison of two forms of intensive treatment for Parkinson disease," J. Speech Hear. Res. **38**, 1232–1251 (1995).
[39]A. M. Goberman, "Correlation between acoustic speech characteristics and non-speech motor performance in Parkinson disease," Med. Sci. Monit. **11**, 109–116 (2005).
[40]R. J. Baken, and R. F. Orlikoff, *Clinical Measurement of Speech and Voice*, 2nd ed. (Singular Thomson Learning, San Diego, CA, 2000), pp. 1–610.
[41]L. Lisker and A. Abrason, "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–442 (1964).
[42]K. Forrest, G. Weismer, and G. Turner, "Kinematic, acoustic and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric males," J. Acoust. Soc. Am. **85**, 2608–2622 (1989).
[43]G. Weismer, "Articulatory characteristics of Parkinsonian dysarthria: Segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination," in *The Dysarthrias: Physiology, Acoustics, Perception, Management*, edited by M. Mcneil, J. Rosenbeck, and A. Aronson (College-Hill Press, San Diego, CA, 1984), pp. 101–130.
[44]J. Logemann and H. Fisher, "Vocal tract control in Parkinson's disease: Phonetics feature analysis of misartriculation," J. Speech. Hear. Disord. **46**, 348–352 (1981).
[45]S. Fletcher, "Time-by-count measurement of diadochokinetic syllable rate," J. Speech. Hear. Disord. **15**, 757–762 (1972).
[46]L. D'Alatri, G. Paludett, M. F. Contarino, S. Galla, M. R. Marchese, and A. R. Bentivoglio, "Effects of bilateral subthalamic nucleus stimulation and medication on Parkinsonian speech impairment," J. Voice **22**, 365–372 (2008).
[47]G. Weismer, J.-Y. Jeg, J. S. Laures, R. D. Kent, and J. F. Kent, "Acoustic and intelligibility characteristic of sentence production in neurogenic speech disorders," Folia Phoniatr. Logop. **53**, 1–18 (2001).
[48]K. Tjaden, "Anticipatory coarculation in multiple sclerosis and Parkinson's disease," J. Speech Lang. Hear. Res. **46**, 990–1008 (2003).
[49]J. Metter and W. Hanson, "Clinical and acoustical variability in hypokinetic dysarthria," J. Commun. Disord. **19**, 347–366 (1986).
[50]A. P. Simpson, "Phonetic differences between male and female speech," Lang. Linguist. Compass **3/2**, 621–640 (2009).
[51]B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 5th ed. (Academic Press, Boston, 2003), pp. 1–413.
[52]J. Mobes, G. Joppich, F. Stiebritz, R. Dengler, and Ch. Schroder, "Emotional speech in Parkinson's disease," Mov. Disord. **23**, 824–829 (2008).
[53]H. Boril and P. Pollak, "Direct time domain fundamental estimation of speech in noisy conditions," in *EUROSIPCO Proceedings*, Vienna, Austria (2004), pp. 1003–1006.
[54]B. Kotnik, H. Hoge, Z. Kacic, "Noise robust F0 determination and epoch-marking algorithms," Signal Process. **89**, 2555–2569 (2009).
[55]X. Wei, L. Zhao, Q. Zhang, J. Dong, "Robust pitch estimation using a wavelet variance analysis model," Signal Process. **89**, 1216–1223 (2009).
[56]P. Boersma and D. Weenink, "PRAAT, a System for Doing Phonetics by Computer," Glot Int. **5**, 341–345 (2001).
[57]A. Novak and J. Vokral, "Acoustic parameters for the evaluation of voice of future voice professionals," Folia Phoniatr. Logop. **47**, 279–285 (1995).

[58]P. Boersma, "Accurate short-term analysis of the fundamental frequency and harmonics-to-noise ratio of a sampled sound," in *Proceedings of the Institute of Phonetics Sciences*, Amsterdam, The Netherlands, vol. 17 (1993), pp. 97–110.
[59]J. Rusz and R. Cmejla, "Speech rate and intensity analysis in Parkinson's disease," (in Czech), Akusticke listy **14**, 13–17 (2008).
[60]C. Ozsancak, P. Auzou, M. Jan, and D. Hannequin, "Measurements of voice onset time in dysarthric patients: Methodological consideration," Folia Phoniatr. Logop. **53**, 48–57 (2001).
[61]E. Fisher and A. M. Goberman, "Voice onset time in Parkinson disease," J. Commun. Disord. **43**, 21–34 (2010).
[62]K. Mustafa and I. C. Bruce, "Robust formant tracking for continuous speech with speaker variability," IEEE Trans. Audio, Speech, Lang. Process. **14**, 435–444 (2006).
[63]H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," IEEE Trans. Acoust. Speech Signal Process. **26**, 43–49 (1978).
[64]D. Duez, "Acoustic analysis of occlusive weakening in Parkinsonian French speech," in *Processing of International Congress of Phonetic Sciences*, Saarbrucken, Germany (2007), pp. 1973–1976.
[65]P. W. Holland, and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," Commun. Stat: Theory Meth. **42**, 152–154 (1988).
[66]R. Cmejla and P. Sovka, "Recursive Bayesian autoregressive changepoint detector for sequential signal segmentation," in *EUROSIPCO Proceedings*, Wien (2004), pp. 245–248.
[67]J. J. K. O. Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing* (Springer-Verlag, New York, NY, 1996), pp. 1–254.
[68]P. Bergl and R. Cmejla, "Improved detection of phonemes in speech databases," in *Proceedings of the Fifth IAESTED International Conference on Biomedical Engineering* (Innsbruck, 2007), pp. 171–174.
[69]A. Wald, *Sequential Analysis* (Wiley, New York, NY, 1947), pp. 1–209.
[70]J. Neyman and E. S. Pearson, "On the use and interpretation of certain test criteria for purposes of statistical inference," Biometrika **20A**, 175–240, 263–294 (1928).
[71]A. W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis* (Oxford University Press, Oxford, United Kingdom, 1997), pp. 1–208.

Appendix A2

Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test

Rusz J, Cmejla R, Ruzickova H, Klempir J, Majerova V, Picmausova J, Roth J, Ruzicka E (2011) Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test. *Mov Disord* **26**:1951-1952.

Journal name (category): **Movement Disorders** (clinical neurology)
Quartile in category: **Q1**
Impact factor: **4.480**
Number of WoS citations (excluding autocitations): **10**

Acoustic Assessment of Voice and Speech Disorders in Parkinson's Disease Through Quick Vocal Test

The disorders of voice and speech in Parkinson's disease (PD) result from involvements in several subsystems including respiration, phonation, articulation, and prosody.[1–3]

We investigated the feasibility of acoustic measures for the identification of voice and speech disorders in PD, using a quick vocal test consisting of sustained phonation, diadochokinetic task, and running speech. Various traditional and novel acoustic measurements have been designed in order to be gender independent, represent all speech subsystems, reduce the time required for voice investigation, and provide a reliable automated assessment in practice.[4]

Patients and Methods

A total of 46 Czech native participants were recruited. Twenty-four of them fulfilled the diagnostic criteria for PD and were examined before the symptomatic treatment was started: 20 men, 4 women; mean age (± SD), 60.9 ± 11.2 years; duration of PD symptoms, 31.3 ± 22.3 months (range, 6–84 months); H&Y stage, 2.2 ± 0.5 (range, 1–3); and UPDRS motor score, 17.4 ± 7.1 (range, 5–32); with UPDRS speech item = 0 in 13 patients and speech item = 1 in 11 patients. As a healthy control (HC) group, 22 persons with no history of neurological or communication disorders were included: 15 men, 7 women; mean age, 58.7 ± 14.6 years. Age distribution did not differ significantly between the groups.

Each participant was instructed to perform 3 vocal tasks: [VT1], sustained phonation at a comfortable pitch and loudness as constant and long as possible, at least 5 seconds on 1 breath; [VT2], diadochokinetic (DDK) task requiring rapid, steady /pa/-/ta/-/ka/ syllable repetition as constant and long as possible, repeated at least 5 times on 1 breath; and [VT3], running speech for approximately 80 seconds. For reproducibility of data, each task was repeated at least 2 times for every subject.

The extracted speech parameters were assessed using measures of *phonation* [VT1] including jitter, shimmer, noise-to-harmonics ratio (NHR), and harmonics-to-noise ratio (HNR)[5]; *respiration* [VT2] including sound pressure level decline (SPLD)[4]; *articulation* [VT2] including robust formant periodicity correlation (RFPC), and spectral distance change variation (SDCV)[4]; and *prosody* [VT3] including voice fundamental frequency variations (F0 SD).[6] Supporting Information Table 1 details the measurements used.

For every subject, average values (speech performances) for each acoustic measurement were calculated. Two-sided Wilcoxon rank-sum and Spearman rank tests were performed to find differences between groups and within-group correlations. Subsequently, an exhaustive search of all possible measure combinations was performed, and a predictive model was built using a kernel support vector machine (SVM) to find the best combination of measurements to differentiate PD from HC subjects. Cross-validation with the leave-one-out method was used to validate reproducibility of the SVM classifier.[7]

Results

In total, 116 vocal recordings were collected and used for classification. Significant differences between the 2 groups were found in all 8 measurements. In addition, from all performed correlations, statistically significant relationships were found between several measures of articulation and phonation and subscores of bradykinesia and rigidity (Supporting Table 2). The best classification performance of 85.0% ± 6.1% was reached in a combination of 4 measures that represent all PD-related affected speech subsystems, including the impaired ability to maintain sound pressure level (SPLD), increased noise components during phonation (NHR), lowered accuracy of articulation (RFPC), and reduced melody of speech (F0 SD); see Figure 1. The maximal classification accuracy using simple task was 81.3% ± 6.9% for running speech, 75.6% ± 8.3% for sustained phonation, and 71.4% ± 8.3% for DDK task; therefore, reduced melody in running speech appeared essential in characterizing the vocal impairment in PD.

Discussion

We have designed a quick 2-minute vocal test and investigated the potential of using acoustic analysis in detecting voice and speech disorders in PD. The method demonstrated that it can accurately differentiate PD patients from HCs. This could be of high clinical relevance as subtle abnormalities such as reduced melody in running speech were detectable from the early stage of PD. Admittedly, the study has certain limitations. Although the uneven gender representation of patients and controls could be offset by gender independence of designed acoustic measurement methods, our sample size remains rather small. Should our results be confirmed on a larger population sample, voice and speech disorders might be considered as early markers of the disease, and acoustic analysis might serve as a simple screening test in view of the expected advent of neuroprotective treatment. In a more

*Correspondence to: Evžen Růžička, Department of Neurology, Charles University in Prague, First Faculty of Medicine and General University Hospital in Prague, Katerinska 30, Praha 2, Czech Republic; eruzi@lf1.cuni.cz

Relevant conflicts of interest/financial disclosures: Nothing to report. The study was partly supported by the Czech Science Foundation, project GACR 102/08/H008, Czech Ministry of Health, project NT 11331-6/2010, Grant Agency of the Czech Technical University in Prague, project SGS 10/180/OHK3/2T/13, and Czech Ministry of Education, projects MSM 0021620849, MSM 6840770012, and MSM 0021620806.

Full financial disclosures and author roles may be found in the online version of this article.

Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.23680

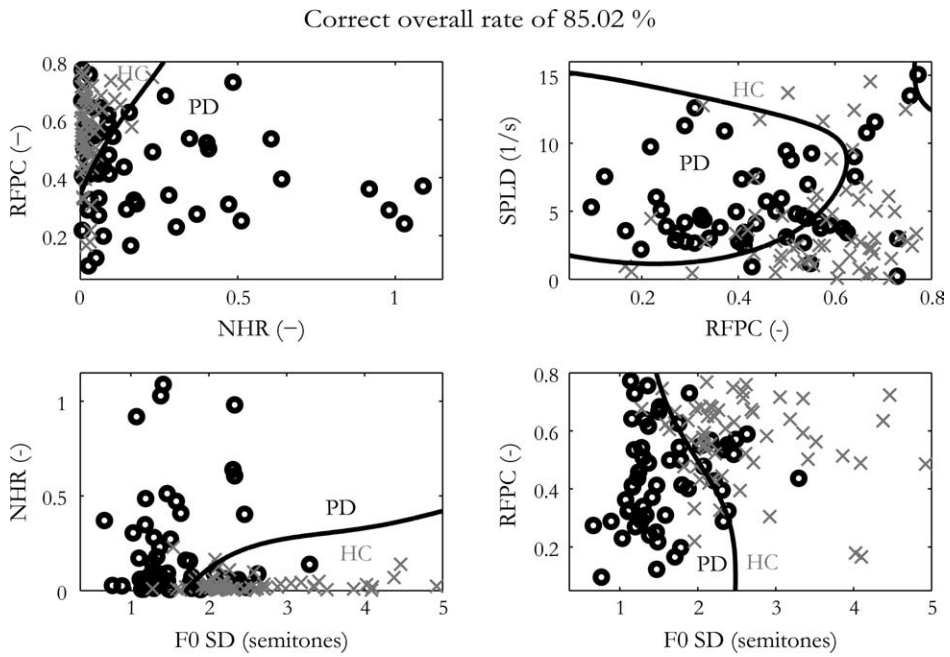


FIG. 1. The results of the SVM-based classifier for selected pairs of the measures combination with best classification accuracy. The "o" marks are for PD, the "x" marks are for HC, and the dark gray curves represent the SVM classification boundaries between both groups.

modest scope, the use of automated acoustic vocal tests can ease the clinical monitoring of voice and speech disorders progression as well as the effects of medication on speech production and can serve as feedback in voice treatment.

Jan Rusz, MSc,[1,2,3] Roman Čmejla, MSc, PhD,[1] Hana Růžičková, MSc,[3] Jiří Klempíř, MD, PhD,[3] Veronika Majerová, MD,[3] Jana Picmausová, MD,[3] Jan Roth, MD, PhD,[3] and Evžen Růžička, MD, DSc[3*]

[1]Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Circuit Theory, Prague, Czech Republic; [2]Charles University in Prague, First Faculty of Medicine, Department of Pathophysiology, Prague, Czech Republic; and [3]Charles University in Prague, First Faculty of Medicine, Department of Neurology and Centre of Clinical Neuroscience, Prague, Czech Republic

References

1. Darley FL, Aronson AE, Brown JR. Differential diagnostic patterns of dysarthria. J Speech Hear Res 1969;12:246–269.
2. Goberman AM, Coelho C. Acoustic analysis of Parkinsonian speech I: Speech characteristics and L-dopa therapy. Neurorehab 2002;17: 237–246.
3. Skodda S, Schlegel U. Speech rate and rhythm in Parkinson's disease. Mov Disord 2008;23:985–992.
4. Rusz J, Cmejla R, Ruzickova H, Ruzicka E. Quantitative acoustic measurements for characterization of voice and speech disorders in early untreated Parkinson's disease. J Acoust Soc Am 2011;129:350–367.
5. Boersma P, Weenink D. Praat, a system for doing phonetics by computer. Glot Int 2001;5:341–345.
6. Boril H, Pollak P. Direct time domain fundamental estimation of speech in noisy conditions. In: EUSIPCO Proceedings, Vienna, Austria; 2004.
7. Franc V, Hlavac V. Statistical pattern recognition toolbox for Matlab. Prague, Czech Republic: Czech Technical University, Faculty of Electrical Engineering, Center for Machine Perception; 2004. Available at http://cmp.felk.cvut.cz/. Accessed January 15, 2011.

Appendix A3

Imprecise vowel articulation as potential early marker of Parkinson's disease: effect of speaking task

Rusz J, Cmejla R, Tykalova T, Ruzickova H, Klempir J, Majerova V, Picmausova J, Roth J, Ruzicka E (2013) Imprecise vowel articulation as potential early marker of Parkinson's disease: effect of speaking task. *J Acoust Soc Am* **134**:2171-2181.

Journal name (category): **Journal of the Acoustical Society of America** (acoustics/audiology & speech-language pathology)
Quartile in category: **Q2/Q2**
Impact factor: **1.646**
Number of WoS citations (excluding autocitations): **25**

Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task

Jan Rusz,[a] Roman Cmejla, and Tereza Tykalova
Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Circuit Theory, Technicka 2, 166 27, Prague 6, Czech Republic

Hana Ruzickova, Jiri Klempir, Veronika Majerova, Jana Picmausova, Jan Roth, and Evzen Ruzicka
Charles University in Prague, First Faculty of Medicine, Department of Neurology and Centre of Clinical Neuroscience, Katerinska 30, 120 00, Prague 2, Czech Republic

(Received 16 October 2012; revised 24 June 2013; accepted 8 July 2013)

The purpose of this study was to analyze vowel articulation across various speaking tasks in a group of 20 early Parkinson's disease (PD) individuals prior to pharmacotherapy. Vowels were extracted from sustained phonation, sentence repetition, reading passage, and monologue. Acoustic analysis was based upon measures of the first (F1) and second (F2) formant of the vowels /a/, /i/, and /u/, vowel space area (VSA), F2i/F2u and vowel articulation index (VAI). Parkinsonian speakers manifested abnormalities in vowel articulation across F2u, VSA, F2i/F2u, and VAI in all speaking tasks except sustained phonation, compared to 15 age-matched healthy control participants. Findings suggest that sustained phonation is an inappropriate task to investigate vowel articulation in early PD. In contrast, monologue was the most sensitive in differentiating between controls and PD patients, with classification accuracy up to 80%. Measurements of vowel articulation were able to capture even minor abnormalities in speech of PD patients with no perceptible dysarthria. In conclusion, impaired vowel articulation may be considered as a possible early marker of PD. A certain type of speaking task can exert significant influence on vowel articulation. Specifically, complex tasks such as monologue are more likely to elicit articulatory deficits in parkinsonian speech, compared to other speaking tasks.
 [http://dx.doi.org/10.1121/1.4816541]

PACS number(s): 43.70.Dn [MAH] Pages: 2171–2181

I. INTRODUCTION

Parkinson's disease (PD) is a neurological illness characterized by the progressive loss of dopaminergic neurons and is associated with a variety of motor and non-motor deficits (Hornykiewicz, 1998). Prior research has demonstrated that approximately 70%–90% of people with PD possess a distinctive alteration of speech termed hypokinetic dysarthria (Logemann *et al.*, 1978), which is a multidimensional impairment affecting various aspects of speech such as respiration, phonation, articulation, and prosody (Darley *et al.*, 1969). Imprecise vowel articulation is a common deficit associated with dysarthria and contributes to reduced speech intelligibility (Kim *et al.*, 2011a). Impairment of vowel articulation, occurring as a consequence of reduced articulatory range of motion ("undershooting" of articulatory gestures) (Forrest *et al.*, 1989; Robertson and Hammerstadt, 1996), has been widely documented in PD (Sapir *et al.*, 2007; Sapir *et al.*, 2010; Skodda *et al.*, 2011). Previous findings of acoustic and kinematic studies support a reduced amplitude and velocity of articulators (lips, tongue, jaw) in parkinsonian speakers (Forrest *et al.*, 1989; Robertson and Hammerstadt, 1996), suggesting that articulation deficits reflect hypokinesia and rigidity of the vocal tract (Forrest *et al.*, 1989; Hunker *et al.*, 1982).

Although imprecise vowel articulation has been observed even in mild PD (Skodda *et al.*, 2011), previous studies have mainly focused on moderate or more advanced stages. As dysarthria can exert significant influence on speech performance in PD (Kim *et al.*, 2011b), one might expect that the extent of vowel articulation impairment is likely to reflect the severity of dysarthria. Moreover, the severity of dysarthria in PD is thought to be influenced by the severity of motor symptoms, disease duration, as well as specific effects of dopaminergic treatment (Goberman and Coelho, 2002; Schulz and Grant, 2000). Generally, findings related to vowel articulation in the course of PD are somewhat ambiguous. With respect to early stages, improvement of vowel articulation performance under dopaminergic therapy has been noted for several PD speakers (Skodda *et al.*, 2010; Rusz *et al.*, 2013). In contrast, recent research has revealed further decline of vowel articulation performance in PD throughout extended treatment periods (Skodda *et al.*, 2012). Considering that medical interventions as well as disease progression may affect speech performance in different ways, the examination of vowel articulation in early PD, before the onset of therapy, is essential to gain more insight into the development of parkinsonian speech disorders.

While deficits of vowel articulation are commonly present in PD speakers (Forrest *et al.*, 1989; Sapir *et al.*, 2007; Skodda *et al.*, 2011), little effort has been given to examine the severity of vowel articulation impairment under various speaking tasks. In treated patients with mild to moderate PD,

[a]Author to whom correspondence should be addressed. Electronic mail: ruszjan@fel.cvut.cz

imprecise vowel articulation has been found in the speaking task of sentence repetition (Sapir *et al.*, 2007; Sapir *et al.*, 2010) as well as reading passage (Skodda *et al.*, 2011). In addition, impaired vowel articulation has been observed in patients with severe PD while performing sustained prolongation of single vowels (Eliasova *et al.*, 2013). However, to the best of our knowledge, investigation of vowel articulation in PD has been primarily based on a single task and limited to simple utterances, and no evidence has been given regarding the sensitivity of vowel articulation under spontaneous speech. Additionally, there is a growing body of evidence that the signs of dysarthria vary across the specific type of speech task performed (Caligiuri, 1989; Rosen *et al.*, 2005). In particular, dysarthric speech performance has been found to be even more significantly altered acoustically during spontaneous speech production when compared to other non-spontaneous tasks (Kempler and Van Lacker, 2002). A study by Weismer (1984) has suggested that the degree of articulatory deviances seem to vary between simple versus complex utterances produced by speakers with parkinsonian dysarthria, which may be explained by the fact that simple speaking tasks do not require the subject's full attention and are likely to be more automatic than structured and complex tasks such as spontaneous speech. Based on these observations, in the present study we endeavored to determine if a certain type of speaking task is more sensitive to disturbed vowel articulation in PD.

The quality and intelligibility of each vowel can be determined primarily by the distinctive acoustic energy peak of the first (F1) and second (F2) formant frequencies. The F1 and F2 frequencies particularly reflect tongue position, with the acoustic-articulatory relationship defined such that the F1 frequency varies inversely with tongue height and the F2 frequency varies directly with tongue advancement (Kent *et al.*, 1999). Thus, limited articulatory range of motion due to PD may result in vowel formant centralization, i.e., formants with naturally higher frequencies tend toward lower frequencies, and formants with naturally lower frequencies tend toward higher frequencies (Kent and Kim, 2003; Sapir *et al.*, 2007). The overall reduction of working space for vowels in PD can be captured well by a reduced size of the vowel space area (VSA), which is constructed by the Euclidean distances between the F1 and F2 coordinates of the corner vowels /a/, /i/, and /u/ in the triangular F1-F2 vowel space (Kent and Kim, 2003), as compared to normal speech. Several studies have reported a relationship between the VSA and the perceptual impression of intelligibility in speakers with dysarthria (Liu *et al.*, 2005; Weismer *et al.*, 2001). Conversely, a study by Kim *et al.* (2011a) demonstrated that intelligibility in dysarthric speakers was better represented by the degree of overlap among vowels than by the vowel space. In fact, a report by Monses (1976) was the first to show correlation between speech intelligibility in speakers with severe hearing impairment and distance among the F2 frequencies of adjacent vowels in the vowel space. Similarly in PD, both the F2i/F2u ratio representing the distance between the vowels /i/ and /u/ and vowel articulation index (VAI) reflecting vowel centralization (considering all F1 and F2 frequencies across corner vowels) have been shown to be more sensitive in differentiating dysarthric speech from normal speech than the VSA (Roy *et al.*, 2009; Sapir *et al.*, 2007; Sapir *et al.*, 2010; Skodda *et al.*, 2011). Therefore, the current study further addresses the question of whether certain formant-based measurements are more sensitive than other measurements in capturing deficits of vowel articulation, and examined early-stage PD speakers as the main focus.

One potential application of the identification of changes in vowel articulation may be related to the early diagnosis of PD. Tetrud (1991) reported that family members and close friends of prospective patients with PD may note changes in speech several years before the diagnosis is established. More recently, Postuma *et al.* (2012) investigated prodromal Parkinsonism-related motor changes in idiopathic rapid eye movement (REM) sleep behavior disorder and revealed that voice and face akinesia were the earliest indicators of Parkinsonism with an estimated prodromal interval of 9.8 yr before diagnosis. However, speech disorders in the early stages of PD are often mild and have a limited effect on speech intelligibility, making them barely perceptible to others or even to the patients themselves (Stewart *et al.*, 1995). On the other hand, acoustic speech abnormalities have been reported even in PD patients with no perceptible dysarthria (Forrest and Weismer, 2009), and several investigators have found impaired speech parameters in early-stage PD using objective acoustic measures (Rusz *et al.*, 2011a; Stewart *et al.*, 1995). In clinical practice, movement disorder specialists (MDS) are responsible for making the early diagnosis of PD, with disability commonly evaluated using the Unified Parkinson's Disease Rating Scale (UPDRS) (Stebbing and Goetz, 1998). A global perceptual description of patient speech is part of the UPDRS III motor score (item 18) and its evaluation represents a part of daily practice for the MDS. As speech deterioration may be a prodromal feature of Parkinsonism and defects of vowel articulation are common findings in PD, the present study was designed to investigate whether changes in vowel articulation captured by objective acoustic analyses are superior to the perceptual impression of disturbed speech raised by the experienced MDS.

In summary, this study was designed to address following questions:

1. Can imprecise vowel articulation be considered as an early marker of PD? We hypothesized that early-stage, untreated PD patients can be differentiated from healthy speakers using objective measurements of vowel articulation.
2. Which speaking task, including sustained phonation, short sentence repetition, reading passage, and monologue is most sensitive to imprecise vowel articulation in PD? We hypothesized that (a) sustained phonation would not be sufficiently sensitive to differentiate healthy speakers from early-stage PD speakers, but (b) spontaneous speaking such as monologue would be more altered in PD speakers when compared to other speaking tasks.
3. Are some formant-based measurements, including F1 and F2 frequencies of each corner vowel (/a/, /i/, and /u/), VSA, F2i/F2u, and VAI more sensitive in capturing

deficits of vowel articulation in early-stage PD speakers in comparison to healthy speakers? We hypothesized that F2i/F2u and/or VAI would be superior to VSA in the description of mildly impaired parkinsonian speech.

4. Are objective measurements of vowel articulation superior to the clinical impression of disturbed speech as determined by the MDS? We hypothesized that changes of vowel articulation captured by objective acoustic analyses would uncover articulatory disorders in PD with greater precision than perceptual evaluation by the MDS.

II. METHODS

A. Participants

A total of 35 male, Czech native speakers were recruited for the study. Twenty subjects were diagnosed with *idiopathic PD*. They fulfilled the diagnostic criteria for PD (Hughes *et al.*, 1992) and were examined immediately after the diagnosis was established and before symptomatic treatment was started. Their age ranged from 34 to 83 yr (mean, 61.0; ± standard deviation (SD) 12.0), the Hoehn and Yahr (HY, Hoehn and Yahr, 1967, disability scale comprised of stages 1 through 5, where 5 is most severe) disease stage ranged from 1 to 3 (2.2 ± 0.5), the UPDRS III (Stebbing and Goetz, 1998, motor rating scaled from 0 to 108, where 108 represents severe motor impairment) ranged from 5 to 32 (17.9 ± 7.3), and the estimated duration of PD manifestations prior to examination ranged from 6 to 82 months (28.6 ± 19.9). None of the participants reported speech, language, or hearing disorders unrelated to their parkinsonian symptoms, nor had a history of speech-language treatment prior to participation in this study. All PD patients were free of depression and cognitive deficits that could interfere with measurements.

The *healthy control* (HC) group consisted of 15 persons of comparable age ranging from 36 to 80 yr (62.6 ± 13.4). None of these individuals reported a history of neurological disorders or other disorders that may affect speech, language, or hearing. Age distributions were not significantly different between the PD and HC groups. All participants provided their consent to the speaking tasks and recording procedure.

B. Recording

Recordings were made in a quiet room with a low ambient noise level using an external condenser microphone placed approximately 15 cm from the subject's mouth and coupled to a Panasonic NV-GS 180 video camera; the video material was not used in subsequent acoustic analysis. The external condenser microphone was manufactured as part of the original video camera set. The gain of the microphone was set to the same optimal level for all participants to ensure comparable recording conditions. The audio data were digitized from the video recording tape to a computer at a sampling rate of 48 kHz and 16-bit quantization using original Panasonic software. All participants were recorded in the same area of the neurological clinic. As the diagnosis of individual PD patients was made at the time, the specific date of recordings for each participant was different but the overall time schedule was the same. Each participant was recorded in a single session with the speech language pathologist. No time limits were imposed during the recording. All of the participants were familiarized with the speaking tasks and recording procedure. In each recording, the participants performed various speaking tasks as a part of the larger protocol. All participants could repeat their performance in case any errors occurred with respect to the speaking task. Neither participant fatigue nor any changes in the quality of voice from the beginning to the end of the session were observed.

C. Speech samples

Four different speaking tasks were evaluated in the present study including sustained phonation, sentence repetition, reading passage, and monologue. In all speaking tasks, the vowels /a/, /i/, and /u/ were of interest. In the first speaking task, subjects were instructed to make a sustained phonation at a comfortable pitch and loudness with one breath, each vowel separately. The second speaking task was multiple repetition of the Czech phrase "*Kolik mate ted u sebe asi penez*," ([ˈkɔlɪkˈmaːtɛ ˈʊsɛbɛˀasɪ ˈpɛɲɛs]; How much money do you have in your wallet?) which was read in one breath and repeated five times.[1] The acoustic analysis of the corner vowels was performed from the part "**u** sebe **a**si". In the third speaking task, each participant read a standardized passage composed of 80 words (see Appendix A). As indicated by the underlined vowels in the text appearing in Appendix A, 30 vowels per passage were studied, including 10 occurrences of /a/, 10 occurrences of /i/, and 10 occurrences of /u/. The fourth speaking task consisted of monologue where the participants were instructed to speak about what they did during the current day or week, their family, their job, or their interests. For each participant, 10 occurrences of the three vowels /a/, /i/, and /u/ were extracted from the monologue. As there is no available methodology for vowel selection from spontaneous speech such as monologue, the inclusion criteria for the entire word and the vowel itself was established as follows:

(a) The word from which the vowel was selected, as a whole, had to be intelligible and perceptually normal.
(b) Only one same corner vowel (/a/, /i/ or /u/) could be extracted from one specific word.
(c) As there is no reduction in vowel duration due to occurrence in non-stressed syllables in the Czech language, the vowels were elicited from both stressed as well as non-stressed syllables equally.
(d) The selected vowel must not have been induced by confounding effects such as coarticulation with surrounding phonemes. To ensure this condition, the vowels were used only if they occurred separately or followed a voiceless consonant.
(e) The minimal length of the vowel had to be 40 ms, with at least a 30 ms segment that could be considered as a stable part of the vowel. The stable part of the vowel refers to the vowel segment where the first two formants were visible and their format contours did not exhibit marked slopes.

(f) The vowels were extracted from the entire duration of the monologue.[2]

D. Acoustic analyses

Acoustic measures were performed using the widely used, specialized speech-analysis software PRAAT (Boersma and Weenink, 2001, available at www.praat.org). Using PRAAT, both the combined wideband spectrographic display and the power spectral density were used to determine F1 and F2 frequencies in Hz. The formant frequencies of vowels /a/, /i/, and /u/ were extracted from the entire duration of sustained phonation, and from a 30-ms segment at the temporal midpoint of the stable part of each vowel (in order to avoid the influence of vowels preceding or following) in speaking tasks of sentence repetition, reading passage, and monologue.[3] The vowel data of F1 and F2 were separately averaged for all corner vowels of each participant and each individual speaking task. The measurements of VSA, F2i/F2u ratio, and VAI were calculated from these averages. The measurement of VSA is expressed in Hz^2, and can be easily calculated using the following formula (Liu *et al.*, 2005):

$$\mathrm{VSA} = 0.5 \times |\mathrm{F1i} \times (\mathrm{F2a} - \mathrm{F2u}) + \mathrm{F1a} \times (\mathrm{F2u} - \mathrm{F2i}) + \mathrm{F1u} \times (\mathrm{F2i} - \mathrm{F2a})|. \quad (1)$$

The measurement of VAI can be expressed using the following formula (Roy *et al.*, 2009):

$$\mathrm{VAI} = \frac{\mathrm{F1a} + \mathrm{F2i}}{\mathrm{F1i} + \mathrm{F1u} + \mathrm{F2a} + \mathrm{F2u}}. \quad (2)$$

E. Measurement reliability

Intra-judge reliability was assessed following the reanalysis of 25% of all vowel data by the investigator that performed the original set of measures. Pearson correlation analysis indicated significant, positive intra-judge correlation for F1 measures ($r = 0.91$; $p < 0.001$) and for F2 measurement ($r = 0.99$; $p < 0.001$). The mean intra-judge standard error of measurement (SEM) was 15 ± 11 Hz for F1 measures and 16 ± 12 Hz for F2 measures. *Inter-judge* reliability was calculated based on the reanalysis of 25% of all vowel data by a second investigator blinded to participant conditions that was well-trained in the analysis method using the same program. Pearson correlation indicated a significant, positive inter-judge correlation for F1 measures ($r = 0.93$; $p < 0.001$) and for F2 measurement ($r = 0.99$; $p < 0.001$). The mean inter-judge SEM was 15 ± 11 Hz for F1 measures and $17\,\mathrm{Hz} \pm 12\,\mathrm{Hz}$ for F2 measures. *Test-retest* reliability was performed following correlation between the second and third set of sentence repetitions. Pearson correlation indicated significant, positive test-retest reliability for F1 ($r = 0.93$; $p < 0.001$) and F2 ($r = 0.97$; $p < 0.001$) measures. Measurement reliability results in this study are in agreement with previous studies on vowel articulation in dysarthric speakers (Tjaden *et al.*, 2005; Sapir *et al.*, 2007; Sapir *et al.*, 2010).

F. Perceptual assessment of speech performance in PD

For further investigation, the PD subjects were separated into two groups according to independent perceptual assessment performed by three equally trained MDS experienced in the early diagnosis of PD. In accordance with clinical practice, the speech of PD patients was evaluated by item 18 of the UPDRS III (global perceptual description of patient speech, ranked from 0 to 4, where 4 represents complete unintelligibility of speech). As a result, the first group consisted of 10 PD subjects *with an absence of perceptible dysarthria* (hereafter, $\mathrm{PD_{ND}}$ ($PD_{no\ dysarthria}$), 0 points on item 18 representing "unaffected speech"), and the second group consisted of 10 PD subjects *with the presence of mild hypokinetic dysarthria* [hereafter, $\mathrm{PD_{MD}}$ ($PD_{mild\ dysarthria}$), 1 point on item 18 representing "slightly impaired speech"]. The patient was designated as $\mathrm{PD_{MD}}$ if at least one MDS raised suspicion about affected speech due to PD.

G. Statistical analysis and classification

As the Kolmogorov-Smirnov test for independent samples showed that acoustic variables were normally distributed, analysis of variance (ANOVA) with *post hoc* Bonferroni adjustment was used to assess group differences across the data. The adjusted level of significance was set at $p < 0.01$.

Although statistical significance provides useful information regarding the difference between group distributions, there are several classification methods that provide a complete picture of the sensitivity of a given measurement in determining subject-group status. To gain reliable classification results, we first removed the statistically insignificant measurements. As a result, only statistically significant measures (hereafter, main indices) were included in the subsequent classification.

The classification experiment was based on the *minimax theorem* (Schlesinger and Hlavac, 2002). The solution of the minimax theorem is established using a strategy which compares the likelihood ratio with the threshold value. Considering that X can be defined as a set of observations and K as a set of object states, the probability distribution $p_{X|K}(x|k)$ using the set X is then in correspondence with each state k. The strategy is based on the decomposition of $X(k)$, $k \in K$, which determines for each observation $x \in X$ that the object is in the state k on condition $x \in X(k)$. Each strategy is described by dividing set X into $|K|$ numbers,

$$\omega(k) = \sum_{x \notin X_k} p_{X|K}(x|k), \quad (3)$$

i.e., by the conditional probabilities of a wrong decision under the condition that the actual true hidden state of the object is k. The minimax task can then be formulated to find a strategy which minimizes $\max_{k \in K}\ \omega(k)$.

The Gaussian kernel density method with automatic data-driven bandwidth was applied to model the probability distribution of the main acoustic indices for the PD and HC groups, and the minimax task was further solved through a

linear programming technique (Schlesinger and Hlavac, 2002). To validate the reproducibility of the minimax classifier, cross-validation with a leave-one-subject-out method was applied, i.e., data from one speaker was used for testing whereas data from the remaining speakers was used to train the classification model.

H. Overall evaluation

The individual steps related to overall evaluation, corresponding with the proposed aims of the study can be summarized as follows:

1. Analysis of variance (ANOVA) was used to find statistically significant differences between the PD and HC groups across all variables and speaking tasks. The significant differences between the PD and HC groups would indicate that imprecise vowel articulation can be considered a marker of early PD.
2. To investigate the suitability of each speaking task in differentiating between the PD and HC groups, we introduced the measure of *task index*, which is computed as the average classification performance (minimax task) across all main acoustic indices. A better task index classification score would indicate greater potential of the speaking task to reveal parkinsonian deficits in vowel articulation.
3. To examine the suitability of the main acoustic indices in differentiating between groups, we designed a measure termed the *acoustic index*, which is calculated as the average classification performance (minimax task) across all speaking tasks. A better acoustic index classification score would imply greater sensitivity of the acoustic variable to capture defects in vowel articulation of mildly impaired PD speech.
4. The minimax task was used to determine whether objective acoustic measures were more sensitive in revealing PD-induced articulation deficits than perceptual evaluation by an experienced MDS, and to predict whether the speakers with perceptible dysarthria (PD_{MD}) as well as speakers with no perceptible dysarthria (PD_{ND}) are correctly identified as PD. A high classification score for PD_{ND} would indicate that objective acoustic measures are able to capture even minor abnormalities in PD vowel articulation, which may be barely distinguishable from the speech of healthy individuals. A high classification performance for PD_{MD} would imply that the severity of vowel articulation deficits contributes to the overall perceptual impression of dysarthric speech.

III. RESULTS

A. Group differences (Objective 1)

Figure 1 shows the mean and SD (error bars), as well as statistically significant differences (stars) between the PD and HC groups, across all formant-based measurements and speaking tasks. The four main acoustic indices (F2u, VSA, F2i/F2u, VAI) were sufficiently sensitive to separate early-stage PD from HC. For single formant measurements, only F2u differed between PD and HC speakers. The significant differences between both groups for F2u were found for the speaking tasks of sentence repetition [$F(1,34) = 16.6$, $p < 0.001$], reading passage [$F(1,34) = 9.8$, $p < 0.01$], and monologue [$F(1,34) = 18.8$, $p < 0.0001$]. The direction of group differences in each case was consistent with the general hypothesis of increased F2u in PD. In addition, all three complex measurements (VSA, F2i/F2u, VAI) were sufficiently sensitive to capture deficits in vowel articulation, with a consistent direction of group differences indicating reduced vowel space as well as F2iF2u and VAI ratios due to PD. Considering reduced vowel space, significant differences between the PD and HC groups captured by VSA were revealed for the speaking tasks of sentence repetition [$F(1,34) = 11.2$, $p < 0.01$] and monologue [$F(1,34) = 8.4$, $p < 0.01$]. In the case of distinction between vowels, altered PD speech performance was found in the measurements of F2i/F2u for the speaking tasks of sentence repetition [$F(1,34) = 12.7$, $p < 0.001$], reading passage [$F(1,34) = 7.5$, $p < 0.01$], and monologue [$F(1,34) = 19.6$, $p < 0.0001$]. Regarding vowel centralization, significant differences between both groups were found in the measurement of VAI for the speaking tasks of sentence repetition [$F(1,34) = 8.2$, $p < 0.01$] and monologue [$F(1,34) = 13.3$, $p < 0.001$]. Therefore, in agreement with our hypothesis, imprecise vowel articulation can be considered an early marker of PD.

B. Differences across speaking tasks (Objective 2)

Figure 2 details classification results with estimated probability distributions for each speaking task across all main acoustic indices. In accordance with the results of statistical analyses, the task index showed that sustained phonation reached the lowest classification performance of 58.7%. All the remaining speaking tasks including sentence repetition, reading passage, and monologue can be considered suitable for the evaluation of vowel articulation in PD. Comparing the results of task index for non-spontaneous and spontaneous speech, sentence repetition and reading passage had an average performance of 69.5% (73.5% for sentence repetition and 65.5% for reading text) whereas monologue reached a score of 76.5%. These findings are consistent with our hypothesis that the performance of vowel articulation in PD speakers is altered to a greater extent in spontaneous rather than non-spontaneous utterances, whereas isolated sustained phonations cannot be considered a suitable task for the investigation of vowel articulation in early PD.

C. Differences across formant-based measurements (Objective 3)

Classification results based on the main acoustic indices (F2u, VSA, F2i/F2u, VAI) through individual speaking tasks are presented in Fig. 2. Considering differences between the PD and HC groups among individual measurements, VSA and F2i/F2u extracted from the monologue reached the best classification performances of 80.4% and 80.0%, respectively. The acoustic index showed very similar classification accuracy across all acoustic indices with a performance of 70.6% for F2u, 70.4% for VSA, 69.1% for F2i/F2u, and

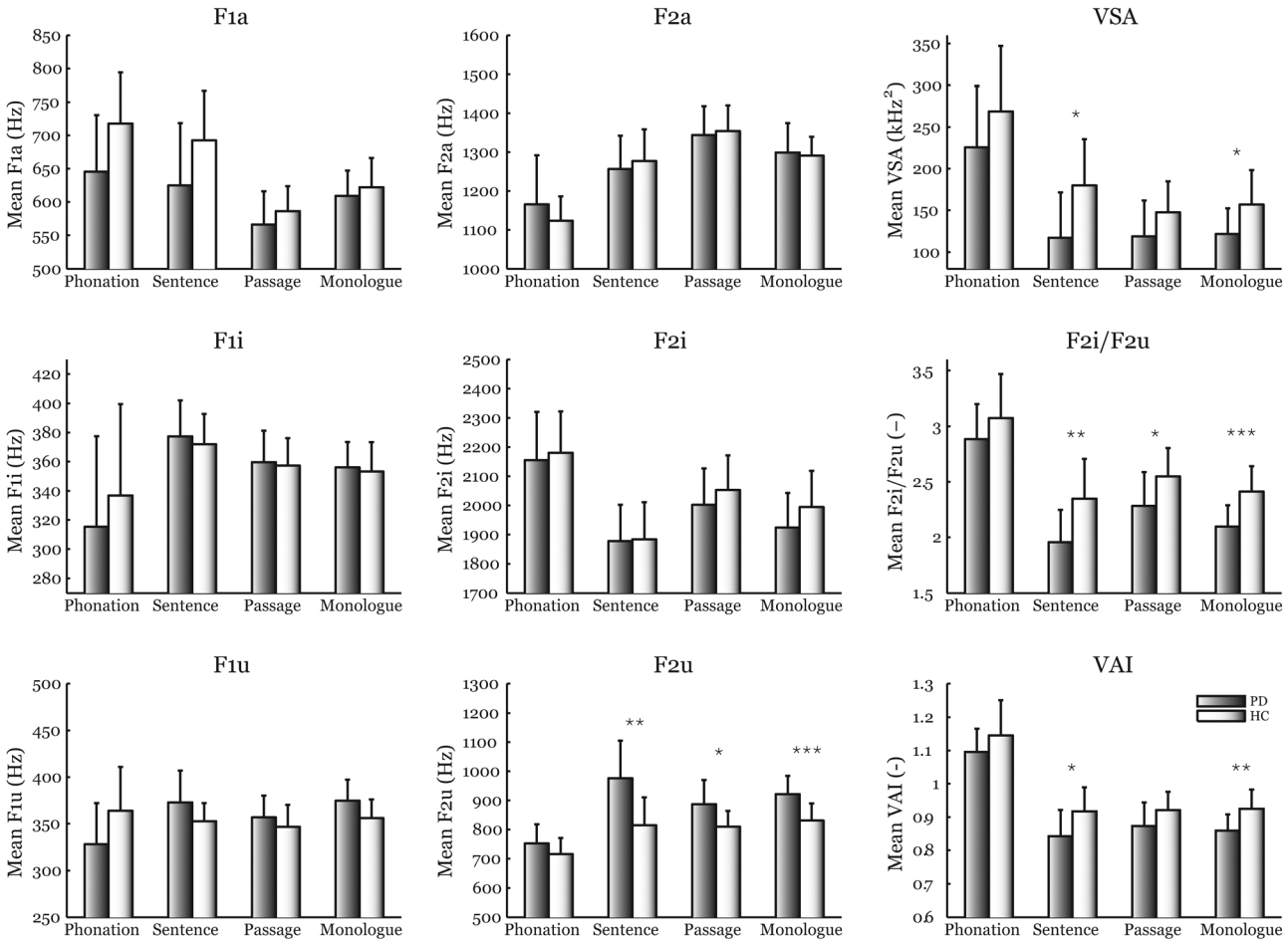


FIG. 1. First (F1) and second (F2) formant frequencies for each corner vowel (/a/, /i/, and /u/), vowel space area (VSA), F2i/F2u ratio, and vowel articulation index (VAI) in PD speakers (gray) and healthy controls (white). Measurements were performed using different types of speech material including sustained phonation, sentence repetition, reading passage, and monologue. The bars represent mean values and error bars standard deviations. Stars indicate significant differences between PD patients and controls: $*p < 0.01$; $**p < 0.001$; $***p < 0.0001$.

64.1% for VAI. Contrary to our hypothesis, distinctions between vowels and vowel centralizations were not revealed to be superior to reduced vowel space in the detection of mildly impaired speech in early stage PD, while articulation of the vowel /u/ captured by F2u seems to be more affected when compared to changes in other vowel frequencies.

D. Comparison between objective measures and perceptual evaluation (Objective 4)

Table I details the results of classification for two groups of PD speakers, with perceptible dysarthria (PD_{MD}) and without perceptible dysarthria (PD_{ND}), across all main acoustic indices and speaking tasks. Altered vowel articulation in PD_{MD} patients was confirmed using all main acoustic indices (F2u, VSA, F2i/F2u, VAI) through all investigated speaking tasks except sustained phonation. This finding, related to imprecise vowel articulation in PD_{MD} subjects, was in agreement with the perceptual evaluation of the MDS. In other words, acoustic metrics generally achieved better results in predicting the group status of PD_{MD} patients than PD_{ND} patients. Interestingly, the measurement of F2 frequency based on the single vowel /u/ (F2u) was best able to capture mild dysarthria, with scores ranging from 81.9% to 88.7%. On the other hand, F2u alone was not sufficiently sensitive to reveal changes in vowel articulation of PD_{ND} patients. The contrary is true for measurements of VSA and F2i/F2u, especially when extracted from monologue, which were able to detect impaired vowel articulation in both PD groups. In fact, the VSA based on monologue was more successful in predicting PD_{ND} group status with a classification score of 80.3% in comparison to 76.3% achieved by F2i/F2u. In contrast, F2i/F2u extracted from monologue reached a higher score of 85.6% for the PD_{MD} group in comparison to 83.7% by VSA. The measurement of VAI was not found to be superior to complex measurements of F2i/F2u and VSA. In correspondence with our hypothesis, the performance of objective acoustic measures of vowel articulation was superior to the subjective clinical evaluation of disturbed speech.

IV. DISCUSSION

In the present study, we investigated various formant-based measures in a group of *de novo* male PD patients in comparison to healthy subjects. Vowel production was examined across different types of speech tasks including sustained phonation, sentence repetition, reading passage, and monologue. The acoustic parameters for subsequent analysis consisted of F1 and F2 for each corner vowel, the

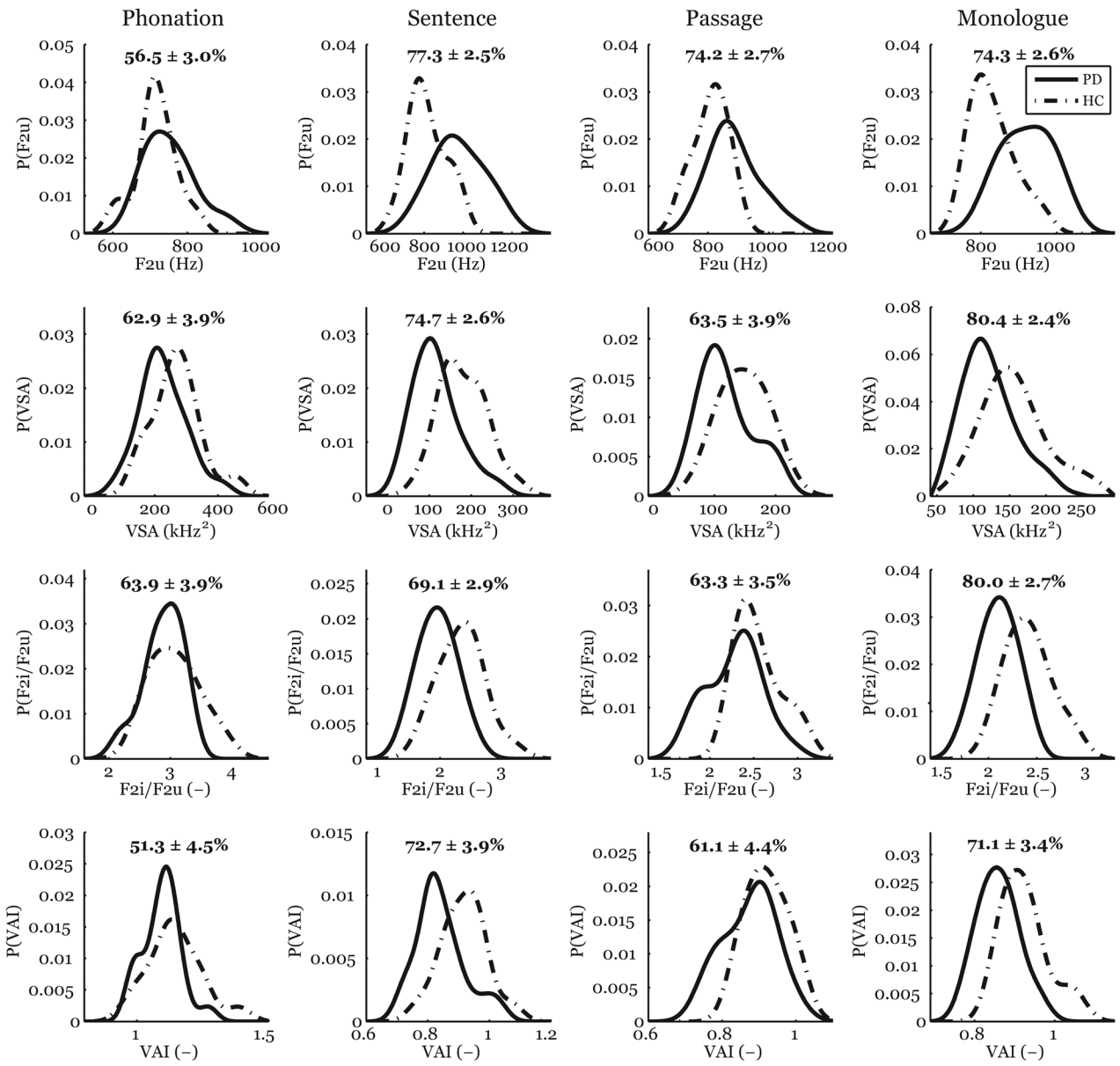


FIG. 2. Probability densities P('measure') with overall classification accuracy according to the minimax task across all main acoustic indices (F2u, VSA, F2i/F2u, VAI) and speaking tasks (sustained phonation, sentence repetition, reading passage, monologue). Solid lines are displayed for PD subjects, the dash-dot lines for HC speakers.

traditional measure of VSA, and the recently introduced measures of F2i/F2u and VAI. Much of what is known about vowel articulation in PD has generally been based on pharmacologically treated patients with various degrees of dysarthria, without comparison between different types of speaking tasks. Thus, the current investigation extends our knowledge related to the characteristics of vowel articulation in the early stages of PD prior to pharmacotherapy, and the usage of more complex speech material.

Our results show that early-stage PD speakers manifest increased F2u, lowered VSA, reduced distinction between vowels captured by F2i/F2u, and abnormalities in formant centralizations measured by VAI across sentence repetition, reading passage, and monologue. In fact, vowel articulation in PD was more acoustically altered during spontaneous speech such as monologue in comparison to the non-spontaneous tasks of sentence repetition and reading passage. Moreover, isolated vowel phonation was found to be an inappropriate task to reveal early changes in parkinsonian articulation. Considering comparisons between measurements, there were no essential differences between the results obtained using various complex formant-based metrics (VSA, F2i/F2u, VAI), although the production of the vowel /u/ (as captured by F2u) was found to be altered to a greater extent when compared to articulation of /a/ and /i/ vowels. Further results of our study indicate that objective acoustic measures are more sensitive in revealing PD-induced articulation deficits than perceptual evaluation by experienced clinicians, and therefore may be helpful in capturing even subclinical signs of speech impairment in PD. To summarize, these findings provide greater insight into impaired vowel articulation in the early stages of PD and may be helpful regarding the advent of neuroprotective treatment as well as speech rehabilitation in PD (Rusz *et al.*, 2011b; Sapir *et al.*, 2010).

TABLE I. List of classification results for PD speakers with mild dysarthria (PD_{MD}) and no perceptible dysarthria (PD_{ND}) across all main acoustic indices (F2u, VSA, F2i/F2u, VAI) and speaking tasks (sustained phonation, sentence repetition, reading passage, monologue).

	Classification score (%)	
Measurement	PD_{MD}	PD_{ND}
Sustained phonation		
F2u	53.5 ± 5.6	60.4 ± 4.8
VSA	67.0 ± 13.3	69.9 ± 12.9
F2i/F2u	72.8 ± 6.7	58.9 ± 4.8
VAI	64.3 ± 6.7	50.2 ± 7.7
Sentence repetition		
F2u	84.0 ± 3.0	72.9 ± 6.5
VSA	84.7 ± 3.5	69.7 ± 3.8
F2i/F2u	78.6 ± 3.2	64.3 ± 3.9
VAI	84.9 ± 3.0	67.2 ± 4.3
Reading passage		
F2u	81.9 ± 2.9	66.0 ± 3.6
VSA	62.2 ± 4.1	68.0 ± 4.1
F2i/F2u	72.3 ± 6.0	59.1 ± 5.9
VAI	69.0 ± 7.3	58.7 ± 7.7
Monologue		
F2u	88.7 ± 3.4	61.5 ± 3.8
VSA	83.7 ± 5.3	80.3 ± 5.5
F2i/F2u	85.6 ± 5.7	76.3 ± 5.2
VAI	71.0 ± 4.0	69.1 ± 3.9

A. Imprecise vowel articulation as an early marker of PD

Although imprecise articulation is considered as one of the most common deficits associated with hypokinetic dysarthria (Darley *et al.*, 1969), previous studies have reported that voice disorders in PD occur more frequently than articulation disorders (Logemann *et al.*, 1978). Moreover, prosodic patterns and features of dysphonia have been suggested to be the most salient early signs of vocal impairment in PD (Rusz *et al.*, 2011a; Stewart *et al.*, 1995). In contrast to previous reports, our results show that impaired vowel articulation was present in 80% of our patients and may be one of the first signs of speech impairment in PD. Individuals with early-stage PD demonstrated significant differences in measurements of F2u, VSA, F2i/F2u, and VAI relative to HC subjects. Our results are generally consistent with previous studies where reduced articulatory movements were investigated in treated PD patients with mild to moderate dysarthria (Sapir *et al.*, 2007; Sapir *et al.*, 2010; Skodda *et al.*, 2011), however, to the best of our knowledge, no previous studies have examined vowel articulation in *de novo* patients with hypokinetic dysarthria or dysarthria of another type.

B. Effect of speaking task

Our results indicate that both spontaneous and non-spontaneous speech is suitable for the assessment of early changes in vowel articulation associated with PD, while isolated vowel phonations were found to be inappropriate. In fact, vowel articulation performance in PD was found to be altered to a greater extent in spontaneous speech such as monologue when compared to the typical non-spontaneous speaking tasks of sentence repetition or reading passage. These findings are in agreement with the general assumption that the efficiency of speech production in PD varies with the task performed (Caligiuri, 1989; Rosen *et al.*, 2005; Weismer, 1984). Furthermore, in a previous study where both spontaneous and non-spontaneous speech were collected from a single patient with PD diagnosed 18 yr prior to investigation (Kempler and Van Lacker, 2002), the intelligibility of spontaneous speech was found to be severely affected when compared to non-spontaneous speech. As imprecise vowel articulation commonly contributes to reduced speech intelligibility in dysarthria (Kim *et al.*, 2011a; Liu *et al.*, 2005; Weismer *et al.*, 2001), the observation by Kempler and Van Lacker (2002) is in accordance with our results showing monologue as the most affected speaking task in parkinsonian patients, even in the early stages of the disease. Thus, it might be rewarding to take into account the type of speaking task during the evaluation of dysarthria.

With respect to differences in articulatory impairment between various types of speaking tasks, one possible explanation is the complexity of the speaking task. Contrary to more advanced stages of PD (Eliasova *et al.*, 2013) where sustained phonation may be a suitable task, in early-stage parkinsonian speakers it is too simple a task to capture subtle changes in vowel articulation. Indeed, the size and centralization of the VSA as well as the ratios of F2i/F2u and VAI obtained from phonation differ in comparison to the results elicited from other speaking tasks (see Fig. 1). Although sentence repetition maintains the advantage that each individual corner vowel is extracted from the same repeated phrase, usage of these extracted vowels does not appear to reflect the variety and flexibility of the entire utterance. Furthermore, the lexical factors of phonological neighborhood density and word frequency can significantly influence the performance of vowel articulation (Watson and Munson, 2008). In contrast to sentence repetition, the variety of utterances is captured well using reading passage where the vowel performances are extracted from a variety of words, and lexical factors are then partially controlled by average measurements. However, in non-spontaneous speech such as reading passage or sentence repetition, the speaker is simply pronouncing ready-made text and thus can provide special attention to articulatory planning (Levelt, 1989). In contrast to non-spontaneous tasks, spontaneous speech represented by monologue requires that the speaker carry out the complete planning process, and therefore the articulatory mechanisms receive relatively less attention. However, in spontaneous speech, there is a limited possibility of using the same words in the same phrases for each participant, and thereby the identification and extraction of vowels must be carefully conducted. In summary, we can assume that the final speech performances are related to the overall articulatory demands and complexity of the individual speaking tasks.

C. Acoustic changes in vowel articulation due to PD

In the present study, we did not detect any fundamental differences between various complex formant-based

measurements. However, this is not unexpected as all of the complex formant-based measurements are based upon the same base measures of F1 and F2 frequencies. On the other hand, subtle differences in findings across formant-based measurements may provide certain clues about acoustic changes in vowels due to the development of PD. Considering the main acoustic indices, F2u is sensitive only to the vowel /u/, F2i/F2u to both /u/ and /i/ vowels, and VAI to all three corner vowels /u/, /i/, and /a/. F2u was the only measurement based on a single vowel that allowed the differentiation of PD and HC groups, and therefore the vowel /u/ can be considered the most affected in mild dysarthria of PD individuals. F2i/F2u was more sensitive in capturing deficits of PD vowel articulation than F2u alone, while VAI, which also takes formant frequencies of the vowel /a/ into account, achieved the worst classification performance in separating PD from HC speakers. The vowel /a/ also contributes to the VSA, but the formants based on the vowel /a/ do not need to negatively contribute to its overall performance (as given by the principle of VSA construction) such as in the case of formant centralization measured by VAI. In light of these observations, we may hypothesize that articulatory deficits are due mainly to alterations of the vowel /u/, followed by the vowel /i/, with the vowel /a/ remaining most resistant to change in the earlier stages of PD. This hypothesis is also in agreement with previous studies reporting the measurement of VAI superior to VSA in parkinsonian patients several years after the diagnosis was established (Sapir *et al.*, 2010; Skodda *et al.*, 2012).

This continuum of articulatory deficits developing through the vowels /u/, /i/, and /a/, respectively, can be also discussed in physiological terms. Considering tongue position and lip posture for the cardinal vowels /a/, /i/, and /u/, the tongue is positioned low for the vowel /a/, high and forward for the vowel /i/, and high and backward for the vowel /u/, whereas lip posture is spread for both the /a/ and /i/ vowels and rounded for the vowel /u/ (Hasegawa-Johnson *et al.*, 2003). Therefore, we may assume that production of the vowel /a/ is a less demanding task than production of the vowels /i/ and /u/. In comparison to the vowel /i/, the articulation of the vowel /u/ requires more demanding involvement of the orofacial muscles to create a tightly rounded lip posture. Admittedly, as the vowel /u/ is characterized by a posterior rise of the tongue root, we may consider that problems with tongue articulation develop in reverse, i.e., from the root to the tip of the tongue, and the resulting tongue restriction may also be related to swallowing abnormalities in PD (Sapir *et al.*, 2008; Tjaden, 2008). In general, concerning the pathophysiological mechanism responsible for the development of speech and other motor deficits in PD, speech impairment has generally been attributed to dopamine deficiency as well as hypokinesia and rigidity of the vocal tract (Schulz and Grant, 2000). On the basis of investigation in PD patients tested and re-tested within a few years, recent innovative studies have found that bradykinesia, rigidity, and axial parkinsonian symptoms are primarily responsible for restricted vowel articulation in PD (Rusz *et al.*, 2013; Skodda *et al.*, 2012).

D. Perceptual and acoustic findings in the speech of early-stage PD individuals

Previous studies have reported a relationship between measurements of vowel articulation and the perceptual impression of intelligibility in dysarthric speakers (Kim *et al.*, 2011a; Liu *et al.*, 2005; Weismer *et al.*, 2001). Although the perceptual classification of dysarthria is typically based on intelligibility rating measures using orthographic transcription (for example, see Kim *et al.*, 2011a; Liu *et al.*, 2005), such methods are not applicable to the evaluation of speech disorders in early-stage PD, where dysarthria is rather mild or even imperceptible, and has no or limited effect on speech intelligibility (Stewart *et al.*, 1995). In the present study, parkinsonian speakers were subjectively separated into two groups (no perceptible dysarthria and mild dysarthria) by experienced clinicians, and a very high classification performance was achieved (up to 90%) when comparing patients with mild dysarthria and controls. This finding suggests that measurements of vowel articulation may be considered suitable for the assessment of speech intelligibility. Subsequently, we observed that the objective acoustic measures of vowel articulation were able to predict articulatory impairment in PD patients with no perceptible dysarthria with relatively high accuracy (up to 80%). Accordingly, it has been proposed that objective acoustic measures may capture even minor abnormalities in PD speech (Forrest and Weismer, 2009; Rusz *et al.*, 2011a; Stewart *et al.*, 1995). Therefore, objective acoustic analyses may be helpful in revealing even subclinical signs of speech impairment in PD.

E. Limitations of the present study

In the course of this study, we investigated only 20 male parkinsonian patients due to limited opportunities in recruiting more early-stage PD individuals prior to dopaminergic treatment. Despite this limitation, we do not believe that there would be any fundamental changes in the overall progression of vowel articulation or that the current findings would differ with a substantial increase in the number of subjects. Previous research has suggested that gender may have an impact on the progression of dysarthria due to sexual dimorphism of laryngeal size (Hertrich and Ackermann, 1995). As our study consisted only of male participants, we cannot exclude that impairment of vowel articulation is influenced by gender-specific aspects of speech. One further limitation of the present study is that we did not investigate relationships between vowel durations and vowel articulation measurements (Tjaden *et al.*, 2005), and hence possible effects related to the speaking rate on vowel articulation in PD cannot be eliminated.

V. CONCLUSION

The present study provides evidence for restricted vowel articulation in early-stage PD speakers prior to dopaminergic treatment. Our results demonstrate that spontaneous speech is more likely to show true deficits in the speech performance of PD patients. Specific changes in speech due to neurological disorders such as PD may have the potential to

contribute to existing assessment batteries. Acoustic measurements of vowel articulation may therefore be useful in the early detection of speech impairment in PD, for monitoring the severity of dysarthria and disease progression, and in the evaluation of treatment response.

ACKNOWLEDGMENT

This research was partly supported by the Czech Science Foundation (GACR 102/12/2230), Czech Ministry of Health (NT 12288-5/2011 and NT14181-3/2013), Czech Ministry of Education (MSM 0021620849), and Charles University in Prague (PRVOUK-P26/LF1/4).

APPENDIX

Reading passage with labeled corner vowels /a/, /i/, and /u/ that was used in acoustical analyses.

Když člověk poprvé vs**a**dí do země s**a**zeničk**u**, chodí se na ni dívat tř**i**krát denně: t**a**k co, povyrostla **u**ž n**e**bo ne? **I** t**a**jí dech, nakl**á**ní se nad ní, př**i**tlač**í** troch**u** p**ů**du u jejich kořínk**ů**, načechrává jí lístky **a** vůbec se obtěžuje různým konáním, které považuje za **u**žitečnou péč**i**. **A** když se s**a**zeničk**a** přesto **u**jme a roste jako z vody, t**u** člověk žasne nad tímto divem přírody, má poc**i**t čehos**i** jako zázrak**u** **a** považuje to za jeden ze svých největš**í**ch osobních **ú**spěchů.

[1]The motivation behind the single exhalatory effort required for each sentence was to differentiate the task of sentence repetition from the connected reading of text, since it has been reported that the performance of PD speakers may differ at the beginning and end of an utterance (Skodda and Schlegel, 2008).

[2]As the previous observation has shown that the performance of PD speakers may differ through an utterance (Skodda and Schlegel, 2008), we suggest extracting vowels throughout the entire length of the monologue. However, in our early-stage PD patients there were no statistically significant differences between formants calculated using the first five occurrences and second five occurrences of corner vowels.

[3]In the sustained phonation speech task, only one repetition for each corner vowel was used to calculate formant frequencies as (a) articulation of isolated vowels is not influenced by the preceding or following phoneme and (b) certain fluctuations of formants are treated by the choice of multiple window length of the analyzed segment when compared to 30-ms segments used in sentence repetition, reading passage, and monologue tasks.

Boersma, P., and Weenink, D. (**2001**). "Praat, a system for doing phonetics by computer," Glot Int. **5**, 341–345.
Caligiuri, M. (**1989**). "The influence of speaking rate on articulatory hypokinesia in Parkinsonian dysarthria," Brain Lang. **36**, 493–502.
Darley, F. L., Aronson, A. E., and Brown, J. R. (**1969**). "Differential diagnosis patterns of dysarthria," J. Speech Hear. Res. **12**, 426–496.
Eliasova, I., Mekyska, J., Kostalova, M., Marecek, R., Smekal, Z., and Rektorova, I. (**2013**). "Acoustic evaluation of short-term effects of repetitive transcranial magnetic stimulation on motor aspects of speech in Parkinson's disease," J. Neural Transm. **120**, 597–605.
Forrest, K., and Weismer, G. (**2009**). "Acoustic analysis of motor speech disorders," in *Clinical Management of Sensorimotor Speech Disorders*, edited by M. R. McNeil (Thieme, New York), pp. 46–63.
Forrest, K., Weismer, G., and Turner, G. S. (**1989**). "Kinematic acoustic and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults," J. Acoust. Soc. Am. **85**, 2608–2622.
Goberman, A. M., and Coelho, C. (**2002**). "Acoustic analysis of Parkinsonian speech I: Speech characteristics and L-Dopa therapy," Neurorehabilitation **17**, 237–246.
Hasegawa-Johnson, M., Pizza, S., Alwan, A., Cha, J. S., and Haker, K. (**2003**). "Vowel category dependence of the relationship between palate height, tongue height, and oral area," J. Speech Hear. Res. **46**, 738–753.
Hertrich, I., and Ackermann, H. (**1995**). "Gender-specific vocal dysfunction in Parkinson's disease: electroglottographic and acoustic analyses," Ann. Oto. Rhinol. Laryngol. **104**, 197–202.
Hoehn, M., and Yahr, M. (**1967**). "Parkinsonism: onset, progression and mortality," Neurology **17**, 427–442.
Hornykiewicz, O. (**1998**). "Biochemical aspects of Parkinson's disease," Neurology **51** (Suppl. 2), S2–S9.
Hughes, A. J., Daniel, S. E., Kilford, L., and Lees, A. J. (**1992**). "Accuracy of clinical diagnoses of idiopathic Parkinson's disease. A clinic-pathological study of 100 cases," J. Neurol. Neurosurg. Psychiatry **55**, 181–184.
Hunker, C., Abbs, J., and Barlow, S. (**1982**). "The relationship between Parkinsonian rigidity and hypokinesia in the orofacial system: a quantitative analysis," Neurology **32**, 749–754.
Kempler, D., and Van Lancker, D. (**2002**). "Effect of speech task on intelligibility in dysarthria: A case study of Parkinson's disease," Brain Lang. **80**, 449–464.
Kent, R., and Kim, Y. (**2003**). "Toward an acoustic typology of motor speech disorders," Clin. Linquist. Phonetics **17**, 427–445.
Kent, R., Weismer, G., Kent, J. Vorperian, H., and Duffy, J. (**1999**). "Acoustic studies of dysarthric speech: methods, progress, and potential," J. Commun. Disord. **32**, 427–445.
Kim, H., Hasegawa-Johnson, M., and Perlman, A. (**2011a**). "Vowel contrast and speech intelligibility in dysarthria," Folia Phoniatr. Logop. **63**, 187–194.
Kim, Y., Kent, R. D., and Weismer, G. (**2011b**). "An acoustic study of the relationships among neurological disease, dysarthria type, and severity of dysarthria," J. Speech Lang. Hear. Res. **54**, 417–429.
Levelt, W. J. M. (**1989**). *Speaking: From Intention to Articulation. A Bradford Book* (The MIT Press, Cambridge, MA), pp. 1–545.
Liu, H., Tsao, F., and Kuhl, P. (**2005**). "The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy," J. Acoust. Soc. Am. **117**, 3879–3889.
Logemann, J. A., Fisher, H. B., Boshes, B., and Blonsky, E. R. (**1978**). "Frequency and cooccurrence of vocal tract dysfunction in the speech of a large sample of Parkinson patients," J. Speech Hear. Disord. **43**, 47–57.
Monses, R. B. (**1976**). "Normal and reduced-phonological space: The production of English vowels by deaf adolescents," J. Phonetics **4**, 189–198.
Postuma, R. B., Lang, A. E., Gagnon, J. F., Pelletier, A., and Montplaisir, J. Y. (**2012**). "How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder," Brain **135**, 1860–1870.
Robertson, L. T., and Hammerstadt, J. P. (**1996**). "Jaw movements dysfunction related to Parkinson's disease and partially modified by levodopa," J. Neurol. Neurosurg. Psychiatry **60**, 41–50.
Rosen, K., Kent, R., and Duffy, J. (**2005**). "Task-based profile of vocal intensity decline in Parkinson's speech," Folia Phoniatr. Logop. **57**, 28–37.
Roy, N., Nissen, S. L., Dromey, C., and Sapir, S. (**2009**). "Articulatory changes in muscle tension dysphonia: evidence for vowel space expansion following manual circumlaryngeal therapy," J. Commun. Disord. **42**, 124–135.
Rusz, J., Cmejla, R., Ruzickova, H., Klempir, J., Majerova, V., Picmausova, J., Roth, J., and Ruzicka, E. (**2011b**). "Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test," Mov. Disord. **26**, 1951–1952.
Rusz, J., Cmejla, R., Ruzickova, H., Klempir, J., Majerova, V., Picmausova, J., Roth, J., and Ruzicka, E. (**2013**). "Evaluation of speech impairment in early stages of Parkinson's disease: a prospective study with the role of pharmacotherapy," J. Neural Transm. **120**, 319–329.
Rusz, J., Cmejla, R., Ruzickova, H., and Ruzicka, E. (**2011a**). "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease," J. Acoust. Soc. Am. **129**, 350–367.
Sapir, S., Ramig, L., and Fox, C. (**2008**). "Speech and swallowing disorders in Parkinson's disease," Curr. Opin. Otolaryngol. Head Neck Surg. **16**, 205–210.
Sapir, S., Ramig, L. O., Spielman, J. L., and Fox, C. (**2010**). "Formant Centralization Ratio: A proposal for a new acoustic measure of dysarthric speech," J. Speech Lang. Hear. Res. **53**, 114–125.
Sapir, S., Spielman, J. L., Ramig, L. O., Stroy, B. H., and Fox, C. (**2007**). "Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: Acoustic and perceptual findings," J. Speech Lang. Hear. Res. **50**, 899–912.

Schlesinger, M. I., and Hlavac, V. (**2002**). *Ten Lectures on Statistical and Structural Pattern Recognition* (Kluwer Academic, Dordrecht), pp. 25–70.
Schulz, G. M., and Grant, M. K. (**2000**). "Effect of speech therapy and pharmacologic and surgical treatments of voice ad speech in Parkinson's disease: A review of the literature," J. Commun. Disord. **33**, 59–88.
Skodda, S., Gronheit, W., and Schlegel, U. (**2012**). "Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease," PLoS ONE **7**, e32132.
Skodda, S., and Schlegel, U. (**2008**). "Speech rate and rhythm in Parkinson's disease," Mov. Disord. **23**, 985–992.
Skodda, S., Visser, W., and Schlegel, U. (**2010**). "Short- and long-term dopaminergic effects on dysarthria in early Parkinson's disease," J. Neural. Transm. **117**, 197–205.
Skodda, S., Visser, W., and Schlegel, U. (**2011**). "Vowel articulation in Parkinson's disease," J. Voice **25**, 467–472.
Stebbing, G., and Goetz, C. (**1998**). "Factor structure of the unified Parkinson's disease rating scale: Motor examination section," Mov. Disord. **13**, 633–636.
Stewart, C., Winfield, L., Hunt, A., Bressman, S. B., Fahn, S., Blitzer, A., and Brin, M. F. (**1995**). "Speech dysfunction in early Parkinson's disease," Mov. Disord. **10**, 562–565.
Tetrud, J. W. (**1991**). "Preclinical Parkinson's disease: detection of motor and nonmotor manifestations," Neurology **41** (Suppl. 2), 69–71.
Tjaden, K. (**2008**). "Speech and swallowing in Parkinson's disease," Top. Geriatr. Rehabil. **24**, 115–126.
Tjaden, K., Rivera, D., and Wilding, G. (**2005**). "Characteristics of the lax vowel space in dysarthria," J. Speech Lang. Hear. Res. **48**, 554–566.
Watson, P. J., and Munson, B. R. (**2008**). "Parkinson's disease and the effect of lexical factors on vowel articulation," J. Acoust. Soc. Am. **116**, 2365–2373.
Weismer, G. (**1984**). "Articulatory characteristics of parkinsonian dysarthria: Segmental and phrase-level timing spirantization and glottal-supraglottal coordination," *The Dysarthrias: Physiology, Acoustic, Perception, Management*, edited by M. R. McNeil, J. C. Rosenbek, and A. E. Aronson (College-Hill Press, San Diego), pp. 101–130.
Weismer, G., Jeng, J.-Y., Laures, J., Kent, R., and Kent, J. (**2001**). "Acoustic and intelligibility of sentence production in neurogenic speech disorders," Folia Phoniatr. Logop. **53**, 1–18.

Appendix A4

Acoustic Investigation of Stress Patterns in Parkinson's Disease

Tykalova T, Rusz J, Cmejla R, Ruzickova H, Ruzicka E (2014) Acoustic Investigation of Stress Patterns in Parkinson's Disease. *J Voice* **28**:129.e1-129.e8.

Journal name (category): **Journal of Voice** (otorhinolaryngology)
Quartile in category: **Q3**
Impact factor: **0.944**
Number of WoS citations (excluding autocitations): **9**

Acoustic Investigation of Stress Patterns in Parkinson's Disease

***Tereza Tykalova, *,†Jan Rusz, *Roman Cmejla, †Hana Ruzickova, and †Evzen Ruzicka,** *†*Prague, Czech Republic*

Summary: Objectives. Although reduced stress is thought to be one of the most deviant speech dimensions in hypokinetic dysarthria associated with Parkinson's disease (PD), the mechanisms of stress production in PD have not been thoroughly explored by objective methods. The aim of the present study was to quantify the effect of PD on prosodic characteristics and to describe contrastive stress patterns in parkinsonian speech.

Methods. The ability of 20 male speakers with early PD and 16 age- and gender-matched healthy controls (HCs) to signal contrastive stress was investigated. Each participant was instructed to unnaturally emphasize five key words while reading a short block of text. Acoustic analyses were based on the measurement of pitch, intensity, and duration. In addition, an innovative measurement termed the stress pattern index (SPI) was designed to mirror the effect of all distinct acoustic cues exploited during stress production.

Results. Although PD patients demonstrated a reduced ability to convey contrastive stress, they could still notably increase pitch, intensity, and duration to emphasize a word within a sentence. No differences were revealed between PD and HC stress productions using the measurements of pitch, intensity, duration, and intensity range. However, restricted SPI and pitch range were evident in the PD group.

Conclusions. A reduced ability to express stress seems to be the distinctive pattern of hypokinetic dysarthria, even in the early stages of PD. Because PD patients were able to consciously improve their speech performance using multiple acoustic cues, the introduction of speech therapy may be rewarding.

Key Words: Parkinson's disease–Speech disorders–Reduced stress–Acoustic analysis–Prosody–Contrastive stress.

INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disorder characterized by the progressive loss of dopaminergic neurons in the substantia nigra, affecting 1–2% of persons over the age of 60 years.[1,2] In addition to cardinal motor manifestations, such as bradykinesia, rigidity, postural instability, and resting tremor, up to 90% of individuals with PD develop an alteration of speech termed hypokinetic dysarthria.[3,4] Moreover, these distinctive speech deficits may be one of the earliest symptoms and appear even several years before the diagnosis is established.[5] Speech deficits commonly reported to be experienced by PD individuals include monoloudness, monopitch, reduced stress, imprecise articulation, variability of speech rate, a breathy and harsh voice, disfluency, voice tremor, and other manifestations that can lead to overall reduced speech intelligibility.[6,7] These changes in speech production may have a significant, negative impact on social interactions and overall quality of the patient's life.[8]

Speech is a unique, complex, dynamic motor activity through which individuals express their thoughts and feelings.[7] From an acoustic perspective, speech can be surveyed with respect to five speech subsystems including respiration, phonation, resonance, articulation, and prosody. Prosody itself is an important aspect of language that is necessary for recovering the intended meanings of an utterance, that is, information that is unavailable in the orthographic transcription. In particular, prosody may serve a variety of functions, including signaling questions or lexical boundaries, conveying contrastive meanings, and expressing emotions and attitudes.[9–11] One of the techniques used by speakers to convey these suprasegmental features is word and sentence stress, representing the relative emphasis given to a certain syllable or word.

In acoustics, there are three main prosodic cues commonly associated with stress: pitch, intensity, and duration.[12–14] In the 1960s, studies of healthy speakers established pitch prominence as the primary marker of stress.[12,15–17] For example, Fry[15] measured pitch and duration changes in lexical stress pairs (eg, HOTdog vs hot DOG) and found pitch to be superior to duration. Another experiment conducted by Fry[13] to determine whether intensity or duration was a better cue to stress showed that duration, on the whole, was a more distinctive cue. On the other hand, some researchers have argued that duration and/or intensity also convey stress and might be at least as important as pitch.[9,10,18,19] Moreover, the way stress manifests itself in the speech stream is partially language dependent.[20,21] From this point of view, the prosodic characteristics of stress are somewhat ambiguous, even considering nonimpaired speech.

Although the manifestation of reduced stress has been well documented in several motor speech disorders,[7,22–24] the mechanisms of stress production in hypokinetic dysarthria of PD have not been thoroughly explored by objective methods. Ma et al[25] analyzed question-statement contrast in 14 Cantonese PD speakers and found that subjects with PD used similar acoustic cues as healthy adults; however, adequate contrast was not observed in all speakers. Cheang and Pell[26] reported

Accepted for publication July 2, 2013.

This research was partly supported by the Czech Science Foundation (GACR 102/12/2230), Czech Ministry of Health (NT 12288-5/2011 and NT14181-3/2013), Czech Ministry of Education (MSM 0021620849), and Charles University in Prague (PRVOUK-P26/LF1/4).

From the *Department of Circuit Theory, Czech Technical University in Prague, Faculty of Electrical Engineering, Prague, Czech Republic; and the †Department of Neurology and Centre of Clinical Neuroscience, Charles University in Prague, First Faculty of Medicine, Prague, Czech Republic

Address correspondence and reprint requests to Jan Rusz, Department of Circuit Theory, Czech Technical University in Prague, Faculty of Electrical Engineering, Technicka 2, 166 27, Prague 6, Czech Republic. E-mail: ruszjan@fel.cvut.cz

Journal of Voice, Vol. 28, No. 1, pp. 129.e1-129.e8

0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2013.07.001

that medicated patients in the early stages of PD exhibit various changes in the speaking tasks of lexical stress, contrastive stress, and emotional prosody. The acoustic results showed that the average amplitude measurement was the most robust parameter, as tokens elicited from PD speakers were lower in amplitude than tokens spoken by healthy participants in all three production tasks. Meanwhile, pitch was found to be aberrant among PD speakers for contrastive stress and emotional prosody; duration measures revealed anomalies between healthy and PD speakers merely in phonemic stress.[26] In addition, there is also some evidence that the ability of PD speakers to express intended stress or emotions through prosody is often poorly understood by listeners.[11]

Reduced stress is thought to be the second most deviant speech dimension in hypokinetic dysarthria[6]; therefore, the effect of PD on prosodic characteristics and the detailed description of stress patterns in parkinsonian speech are of principal concern in this investigation. To further examine this issue, we chose a speaking task of *contrastive stress* because lexical stress is not inherent in all languages. Contrastive stress refers to a production task in which the information conveyed is altered by the location of syllabic stress. As acoustic analysis has the potential to provide a cheap, precise, and noninvasive method for the evaluation and support of speech therapy, further aims of the present study were to verify the suitability of commonly used measurements for the evaluation of stress in PD and to design an innovative measurement that would reflect the effect of all main acoustic cues exploited during stress production. We hypothesized that PD subjects would have a reduced ability to convey contrastive stress and show abnormal patterns of pitch, amplitude, and duration in both stressed and nonstressed tokens.

PATIENTS AND METHODS

The participants of this study were originally recruited as a part of an earlier study.[27] No study of contrastive stress has been previously published on the current participants. A total of 36 male Czech native speakers volunteered for the study. The PD group consisted of 20 individuals with idiopathic PD, whose age ranged from 34 to 82 years (mean = 60.5; standard deviation [SD] = 11.3). The diagnosis of idiopathic PD was made in a specialized center and was based on accepted criteria.[28] All patients were recruited immediately after the diagnosis was established and before symptomatic treatment was started. Before the recording procedure, each patient underwent a neurologic examination including the Unified Parkinson's Disease Rating Scale part III (UPDRS III, an objective measure of parkinsonian motor signs, ranging from 0 to 108, where a higher score indicates more severe disability), and Hoehn and Yahr (HY) staging (ranging from 1 to 5, where a higher stage indicates more severe disability). In our patients, the UPDRS III score ranged from 5 to 32 (mean = 17.8; SD = 7.2) and the HY ranged from 1 to 3 (mean = 2.2; SD = 0.5). In addition, the estimated duration of PD manifestations before the examination was surveyed and ranged from 6 to 82 months (mean = 31.9; SD = 21.4). No patient had a history of speech, language, or hearing disorders unrelated to parkinsonian symptoms or underwent speech-language treatment before participation in this study. All subjects were free of depression and cognitive deficits that could interfere with the measurements.

The healthy control (HC) group consisted of 16 male participants of comparable age, ranging from 36 to 80 years (mean = 61.8; SD = 13.3). None of these individuals reported a history of neurologic difficulties or any disorders that may affect speech, language, or hearing. No significant differences in age distribution were detected between the PD and HC groups. The study was approved by the local ethics committee and all participants provided written, informed consent for the speaking task and recording procedure.

The recordings were obtained during one session with a speech therapist who conveyed instructions to the subjects. Each participant completed a series of speaking tasks as part of the larger protocol. There were no time limits during the recordings. All participants were asked to repeat their performance at any time if they or the examiner were not fully satisfied with their initial attempt. The performance of the task including contrastive stress was selected for further investigation. The task was designed to evaluate whether speakers could highlight the semantic importance of information in their utterances using prosody. During the recording, each patient read a short block of text composed of five similar sentences and was required to unnaturally emphasize certain "key words" included in the text (Table 1). The first part of each sentence was variable and determined the linguistic context to identify the key word in the second part of the sentence, which was uniform and highlighted one of five key words (eg, *Dnes jsme to již nestihli*, možná ZÍTRA navštívíme všechny své známé./*Today we did not have enough time* but TOMORROW we will visit all our acquaintances; the part that determined the linguistic context of the sentence is indicated in italics). To ensure a better understanding of the task, the key words were underlined and written in capital letters and the entire task was demonstrated by a speech therapist (H.R.). As a result, for five different key words (ie, zítra/tomorrow; navštívíme/visit; všechny/all; své/our; and známé/acquaintances), we elicited one emphasized and four normally read tokens that were subjected to further investigation.

The speech samples were recorded in a quiet room with a low level of ambient noise using an external condenser microphone placed approximately 15 cm from the subject's mouth and coupled to a Panasonic NV-GS 180 video camera (Panasonic Corporation, Osaka, Japan). The audio data were digitized from the videotape to a computer at a sampling rate of 48 kHz and 16-bit quantization using original *Panasonic* software (Panasonic Corporation, Osaka, Japan).

Acoustic analysis was completed using the widely used specialized speech software *PRAAT* [available at: www.praat.org (Phonetic Sciences, University of Amsterdam, The Netherlands)].[29] To ensure the correctness of the automatic detecting procedure, the results of automatic analysis were verified by the examiner (T.T.) and manually adjusted if necessary. For the entire duration of each token, three standard acoustic parameters were assigned: fundamental frequency (F_0) in hertz; intensity in decibels; and duration in milliseconds. F_0 as well as intensity were expressed as the mean and range, that is, the difference between the maximum and minimum values. Duration was

TABLE 1.
Text of the Contrastive Stress Production Task Obtained and Recorded From all Participants

Key word	Key word Order	Part of Sentence	Text of Contrastive Stress Tokens
zítra	First	Context	Dnes jsme to již nestihli
		Stress	možná ZÍTRA *navštívíme všechny své známé*, zbude-li čas
navštívíme	Second	Context	I když jsme s nimi už mluvili telefonicky
		Stress	možná *zítra* NAVŠTÍVÍME *všechny své známé*, zbude-li čas
všechny	Third	Context	I když jsme se s některými známými již setkali
		Stress	možná *zítra navštívíme* VŠECHNY *své známé*, zbude-li čas
své	Fourth	Context	Jeho známé jsme již navštívili
		Stress	možná *zítra navštívíme všechny* SVÉ *známé*, zbude-li čas
známé	Fifth	Context	Příbuzné jsme již navštívili
		Stress	možná *zítra navštívíme všechny své* ZNÁMÉ, zbude-li čas

Notes: Words written in capital letters denote tokens upon which speakers were predisposed by the examiner to place emphasis. Normally read tokens were acquired for comparative purposes and are here denoted by italics.

measured as the total period between the onset and offset of each word. In addition, we developed an innovative parameter termed the *stress pattern index* (SPI), which was designed to reflect the effect of all three distinct acoustic characteristics exploited during stress production. To minimize the effects of individual differences such as loudness of voice, the intensity range was normalized by adjusting the minimal amplitude to the reference 0 dB before SPI calculation. The SPI is defined as the cumulative sum of signal energy (E_n) multiplied by the logarithmic expression of the maximum F_0 divided by the minimum F_0, that is, $\mathrm{SPI} = (1 + \ln(F_{0_{\max}}/F_{0_{\min}})) \sum E_n$. To minimize the effect of intersubject variability, such as individual pitch range of 70–120 Hz compared with 120–200 Hz, the logarithmic expression of the F_0 was also arranged.[27,30] Figure 1 shows the mechanism of stress production leading to the design of the SPI, demonstrated on a sample of nonstressed and stressed words. It can be seen that the mutual effect of exaggerated pitch, intensity, and duration is well represented by the SPI, where both increased intensity range and duration contribute to a greater cumulative sum of intensity and the overall effect is further amplified when the cumulative sum is multiplied by pitch range.

Statistics

Before statistical analyses, the assigned values of every individual parameter were divided into two groups for each participant and key word: the first group, indexed as "stressed," contained one emphasized token; the second group, indexed as "nonstressed," contained an average value calculated from four normally read tokens. As acoustic variables were normally distributed (Kolmogorov-Smirnov test), statistical analyses were performed using $2 \times 2 \times 5$ repeated measures analysis of variance (RM-ANOVA) involving the factors of SPEAKER GROUP (HC and PD), STRESS CONDITION (stressed and nonstressed), and KEY WORD ORDER (first, second, third, fourth, and fifth). *Post hoc* Bonferroni adjustment was used to determine between-group differences. A nominal alpha level was adjusted at .05.

RESULTS

The results for each individual speech variable as a function of key word order for each group (HC and PD) and both contrastive stress conditions (stressed and nonstressed) are presented in Figure 2. To examine the acoustic correlates of contrastive stress in the production experiment, RM-ANOVA was conducted for each dependent variable. The main effects and interactions may be interpreted as follows: (a) a main effect of SPEAKER GROUP would be statistically significant if PD speakers had already manifested impaired speech performance and, therefore, were not able to use the acoustic cues as effectively as HC subjects; (b) a main effect of STRESS CONDITION would be statistically significant if the parameter had been conveying stress; (c) a main effect of KEY WORD ORDER would be statistically significant if the parameter varied depending on the location of the key word within a sentence; (d) an interaction involving SPEAKER GROUP and STRESS CONDITION would be statistically significant if PD speakers were not able to express stress as effectively as HC speakers; (e) an interaction involving SPEAKER GROUP and KEY WORD ORDER would be statistically significant if some word was more suitable for differentiation between groups; and (f) an interaction involving STRESS CONDITION and KEY WORD ORDER would be statistically significant if some word was more appropriate for signaling stress. All statistically significant findings and their interpretations are listed below.

Mean F_0

Statistical analyses of mean F_0 revealed a significant main effect of STRESS CONDITION [$F(1,34) = 38.79$, $P < 0.001$]

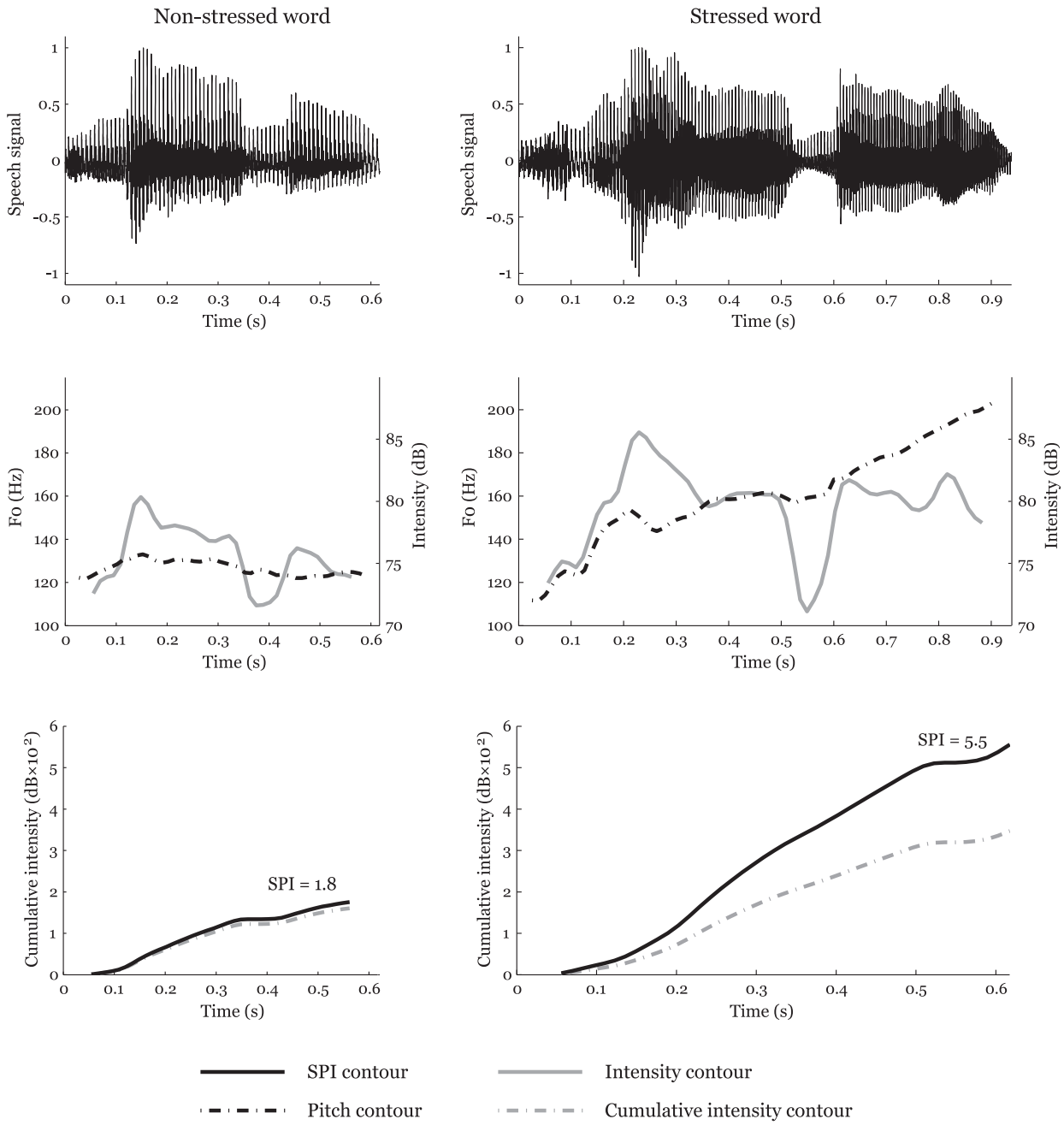


FIGURE 1. Mechanism of stress production leading to the design of the SPI shown on a sample of the nonstressed and stressed word "známé."

and KEY WORD ORDER [$F(4,136) = 6.58$, $P < 0.001$]. Surprisingly, there was no effect of SPEAKER GROUP. This finding suggests that the mean F_0 is elevated under stress conditions and varies across key word positions but is inappropriate for differentiation between PD and HC speakers.

Mean intensity

Considering the measurement of mean intensity, RM-ANOVA showed a significant main effect of STRESS CONDITION [$F(1,34) = 22.81$, $P < 0.001$] and SPEAKER GROUP [$F(1,34) = 11.34$, $P < 0.001$] that was attributable to PD speakers producing tokens with a significantly lower intensity than HC speakers.

Duration

A statistically significant main effect of duration was revealed for SPEAKER GROUP [$F(1,34) = 35.43$, $P < 0.001$], STRESS CONDITION [$F(1,34) = 611.88$, $P < 0.001$], and KEY WORD ORDER [$F(4,136) = 121.69$, $P < 0.001$]. A similar pattern of signaling contrastive stress was evident in both groups. Within a sentence, the stressed words were dependent on the original length approximately 30–45 milliseconds longer than nonstressed words. In addition, an interaction between STRESS CONDITION and KEY WORD ORDER [$F(4,136) = 5.71$, $P < 0.001$] was found. It is worthy to mention that one of the key words was monosyllabic, three were disyllabic, and one was quadrisyllabic. Therefore, as one might expect, the

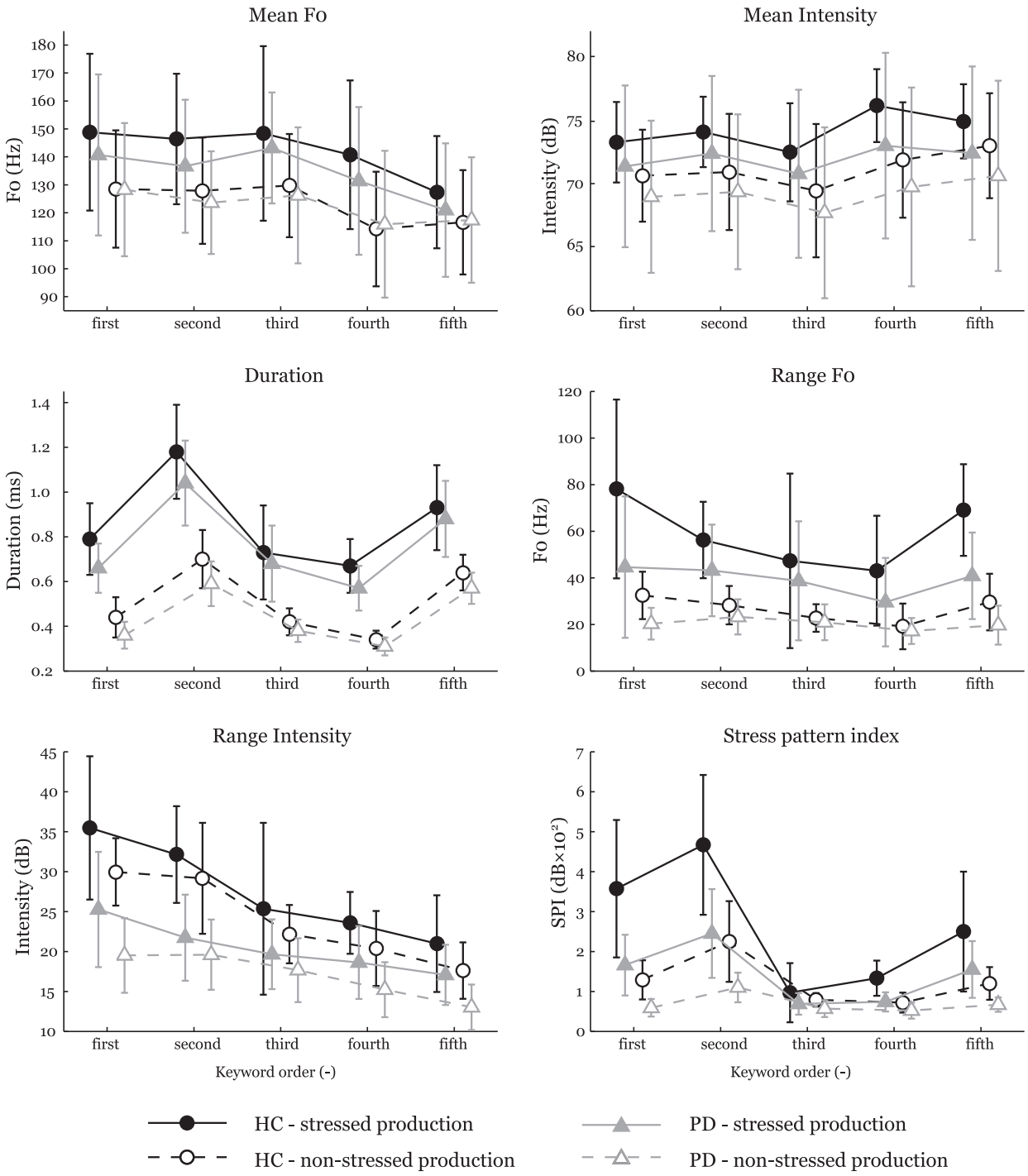


FIGURE 2. Comparison of speech measurements between parkinsonian speakers and HCs. Mean values and SDs (error bars) are depicted for both groups (HC and PD) as well as stress conditions (stressed and nonstressed), presented as a function of key word order.

multisyllabic words encouraged greater elongation during stress production than shorter words.

F_0 range

For the measurement of F_0 range, we revealed a main effect of SPEAKER GROUP [$F(1,34) = 39.20$, $P < 0.001$], STRESS CONDITION [$F(1,34) = 158.65$, $P < 0.001$], and also their interaction: SPEAKER GROUP × STRESS CONDITION [$F(1,34) = 10.35$, $P = 0.0014$]. Both PD and HC speakers markedly increased the F_0 range to indicate contrastive stress; however, PD speakers were less successful in elevating the F_0 range. In fact, when they were required to emphasize a key word within a sentence, they produced an F_0 range that was approximately 10–20 Hz narrower compared with that of HC speakers. We further found a main effect of KEY WORD ORDER [$F(4,136) = 8.12$, $P < 0.001$] that can be explained by consistent narrowing of the F_0 range, corresponding to the order in which the key words appeared (ie, first key word F_0

range > second key word F_0 range > third key word F_0 range > fourth key word F_0 range). The slight increase in F_0 range presented in the last key word may normally occur in the final word of a Czech declarative sentence and helps listeners to identify the boundaries between individual clauses.

Intensity range

Considering the measurement of intensity range, we proved a significant main effect of SPEAKER GROUP [$F(1,34) = 143.79$, $P < 0.001$], STRESS CONDITION [$F(1,34) = 37.78$, $P < 0.001$], and also KEY WORD ORDER [$F(4,136) = 44.10$, $P < 0.001$] that was due to a significant difference of intensity range between the first and third word, the second and fourth word, and also the third and fifth word. Taken together, both PD and HC utterances were characterized by a constantly narrowing intensity range within a sentence, which is assumed to be typical for the declarative clauses. Moreover, we found a significant SPEAKER GROUP × KEY WORD ORDER [$F(4,136) = 5.21$, $P < 0.001$] interaction. Regardless of stress condition, the intensity range of the first and second word produced by PD speakers was approximately 10 dB narrower, whereas the last three words of the sentence were only 5 dB narrower than those of HC speakers.

Stress pattern index

With respect to the measurement of SPI, we revealed a main effect of SPEAKER GROUP [$F(1,34) = 106.67$, $P < 0.001$], STRESS CONDITION [$F(1,34) = 152.38$, $P < 0.001$], and also their interaction: SPEAKER GROUP × STRESS CONDITION [$F(1,34) = 14.16$, $P < 0.001$]. Both PD and HC speakers markedly increased SPI values to signal contrastive stress; however, PD subjects achieved significantly lower values. Furthermore, we found a main effect of KEY WORD ORDER [$F(4,136) = 66.13$, $P < 0.001$] and also its interactions: KEY WORD ORDER × SPEAKER GROUP [$F(4,136) = 10.49$, $P < 0.001$] as well as KEY WORD ORDER × STRESS CONDITION [$F(4,136) = 15.49$, $P < 0.001$]. These interactions may be explained by individual distinctions in length and voicelessness among particular key words. In fact, between-group differences arose during the first, second, and last key word performances when compared with the productions of key words in the middle positions.

DISCUSSION

Reduced stress is thought to be one of the most deviant speech dimensions in hypokinetic dysarthria associated with PD.[6] Using objective assessment by acoustic analysis, the present study strives to contribute to a better understanding of prosodic cues relating to stress production in PD and to determine how these patterns differ from those of healthy speakers. Contrary to previous reports,[23–26] this study is not only limited to the evaluation of stress patterns using standard measurements such as amplitude, pitch, and duration but also introduces an innovative measurement termed the SPI that is designed to reflect the effect of all basic acoustic cues exploited during stress production. Our results indicate that persons in the early stages of PD maintain the ability to signal contrastive stress using exaggeration of pitch, intensity, and duration. However, PD subjects were not able to modulate these acoustic cues to the same extent as HC. From this point of view, the measurement of pitch range and SPI proved their feasibility for the assessment of contrastive stress.

Our general findings relating to parkinsonian speech, regardless of stress prominence, are consistent with previous studies reporting dysprosody as a common part of speech deficits experienced by PD patients.[27,31,32] In comparison with HC, PD speakers exhibited a significantly lower intensity, narrower pitch range, reduced intensity range, as well as shorter word duration. Despite these discrepancies, the within-sentence patterns of PD performances were similar to those of HC speakers for all measurements, with the exception of F_0 range (Figure 2). For example, as the average pitch contours of HC utterances fell toward the end of the sentence, PD utterances showed the same trend. This intonation drop at the end of a sentence is typical for declarative clauses. These findings maintain the hypothesis that (a) prosodic cues vary depending on the position of the stressed word within an utterance because these cues are used to convey multiple linguistic functions[10,12] and (b) PD speakers are able to naturally alter prosodic cues to express a number of linguistic meanings. It is also noteworthy to mention the measurement of F_0 range because the pitch contours of PD productions were notably flatter than those of HC, especially at the beginning and end of a phrase. These findings support the general conclusion that monopitch is the primary feature of dysprosody in PD,[27,33] followed by monoloudness and speech rate abnormalities.

In agreement with previous research,[26] our results relating to stress production in PD indicate that although patients in the early stages of the disease have a reduced ability to convey contrastive stress, they can still notably increase pitch, intensity, and duration to emphasize a word within a sentence. In fact, we did not observe any significant differences between PD and HC stress productions using the measurements of F_0, intensity, duration, and intensity range. However, restricted F_0 range was evident in the PD group, particularly due to the significantly lower values in the first and last words of a phrase. Remarkably, Pell et al[11] investigated how listeners perceive the contrastive stress utterances produced by PD speakers and observed that listeners were least accurate at identifying sentence initial- and final-word emphasis compared with middle-word emphasis.

One further aim of the present study was to verify the feasibility of commonly used measurements for the evaluation of stress in PD speech. According to our data, only the measurement of F_0 range seems to be suitable for the assessment of stress patterns because none of the other measurements (mean F_0, mean intensity, intensity range, or duration) led to differentiation between PD and HC stress production. These findings are not in complete agreement with previous research by Cheang and Pell,[26] which reported intensity as the most robust parameter, followed by F_0. However, the authors subjected all acoustic data to several normalizations, which makes a comparison of their findings with those of the present study hardly possible. Finally, the innovative parameter SPI achieved the best result in separating PD and HC groups and proved its suitability

for the evaluation of stress patterns in hypokinetic dysarthria. It may be further concluded that, in comparison with HC subjects, PD speakers have greater difficulties in altering pitch, intensity, and duration when required to modify them simultaneously, suggesting somewhat restricted phonatory flexibility.

During stress production, PD subjects in the present study were able to consciously improve their speech performance using multiple acoustic cues including increased pitch and loudness as well as prolonged duration. Similar results were noted by Goberman and Elmer,[34] who investigated clear versus conversational speech in PD individuals with varying severity of dysarthria, based on analysis of five acoustic characteristics. It has been demonstrated that PD patients are able to consciously achieve significant improvements in articulation rate, average F_0, and speaking F_0 variability with the simple request to produce speech clearly.[34] These findings, altogether, may indicate the important role of voice and speech therapy in the course of PD. Currently, the Lee Silverman Voice Treatment (LSVTLOUD, LSVT Global Inc., Tuscon, AZ) has been established as an efficacious behavioral treatment for voice and speech problems associated with PD.[35–37] Furthermore, several studies have reported that the effect of an acquired increase in vocal loudness relating to LSVTLOUD might be generalized to improvements not only in voice production but also in speech articulation and intelligibility.[38,39] Accordingly, contrastive stress drills might be used in speech therapy to increase prosodic adequacy, intelligibility, and communicative efficacy of persons suffering from motor speech disorders.[40]

Admittedly, there are some limitations to the present study. Because previous studies[41,42] have proposed that gender may have an impact on prosodic characteristics and our study consisted of only male participants, we cannot exclude that stress production may be influenced by gender-specific aspects of speech. Another point to consider is that only a contrastive stress task was used in the present investigation, and therefore, we cannot provide any information regarding possible variable patterns of speech impairment depending on the type of stress-based speaking task.[26]

In conclusion, the present study illustrates the potential of acoustic analyses to document stress patterns in hypokinetic dysarthria of PD. Although PD speakers used the same acoustic cues and strategies as HC, they exhibited a reduced ability to convey contrastive stress. However, well-controlled prosody may contribute to better speech intelligibility, and therefore, the effect of medical intervention and speech therapy on stress patterns should be investigated in future studies. In addition, a qualitative description of discrepancies in stress production may be helpful in differential diagnosis, as abnormalities in stress patterns were observed to vary depending on the subtype of dysarthria.

Acknowledgments

We are obliged to doctors Jan Roth, Jiri Klempir, Veronika Majerova, and Jana Picmausova for provision of clinical data.

REFERENCES

1. Hornykiewicz O. Biochemical aspects of Parkinson's disease. *Neurology*. 1998;51:S2–S9.
2. De Rijk MC, Tzourio C, Breteler MM, et al. Prevalence of parkinsonism and Parkinson's disease in Europe: The EUROPARKINSON collaborative study. European Community Concerted Action on the Epidemiology of Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 1997;62:10–15.
3. Ho AK, Iansek R, Marigliani C, Bradshaw J, Gates S. Speech impairment in a large sample of patients with Parkinson's disease. *Behav Neurol*. 1998;11:131–137.
4. Logemann JA, Fisher HB, Boshes B, Blonsky ER. Frequency and cooccurrence of vocal tract dysfunction in the speech of a large sample of Parkinson patients. *J Speech Hear Disord*. 1978;43:47–57.
5. 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;135:1860–1870.
6. Darley FL, Aronson AE, Brown JR. Differential diagnostic patterns of dysarthria. *J Speech Hear Res*. 1969;12:246–269.
7. Duffy JR. *Motor Speech Disorders: Substrates, Differential Diagnosis and Management*. 2nd ed. New York, NY: Mosby; 2005.
8. Miller N, Noble E, Jones D, Burn D. Life with communication changes in Parkinson's disease. *Age Ageing*. 2006;35:235–239.
9. Klatt DH. Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *J Acoust Soc Am*. 1976;59:1208–1221.
10. Cooper WE, Stephen EJ, Pamela MR. Acoustical aspects of contrastive stress in question-answer contexts. *J Acoust Soc Am*. 1985;77:2142–2156.
11. Pell MD, Cheang HS, Leonard CL. The impact of Parkinson's disease on vocal-prosodic communication from the perspective of listeners. *Brain Lang*. 2006;97:123–134.
12. Bolinger DL. A theory of pitch accent in English. *Word*. 1958;14:109–149.
13. Fry D. Duration and intensity as physical correlates of linguistic stress. *J Acoust Soc Am*. 1955;27:765–768.
14. Lehiste I. Suprasegmental features of speech. In: Lass NJ, ed. *Contemporary Issues in Experimental Phonetics*. New York, NY: Academic Press; 1976:225–235.
15. Fry D. Experiments in the perception of stress. *Lang Speech*. 1958;81:126–152.
16. Morton J, Jassem W. Acoustic correlates of stress. *Lang Speech*. 1965;8:159–181.
17. O'Shaughnessy D. Linguistic features in fundamental frequency patterns. *J Phon*. 1979;7:119–145.
18. Sluijter AMC, Van Heuven VJ, Pacilly JA. Spectral balance as a cue in the perception of linguistic stress. *J Acoust Soc Am*. 1997;101:503–513.
19. Howell P. Cue trading in the production and perception of vowel stress. *J Acoust Soc Am*. 1993;94:2063–2073.
20. Fidelholtz JL. Stress in Polish—with some comparisons to English stress. *Papers and Studies in Contrastive Linguistics*. 1979;9:47–61.
21. Ladd RD, Beckman EM, Pierrehumbert JB. Intonational structure in Japanese and English. *Phonology*. 1986;3:255–309.
22. Liss JM, Weismer G. Selected acoustic characteristics of contrastive stress in control geriatric, apraxic and ataxic dysarthria speakers. *Clin Linguist Phon*. 1994;8:43–66.
23. Patel R, Campellone P. Acoustic and perceptual cues to contrastive stress in dysarthria. *J Speech Lang Hear Res*. 2009;52:206–222.
24. Wang YT, Kent RD, Duffy JR, Thomas JE. Dysarthria associated with traumatic brain injury: speaking rate and emphatic stress. *J Commun Disord*. 2005;38:231–260.
25. Ma JKY, Whitehill TL, So SYS. Intonation contrast in Cantonese speakers with hypokinetic dysarthria associated with Parkinson's disease. *J Speech Lang Hear Res*. 2010;53:836–849.
26. Cheang HS, Pell MD. An acoustic investigation of Parkinsonian speech in linguistic and emotional contexts. *J Neurolinguist*. 2007;20:221–241.
27. Rusz J, Cmejla R, Ruzickova H, Ruzicka E. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *J Acoust Soc Am*. 2011;129:350–367.
28. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnoses of idiopathic Parkinson's disease. A clinic-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry*. 1992;55:181–184.
29. Boersma P, Weenink D. Praat, a system for doing phonetics by computer. *Glot Int*. 2001;5:341–345.

30. Moore BCJ. *An Introduction to the Psychology of Hearing*. 5th ed. Boston, MA: Academic Press; 2003.
31. Caekebeke JFV, Jennekens-Schinkel A, Van der Linden ME, Buruma OJS, Roos RAC. The interpretation of dysprosody in patients with Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 1991;54:145–148.
32. Skodda S, Grönheit W, Schlegel U. Intonation and speech rate in Parkinson's disease: general and dynamic aspects and responsiveness to levodopa admission. *J Voice*. 2011;25:199–205.
33. Rusz J, Cmejla R, Ruzickova H, et al. Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test. *Mov Disord*. 2011;26:1951–1952.
34. Goberman AM, Elmer LW. Acoustic analysis of clear versus conversational speech in individuals with Parkinson disease. *J Commun Disord*. 2005;38:215–230.
35. Pinto S, Ozsancak C, Tripoliti E, Thobois S, Limousin-Dowsey P, Auzou P. Treatments for dysarthria in Parkinson's disease. *Lancet Neurol*. 2004;3:547–556.
36. Ramig LO, Countryman S, Thompson LL, Horii Y. Comparison of two forms of intensive speech treatment for Parkinson disease. *J Speech Hear Res*. 1995;38:1232–1251.
37. Spielman J, Mahler L, Halpern A, Gilley P, Klepitskaya O, Ramig L. Intensive voice treatment (LSVT®LOUD) for Parkinson's disease following deep brain stimulation of the subthalamic nucleus. *J Commun Disord*. 2011;44:688–700.
38. Halpern A, Spielman J, Ramig L, Sharpley A, Panzer I. The effects of loudness and noise on speech intelligibility in Parkinson disease. Presented at: 14th Biennial Conference on Motor Speech. Monterey, CA. March 6,2008.
39. Sapir S, Spielman JL, Ramig LO, Story BH, Fox C. Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: acoustic and perceptual findings. *J Speech Lang Hear Res*. 2007;50:899–912.
40. Murdoch BE. *Dysarthria: A Physiological Approach to Assessment and Treatment*. 2nd ed. Cheltenham, United Kingdom: Nelson Thornes; 1998.
41. Hertrich I, Ackermann H. Gender-specific vocal dysfunction in Parkinson's disease: electroglottographic and acoustic analyses. *Ann Otol Rhinol Laryngol*. 1995;104:197–202.
42. Skodda S, Visser W, Schlegel U. Gender-related patterns of dysprosody in Parkinson's disease and correlation between speech variables and motor symptoms. *J Voice*. 2011;25:76–82.

Appendix A5

Automatic evaluation of articulatory disorders in Parkinson's disease

Novotny M, Rusz J, Cmejla R, Ruzicka E (2014) Automatic evaluation of articulatory disorders in Parkinson's disease. *IEEE/ACM T Audio Speech Lang Process* **22**:1366-1378.

Journal name (category): **IEEE/ACM Transactions on Audio Speech and Language Processing** (acoustics/engineering, electrical & electronic)
Quartile in category: **Q1/Q1**
Impact factor: **2.625**
Number of WoS citations (excluding autocitations): **11**

Automatic Evaluation of Articulatory Disorders in Parkinson's Disease

Michal Novotný, Jan Rusz, Roman Čmejla, and Evžen Růžička

***Abstract*—Although articulatory deficits represent an important manifestation of dysarthria in Parkinson's disease (PD), the most widely used methods currently available for the automatic evaluation of speech performance are focused on the assessment of dysphonia. The aim of the present study was to design a reliable automatic approach for the precise estimation of articulatory deficits in PD. Twenty-four individuals diagnosed with de novo PD and twenty-two age-matched healthy controls were recruited. Each participant performed diadochokinetic tasks based upon the fast repetition of /pa/-/ta/-/ka/ syllables. All phonemes were manually labeled and an algorithm for their automatic detection was designed. Subsequently, 13 features describing six different articulatory aspects of speech including vowel quality, coordination of laryngeal and supralaryngeal activity, precision of consonant articulation, tongue movement, occlusion weakening, and speech timing were analyzed. In addition, a classification experiment using a support vector machine based on articulatory features was proposed to differentiate between PD patients and healthy controls. The proposed detection algorithm reached approximately 80% accuracy for a 5 ms threshold of absolute difference between manually labeled references and automatically detected positions. When compared to controls, PD patients showed impaired articulatory performance in all investigated speech dimensions ($p < 0.05$). Moreover, using the six features representing different aspects of articulation, the best overall classification result attained a success rate of 88% in separating PD from controls. Imprecise consonant articulation was found to be the most powerful indicator of PD-related dysarthria. We envisage our approach as the first step towards development of acoustic methods allowing the automated assessment of articulatory features in dysarthrias.**

***Index Terms*—Acoustic analysis, automatic segmentation, diadochokinetic task, hypokinetic dysarthria, Parkinson's disease, speech disorders.**

Manuscript received November 12, 2013; revised April 13, 2014; accepted June 01, 2014. This work was supported in part by the Czech Grant Agency under Grant 102/12/2230, the Czech Ministry of Health under Grant NT14181-3/2013, and Charles University in Prague under Grant No. PRVOUK-P26/LF1/4. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Robert Togneri.

M. Novotný and R. Čmejla are with the Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, 160 00 Prague 6, Czech Republic (e-mail: novotm26@fel.cvut.cz; cmejla@fel.cvut.cz).

J. Rusz is with the Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, 160 00 Prague 6, Czech Republic, and also with the Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University in Prague, 120 00 Prague 2, Czech Republic (e-mail: ruszjan@fel.cvut.cz).

E. Růžička is with the Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University in Prague, 120 00 Prague 2, Czech Republic (e-mail: eruzi@lf1.cuni.cz).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASLP.2014.2329734

I. Introduction

PARKINSON's disease (PD) is a progressive, idiopathic disorder which primarily affects dopaminergic neurons in the substantia nigra pars compacta and causes dopaminergic striatal loss [1]. Low levels of dopamine lead to dysfunction of the basal ganglia and primarily account for motor deficits. The cardinal features of PD include tremor at rest, rigidity, bradykinesia, and postural instability. In addition to the most common motor manifestations, other non-motor manifestations such as autonomic dysfunction, cognitive and neurobehavioral abnormalities, sleep alterations, and sensory disruptions may be evident [2], [3].

The diagnosis of PD is based upon the presence of primary motor manifestations, which develop after 60–70% of dopaminergic neurons degenerate and dopamine levels are reduced by 80% [3], [4]. Due to slow, gradually progressive nature of the disorder, PD has a prodromal interval during which the main motor manifestations are not clearly evident. The duration of the PD prodromal period has been documented as 3-15 years [5]. Although some dopamine agonists may have a neuroprotective effect [6], pharmacotherapy and neurosurgical interventions that are currently available only offer alleviation of certain parkinsonian manifestations. Despite the fact that medication generally prolongs active life expectancy, the effect of treatment depends upon the stage of the disease during which it is initiated. Furthermore, there is no treatment that can cure PD or halt its progression. Therefore, the early diagnosis of PD plays a vital role in improving the patient's quality of life [7], [3].

Several studies have found speech to be one of the earliest disrupted modalities in PD [5], [8]. In addition, previous research has indicated that up to 90% of PD sufferers display vocal impairment [9], with the most salient impact on phonatory and articulatory features of speech [10]. These vocal deficits can be generally described as hypokinetic dysarthria [11], [12]. Signs of hypokinetic dysarthria involve reduced loudness, breathiness, roughness, decreased energy in the higher parts of the harmonic spectrum, exaggerated tremor, imprecise articulation of vowels and consonants, monopitch, monoloudness, disturbances in speech timing, and dysrhythmia, which together lead to overall reduced speech intelligibility [13]–[16].

The analysis of speech is therefore an attractive method for monitoring disease onset and progression, as well as treatment efficacy [5], [8], [13], [17]. Recent studies have identified speech analysis as an affordable, objective and widely available approach, which could significantly reduce demands on PD

patient investigation [13], [18]. A wide range of speech tests, including fast syllable repetition, sustained phonation, various readings and freely spoken monologue have been designed to assess the extent of speech manifestations. To precisely analyze speech performances, recorded utterances are commonly subjected to traditional methods including the assessment of sound pressure levels, fundamental frequency, formant frequencies, speech rate and rhythm [19]–[24].

Increasing computational power is currently leading to higher levels of automation, and therefore, novel methods of automatic speech analysis have been introduced [13], [18], [25]. However, due to the confounding effects of articulatory and linguistic components, new approaches for automatic speech analysis are often limited to the use of sustained phonation, enabling the measurement of dysphonic aspects of speech [13], [18]. Nevertheless, the importance of articulatory knowledge in dysarthric speech recognition has been noted [15], and hypokinetic dysarthria in PD is primarily a disorder of articulation affecting various aspects of speech [16].

According to previous research [26], PD-induced articulatory impairment may be clearly apparent when patients perform diadochokinetic (DDK) tasks. The most typical DDK utterance includes repetition of the /pa/-/ta/-/ka/ syllable train. This DDK task is widely preferred because it consists of fast syllable-train repetitions with bilabial, alveolar and velar places of articulation [27]. Such an approach requires complex movements of the articulators (lips, jaws, and tongue) during a task with well-defined structure, which contributes to a reduction in data processing complexity. Such tasks may allow the automatic detection of a variety of relevant features that would be difficult to reliably assess from running speech. For example, one of the most common signs of dysarthria in PD is imprecise consonant coordination, which can be evaluated using voice onset time (VOT), typically determined as the duration between an initial burst and vowel onset [28]. Although the assessment of consonant articulation contributes significantly to an accurate, subjective diagnosis of dysarthria, to the best of our knowledge, there is no algorithm for the automatic detection of VOT in dysarthric speakers [22], [29].

The main goal of the present study was therefore to develop an automatic segmentation algorithm allowing the accurate detection of the initial burst, vowel onset, and occlusion. Using the proposed segmentation algorithm, we further endeavored to introduce several acoustic features sensitive to possible articulatory deficits due to dysarthria. To explore the suitability of the designed acoustic features in capturing parkinsonian articulatory disorder, an additional aim was to propose classification experiment in order to differentiate PD subjects from controls.

The present text is divided into several sections as follows: The "Methods" section comprises a description of the recruited subjects, recorded utterances, data processing, statistical evaluation and classification experiment. The "Results" section evaluates the performance of the automatic segmentation, presents the correlation between the obtained results and reference hand labels, illustrates the statistical significance of each feature and lists success rates of the classification. The "Discussion and Conclusion" sections provide a discussion and summary of our general findings.

II. Methods

A. Subjects

Data were collected as part of an original study [30]; the methods of automatic segmentation as well as speech characteristics based on automatic segmentation have not previously been reported. Recordings were obtained from 46 native Czech speakers with no history of speech therapy. The PD group consisted of 24 participants (20 men, 4 women), all of whom fulfilled the diagnostic criteria for PD[1] [31]. All PD speakers were examined immediately after the diagnosis was made and before symptomatic treatment was initiated. The mean age of PD participants was $60.9\pm$ standard deviation 12.6 years, mean disease duration 31.3 ± 22.3 months, disease stage 2.2 ± 0.5 according to the Hoehn & Yahr (H&Y)[2] scale [32], mean motor score 17.4 ± 7.1 according to the Unified Parkinson's Disease Rating Scale (UPDRS) III[3][33]. In agreement with perceptual evaluations based on UPDRS III item 18[4], 13 patients obtained a score of 0 and 11 patients a score of 1, suggesting no-to-mild speech impairment. None of the PD patients reported previous speech disorders unrelated to the present illness.

The healthy control (HC) group was comprised of 22 volunteers (15 men, 7 women; mean age 58.7 ± 4.6 years) with no history of neurological disease. No differences in age between the PD and HC groups were observed (two-sample t-test; $t(44) = -0.89$, confidence interval $(CI) = [-10.13, 3.94]$, $p = 0.38$). All participants (PD and HC) had no history of speech therapy. The study was approved by the Ethics Committee of the General University Hospital in Prague and all participants provided written, informed consent.

B. Protocol

C. 1) Recording

Recordings were taken in a quiet room with a low ambient noise level using a condenser microphone at a distance of approximately 15 cm from the subject's mouth. Data were transferred to a personal computer with a sampling frequency of 48 kHz and 16 bit quantization. All participants were recorded in an examination room within the neurological department. All PD patients were recorded shortly after the diagnosis was established, before starting dopaminergic treatment. Each utterance was recorded during a single session by a speech-language

[1]UK Parkinson's Disease Society Brain Bank clinical diagnostic criteria consist of Step 1: presence of bradykinesia (slowness of initiation of voluntary movement with progressive reduction in speed and amplitude of repetitive actions) and at least one of the following: muscular rigidity, 4-6 Hz rest tremor, postural instability not caused by primary visual, vestibular, cerebellar, or proprioceptive dysfunction; Step 2: exclusion criteria for Parkinson's disease; and Step 3: supportive prospective positive criteria for Parkinson's disease including excellent response to levodopa.

[2]Hoehn & Yahr scale contains five grades of PD severity and is commonly used for the description of PD progression. The scale comprehends severity from a mild unilateral motor disorder as the first grade, to confinement to bed or wheelchair as the fifth grade.

[3]The UPDRS III is scaled from 0 to 108, with 0 for no motor manifestations and 108 representing severe motor manifestations.

[4]The UPDRS III item 18 is concerned with the assessment of speech, and is ranked from 0 to 4, where 0 represents normal speech; 1 slight loss of expression, diction and volume; 2 monotone slurred but understandable speech, moderately impaired; 3 marked speech impairment, difficult to understand; and 4 unintelligible speech.

TABLE I
LABELING CRITERIA BASED ON [22]

Position	Description	Frequency domain	Time domain
Initial Burst	Abrupt onset of noise energy caused by turbulent airflow during stop release. Good contrast in time and moderate energy contrast.	Moderate excitation of one or a few time windows of the spectrogram over the entire frequency range.	Used for specification of burst onset. In the case of multiple bursts the initial burst is marked [34].
Vowel Onset	Abrupt onset of periodic signal with highest acoustic energy caused by vocal fold vibration. Good contrast in time and best energy contrast.	Onset of fundamental (F0) and first formant frequencies (F1, F2, F3) [35], [36]. Energy is concentrated to these frequencies.	Position with highest contrast. If the abrupt onset of energy is not clearly apparent, the F0, F1, and F2 onset is sought.
Occlusion	Slow voice weakening, and therefore slow weakening of F0, F1, F2, and F3. Fuzzy due to weak time and energy contrast.	Energy of F0, F1, F2, and F3 slowly weakens. The F2-vowel offset is considered the best indicator of occlusion onset [37].	Used especially to boost the robustness of labeling. Needed especially due to slow energy weakening.

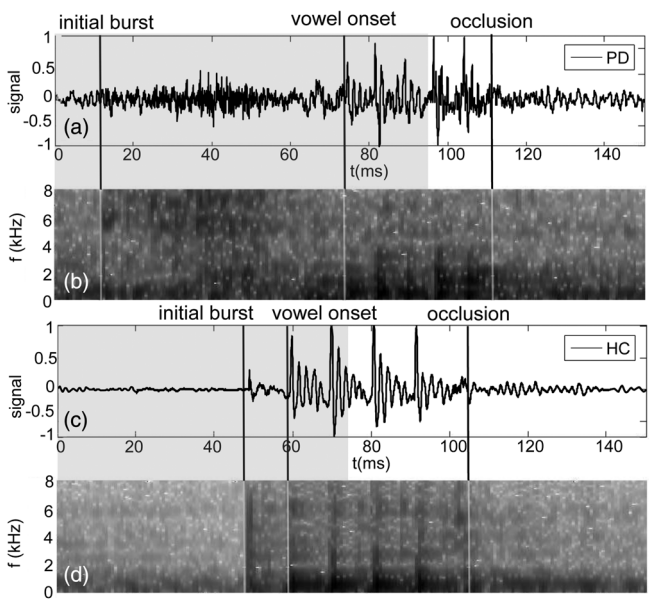


Fig. 1. Examples of syllable and its wideband spectrogram for Parkinsonian (a), (b) and healthy (c), (d) speakers, with marked positions of the initial burst, vowel onset and occlusion. The gray background shows the front part of the syllable and the white background refers to the rear part of the syllable.

pathologist. All participants were instructed to perform rapid, steady /pa/-/ta/-/ka/ syllable train repetitions as constantly and as quickly as possible, where each performance consisted of at least five syllable train repetitions. No time limits were imposed during the recording. Each participant, with the exception of three controls and three patients, repeated this task two times resulting in the acquisition of 80 utterances in total. As a result, a total of 1644 tokens (syllables) were collected, 753 for PD and 891 for HC.

2) Reference Labels: As can be seen in Fig. 1, each /pa/, /ta/, or /ka/ syllable consists of an initial burst, vowel onset, and occlusion. These three basic events generally describe the timing of articulation, and their positions must be detected in each syllable to analyze articulation deficits. Thus, reference labels must first be established, and this procedure requires the manual segmentation of each utterance. However, segmentation may be challenging even for manual labeling and therefore, the criteria according to which labeling was performed must be stated. Fischer and Goberman [22] summarize three basic rules based on previous research [34]–[37] which were used as a foundation for our labeling criteria (see Table I).

D. Algorithm for Automatic Segmentation

Manual labeling is a time consuming process and may be biased by subjective evaluation. To decrease time demands and provide objective results, a deterministic detector of the initial burst, vowel onset, and occlusion was designed. The algorithm is presented in several subsections describing pre-processing, rough segmentation, detection of the initial burst, vowel onset, and occlusion.

1) Pre-Processing: The pre-processing step consists of re-sampling the signal to 20 kHz, which lowers the computational complexity and maintains useful speech information [38]. The pre-processing step also includes DC offset removal and normalization of the signal to the interval $[-1, 1]$.

2) Rough Segmentation: The first problem encountered in automatic processing was the unknown number of syllables. This problem was solved by rough segmentation, which divided an utterance to single syllables (see Fig. 2). These syllables were then processed separately. To split the signal into single syllables, the approximate position of each syllabic nucleus had to be estimated. We may assume that in the DDK task, each syllable consists of one low-energy consonant and one high-energy vowel. Therefore, positions of syllabic nuclei may be identified by high-energy vowel peaks. However, the presence of a higher noise component in PD utterances may bias the nuclei search, and therefore filtering must be performed. Filtering was accomplished by a low-pass FIR filter with a linear phase and order of 500 with a 300 Hz cut-off frequency. The filtered signal is squared and smoothed by the moving average filter of order 800 and local energy maxima are detected. We noted that when one syllable has considerably lower energy than its neighboring syllables, detection based on 300 Hz filtering tends to omit the syllable. Hence, the same detection based on a low-pass filter with a 1000 Hz cut-off frequency was used. The detector based only on 1000 Hz filtering was more vulnerable to the higher noise component included in PD utterances, and therefore it was used only as a complement to more robust, 300 Hz filter-based detector. The maximal distance between two consecutive nuclei was estimated and enlarged 1.1 times, providing the length of a single syllable segment. This length was distributed before and behind the energy peaks, providing the approximate borders for each syllable.

To avoid false detections due to the high sensitivity of the detector, the elimination of false positions must be implemented as the second step of rough segmentation. The elimination was based on the comparison of high and low energy centroid positions obtained from the filtered spectrogram around the vowel

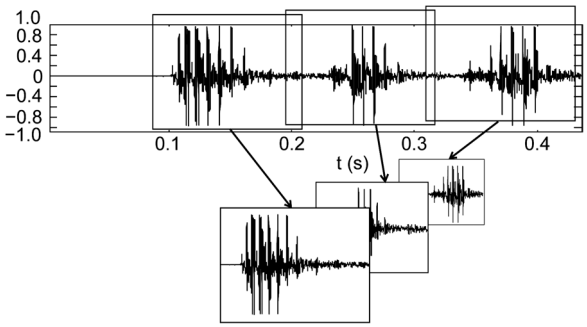


Fig. 2. Detail of an utterance divided by rough segmentation into single syllabic segments.

onset. Due to higher computational complexity, the spectrogram was also utilized during the detection of the initial burst, which was also spectrogram-based.

The spectrogram window length was defined as the length of the processed signal divided by 120, and the overlap was equal to one half of a window. To increase efficiency, the unnecessary rear part was omitted and the spectrogram was computed only from the front part of the syllable (gray part of the syllable highlighted in Fig. 1).

Spectrogram processing consisted of the elimination of negligible values and computation of energy envelopes. To determine which value was negligible, the spectrogram was treated as a matrix P with m rows for frequency bins and n columns for time bins. The threshold matrix $\mathbf{T}$ was an m by n matrix, where the i-th row was computed from the i-th frequency bin of the spectrogram according to equation (1).

$$\mathbf{T}(i, 1 \ldots n) = 0.8 \frac{1}{n} \sum_{k=0}^{n} \mathbf{P}(i, 1 \ldots n). \tag{1}$$

This equation sets each row of the threshold matrix $\mathbf{T}$ as the weighted mean value of energy contained in the equivalent frequency row of matrix $\mathbf{P}$. Filtering was then performed as shown in equation (2).

$$P_{RoughSegm}(i,j) = \begin{cases} P(i,j) & P(i,j) \geq T(i,j) \\ 0 & P(i,j) < T(i,j) \end{cases}, \tag{2}$$

where $P_{RoughSegm}(i,j)$ denotes element contained in the i-th frequency bin and the j-th time bin of the filtered matrix. An example of a filtered spectrogram can be seen in Fig. 3.

The next processing step was the computation of two energy envelopes. The first was calculated by summing the values in each column (Fig. 3(c)), while the second was determined by summing values only in the upper half of each column (Fig. 3(d)). The first envelope considers the high energy of vowels contained mostly in low frequencies; the second emphasizes high frequencies generated during the initial burst. Centroids were computed from these envelopes and their absolute and mutual positions were used for the elimination of false detections. The centroid positions are marked as black arrows in Fig. 3(c) and Fig. 3(d).

The energy envelope comprising the entire frequency bandwidth provides facilities for rough vowel onset estimation. The position of vowel onset was set as the first peak of the voicing periodic sequence. This approach was based on the assumption that, during the voicing, the vocal tract is excited by quasi-periodically repeating glottal pulses [39]. In processing the front part of the syllable (see Fig. 1), peaks may be traced from the end of the envelope (see Fig. 3(c)). However, this estimation sometimes marks the accentuated initial burst instead of vowel onset, and therefore, it is sufficient only for the correction of syllable position.

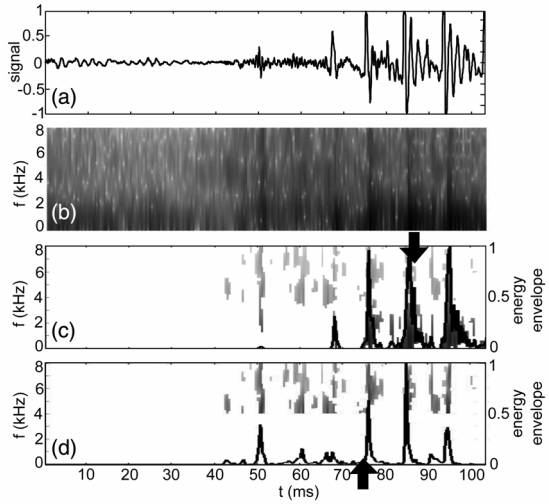


Fig. 3. Signal in the time domain (a), signal spectrogram (b), filtered spectrograms with marked energy envelopes and arrows pointing to spectral centroid positions in the entire frequency range (c), and the upper half of the frequency range (d).

3) Detection of the Initial Burst: After the elimination of false detections and correction of the segment borders, the noise burst connected with the initial stop release was sought. For the purposes of burst detection, the previously computed spectrogram was processed according to a modification of eq. (2) (see eq. (3)),

$$P_{InitialBurst}(i,j) = \begin{cases} 1 & P(i,j) \geq T(i,j) \\ 0 & P(i,j) < T(i,j) \end{cases}, \tag{3}$$

where the $\mathbf{T}$ matrix is given by eq. (1). The result of this filtering can be seen in Fig. 4.

The envelope, given by summing all values in each time window of the matrix $\mathbf{P_{InitialBurst}}$, emphasizes information about frequency bandwidth at the expense of information about energy distribution. This method emphasizes the noise burst, which has lower energy uniformly distributed through the entire spectrum. Furthermore, due to abrupt onset, the difference of the envelope highlights and specifies the stop release position as shown in Fig. 4.

4) Detection of Vowel Onset: The quasi-periodic character of a vowel with an abrupt onset of energy was detected using the Bayesian Step Changepoint Detector (BSCD) [40], [41]. In general, the BSCD assumes that (i) the signal is composed of two different constant values (e.g., 0.05 and 0.3 marked as lines in Fig. 5(b)), and (ii) that it is possible to calculate the posteriori probability of changes in the signal through Bayesian marginalization. Whereas the approach with the matrix $\mathbf{P_{InitialBurst}}$ emphasizes the abrupt noise burst, the assumption of signal being composed of two different constant steps emphasizes a boundary between two different signals.

The input of the detector represents the first part of the syllable from the initial burst to the end of the front part of the

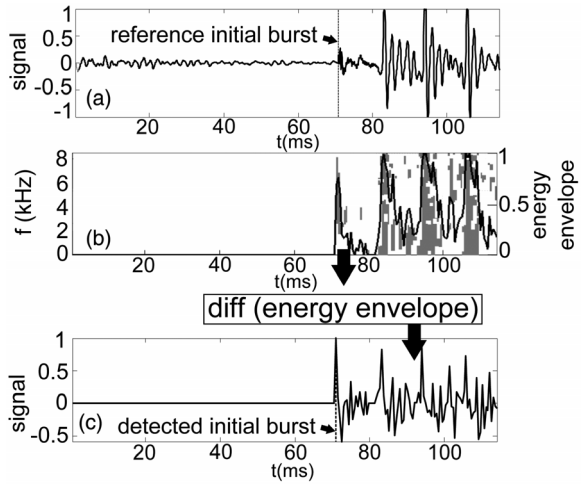


Fig. 4. Fig. 4 Front part of a syllable in the time domain (a), filtered spectrogram with the gray color denoting 1 and the white color denoting 0 and its marked energy envelope (b), and the normalized difference of the energy envelope used for the final initial burst detection (c).

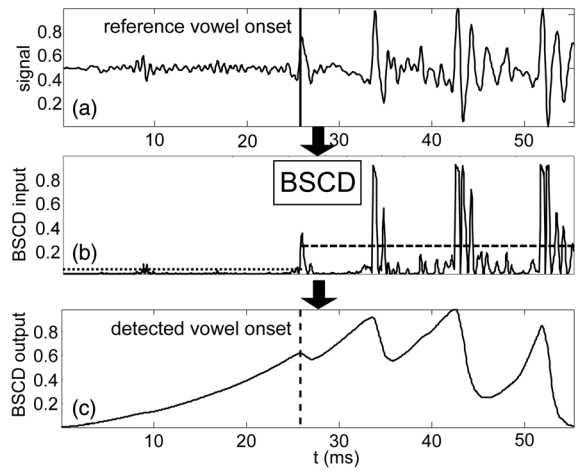


Fig. 5. Original signal with marked reference position (a), input of the BSCD detector represented by the squared original signal and marked BSCD steps (b), and output of the BSCD detector with marked positions of reference and detected V position (c).

signal (see Fig. 1); this can be seen in Fig. 5(a), where the reference position of the vowel onset is highlighted. Subsequently, due to the differing character of consonants and vowels, we may assume that the position of vowel onset is located in one of several local maxima of the BSCD output. This output is depicted in Fig. 5(c), where the detected position is marked.

To detect the local maximum corresponding to vowel onset, we may assume that the entire consonant is longer than the distance between single glottal pulses. This presumption allows delineation of the local maximum, following the largest gap between two consecutive maxima, as the position of vowel onset.

5) Detection of Occlusion: The position of occlusion is the most difficult to detect due to its slow subsidence and fuzzy borders. Due to decreased voice quality of PD speakers, a low-pass FIR filter with an order of one quarter of the signal length including the 1.5 kHz cut-off frequency was used. Contrary to the noise component, the fundamental frequency (F0) and first two formant frequencies (F1 and F2) provide a major contribution to signal energy in this frequency band.

Signal energy was estimated from the filtered rear-part of the signal (see Fig. 6(a)) as the squared signal (see Fig. 6(b)). Subsequently, the flexible threshold was adjusted for occlusion detection. The threshold was given as an inverted polynomial energy approximation, and therefore the threshold was lowered with an increase in energy and vice versa, as illustrated by Fig. 6(b). The definition of the threshold may be written as

$$T_{Occlusion} = \prod_{j=0}^{k} c_k x^k + 2\bar{E}, \tag{4}$$

where c_k denotes the k-th coefficient, $\bar{E}$ gives the mean value of energy, and k is the order of polynomial approximation. The order was experimentally set at nine, providing a good compromise between threshold elasticity and boundary fuzziness. The exact occlusion position was then marked as the place of the last intersection of energy and the threshold, which is no further than 20 ms from the preceding intersection. The 20 ms rule eliminates false detections connected with abrupt noises in distant parts of the signal.

E. Articulatory Features

To evaluate the impact of PD on speaker performance, we propose 13 features representing six aspects of speech. The features describing *Voice Quality*, *Coordination of Laryngeal and Supralaryngeal Activity*, *Precision of Consonant Articulation*, *Tongue Movement*, *Occlusive Weakening*, and *Speech Timing* are listed in Table II. Due to the differing spectral characteristics of /p/, /t/, and /k/ consonants and their following vowels, features describing the precision of consonant articulation and tongue movement were performed on different types of syllables (bilabial /pa/, alveolar /ta/, and velar /ka/), separately. Moreover, the measurements connected with the coordination of laryngeal and supralaryngeal activity were performed for separate and mixed syllables. Therefore, the final number of measurements performed was 27. All of the measurements ranked each utterance with an average feature value computed from the first 5 syllabic trains (15 syllables overall). This approach helps to separate the involvement a of single speech feature from the impact of varying speech length.

1) Voice Quality: One of the muscle groups affected by PD is the group of laryngeal muscles. Distortion of this muscle group may lead to decreased vocal fold adduction and decreased ability to keep laryngeal muscles in a fixed position, which may result in increased jitter, shimmer, noise, distortion of F0 in general, and voice tremor [23], [42]. It is beyond the scope of this article to provide a complex overview of voice quality estimation methods. Nevertheless, to obtain general information about voice quality, two vowel similarity quotients and one vowel variability quotient were utilized. The vowel similarity quotient of the entire voicing (VSQ) and the vowel similarity quotient of the first 30 ms of voicing (VSQ_{30}) are defined as the first autocorrelation coefficients, and estimate the ability to produce a steady vocal tone. The motivation behind a 30 ms window in VSQ_{30} was based on a previous study on vowel articulation in PD [43]; in the present study, the 30 ms window represented the midpoint of the vowel that should manifest the greatest periodicity through the entire vowel duration. The vowel variability

TABLE II
DEFINITIONS OF ARTICULATORY FEATURES

Name	Defined in interval from	Definition
Voice Quality		
VSQ	vowel onset to occlusion	Vowel similarity quotient, the autocorrelation of the entire vowel duration, representing the rate of regularity of the vowel
VSQ_{30}	first 30ms after vowel onset	VSQ of the first 30 ms of the vowel, representing the rate of regularity of the vowel beginning
VVQ	vowel onset to occlusion	Vowel variability quotient, the level of variability in vowel lengths
Coordination of Laryngeal and Supralaryngeal Activity		
VOT	initial burst to vowel onset	Voice onset time defining the length of the entire consonant
VOT ratio	initial burst to vowel onset	The voice onset time ratio defining the length of the entire consonant relative to syllable length
Precision of Consonant Articulation		
CST	initial burst to vowel onset	Consonant spectral trend, the regression of consonant spectrum computed in defined intervals
CSM	initial burst to vowel onset	Consonant spectral moment, the first spectral moment of the consonant
Tongue Movement		
1FT	vowel onset to occlusion	First formant trend, regression of the first format frequency
2FT	vowel onset to occlusion	Second formant trend, regression of the second format frequency
Occlusion Weakening		
SNR	vowel onset to occlusion (harmonic signal) as compared to occlusion to subsequent initial burst (noise signal)	Signal-to-noise ratio, representing the amplitude of tonal to noise components
Speech Timing		
DDK rate	entire utterance	Diadochokinetic rate, the number of syllables per second
DDK pace	occlusion to subsequent initial burst	Diadochokinetic pace, the mean length of silent gaps between syllables
DDK fluctuation	occlusion to subsequent initial burst	Diadochokinetic instability, the level of instability of silent gaps between syllables

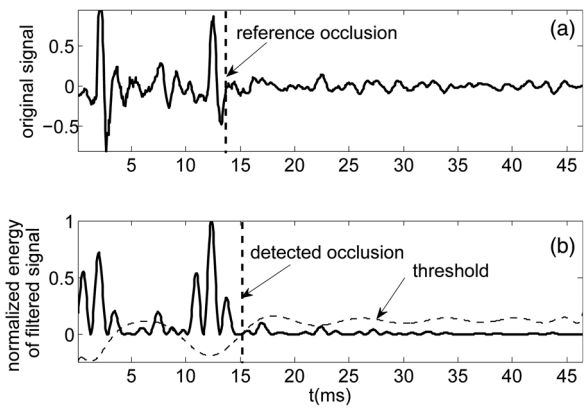


Fig. 6. Rear part of a syllable with marked reference position O (a), and energy of the filtered signal with polynomial threshold and marked position O (b).

quotient (VVQ) is given as the standard deviation of vowel duration, which reflects the stability of the timing of vocal fold abduction and adduction.

2) Coordination of Laryngeal and Supralaryngeal Activity: The PD-induced disruption of movement patterns may lead to disturbances in muscle group coordination. To evaluate the impact of PD on the coordination of laryngeal and supralaryngeal muscle groups, the voice onset time (VOT) and the VOT ratio were used. The VOT parameter, defined as the duration between stop release and the onset of voicing [44], was motivated by the assumption that acoustic events, including the initial burst and vowel onset, are associated with articulatory gestures (i.e., the release of consonant constriction, the onset of vocal fold vibration) [44]. In addition, the VOT ratio, defined as VOT divided by the length of entire syllable, was estimated as the parameter suppressing the effect of speech rate [22].

3) Precision of Consonant Articulation: Effort to achieve a normal repetition rate may lead to reduced articulatory displacement. This reduced movement may manifest as airflow leaking around insufficiently closed articulators as well as decreased energy during the initial burst. To assess the impact of imprecise articulator setup, spectral characteristics describing a consonant spectral trend (CST) and a consonant spectral moment (CSM) were employed. The consonant spectral trend is computed as the slope of the line obtained using Fourier spectrum regression in a certain frequency interval. To emphasize the different spectral characteristics of /p/, /t/, and /k/ consonants, three different frequency bands were selected as: /p/ [2500, 3500] Hz; /t/ [2000, 3000] Hz; /k/ [1500, 2500] Hz [44]. The CSM represents the first spectral moment describing a centroid of energy contained in the entire Fourier spectrum of the consonant.

4) Tongue Movement: As one of the major articulators, the tongue has a crucial influence on the shape of the oral cavity and formant frequencies, and therefore, change of formant frequency behavior may reveal PD-induced disruption of tongue movement. In general, the acoustic-articulatory relationship can be easily understood, as the F1 frequency varies inversely with tongue height and the F2 frequency varies directly with tongue advancement [28], [42], [43]. To assess tongue movement during vocalization, the first formant trend (1FT) and the second formant trend (2FT) were computed as the angle of the linear regression line of F1 or F2 tracked in the vowel.

5) Occlusion Weakening: Reduced articulatory movements may also be present during the silent gap between two syllables. Reduced movements may lead to the leakage of turbulent airflow, which results in increased noise during the silent gap [37]. To describe the noise contained in the silent gap, the signal-to-noise ratio (SNR) defined a according to equation (5)

$$SNR = 10log_{10}\frac{P_S}{P_N}, \tag{5}$$

where P_S represents power contained in voicing and P_N represents power obtained in the signal during the silent gap.

6) Speech Timing: Disrupted movement patterns do not only influence two particular muscle groups separately (e.g., coordination of laryngeal and supralaryngeal activity), but may also

affect all aspects of speech timing. Therefore, three parameters were proposed to evaluate the impact of PD on speech timing. The first designed parameter investigates the overall DDK speech rate (DDK rate). The DDK rate is defined as the number of syllables per second and is computed as the number of initial bursts across the entire utterance. The second parameter estimates the ratio of silent gaps during the DDK task (DDK pace), and it is defined as the average value of silent gaps obtained in each utterance. The DDK pace, in connection with the DDK rate, provides information about the speech-silence duration ratio. The third parameter reflects the subject's ability to maintain a steady rhythm during the DDK speech task (DDK fluctuation), and is computed as the standard deviation of the duration of silent gaps in an utterance.

F. Statistics

Statistical analyses were performed in three separate parts: algorithm performance evaluation for automatic segmentation of an utterance, the evaluation of group differences across articulatory features estimated from segmented utterances, evaluation of the classification experiment based on previously computed articulatory features. Although these three parts are interconnected, the evaluation of each was performed separately, i.e., single syllables were used in the evaluation of algorithm performance, average performances of each participant for group difference estimation, and single utterances for the classification task (two per subject).

1) Algorithm Performance: Algorithm performance is illustrated by the cumulative distributions of absolute differences between reference-manual labels and automatically detected positions. For each syllable's event (i.e., initial burst, vowel onset, occlusion), three cumulative distributions were computed. The first was based on all 1644 tokens (across both PD and HC groups). Two other distributions were based on PD or HC tokens separately (753 tokens for PD and 891 for HC).

Furthermore, to compare the performance of our algorithm with previous results, a method based on the teager energy operator (TEO) published by Hansen *et al.* was implemented [44]. This approach uses the amplitude modulation component (AMC), which is derived from the TEO, to detect the initial burst and vowel onset in single words. The TEO-based algorithm is not designed for the detection of occlusion. The AMC was applied on the filtered signal, whereas the parameters of the filter were set according to the event (i.e., initial burst or voice onset), and also according to the type of consonant (i.e., /p/, /t/, /k/) when considering burst. The TEO-based algorithm was used to detect the initial burst and voice onset in our data and the cumulative distributions of absolute differences for all PD and HC syllables.

2) Group Differences and Relationships Between Metrics: For assessment of group differences, the average feature values were calculated for each participant prior to analyses. As the one-sample Kolmogorov-Smirnov test ($D = 0.08$ to 0.20, $p > 0.05$) showed that articulatory features were normally distributed, the two-sample t-test was used to assess group differences. Cohen's effect size (ES) was additionally calculated to assess the strength of differences between the PD and HC groups. Finally, the Pearson correlation coefficient was used to evaluate the correlation between results obtained by automatic detection and reference values, as well as the extent to which single measurements were correlated.

3) Classification Experiment: The experiment based on the support vector machine (SVM) classifier was performed using all utterances (two per subject) in order to obtain more robust classifier estimates, i.e., the utterances provided by the same participant were not averaged as in the evaluation of group differences. The aim of the experiment was to separate two classes of PD and HC participants, based on automatically extracted articulatory features, which were pre-selected using Pearson's correlation and distance correlation.

Being linearly inseparable, the features had to be mapped to the space with higher dimensionality, where the linear separability was achieved. For this purpose a Gaussian radial basis function (RBF) kernel was used. The RBF is defined as

$$\mathbf{K}(\mathbf{z}, \mathbf{z}') = exp(-\gamma \|\mathbf{z} - \mathbf{z}'\|^2), \tag{6}$$

where $\|\mathbf{z} - \mathbf{z}'\|$ is euclidean distance of the input vectors and the kernel parameter γ is used to set width of Gaussians approximating the decision boundary. The SVM model may be then written as

$$\text{sign}\left(\sum_{\alpha_n > 0} \alpha_n y_n \mathbf{K}(\mathbf{z}, \mathbf{z}') + \beta\right), \tag{7}$$

where $\mathbf{z}$ and $\mathbf{z}'$ are vectors of input features, y_n are labels of data used for training and α_n are Lagrange multipliers based on the Lagrange formulation of the optimization task. To prevent overfitting the penalty coefficient C was used to constrain the maximal value of Lagrange multipliers.

The determination of the optimal parameter C and γ was performed using a grid search over the sets $\mathrm{C} = [2^{-15}, 2^{-13}, \ldots, 2^{15}]$ and $\gamma = [2^{-15}, 2^{-13}, \ldots, 2^{3}]$ [18]. Once the optimal parameters C and γ were found the classifier was trained and tested using these values.

To validate the generalization, empirical findings of previous studies suggest cross-validation or bootstrap methods as the most reliable [18], [45]. For the purposes of the generalization estimation the standard cross-validation splitting entire dataset (80 utterances) to the training set containing only 60% of the data (48 utterances) and the testing set containing 40% of all recordings (32 utterances) was employed. For the purposes of the cross-validation a total number of 20 repetitions were performed, with random permutation of the data prior to splitting into training and test subsets. Furthermore, leave-one-subject-out (LOSO) cross-validation, excluding all utterances of the subject used for testing, was utilized and run throughout the entire data.

The testing error was estimated during each iteration of both cross-validations [46]. Subsequently, the errors were averaged over all repetitions and the overall performance was determined as the average percentage of correctly classified utterances. Furthermore, the true positive (number of correctly classified PD participants) and true negative (number of correctly classified HC participants) classification performances were assessed.

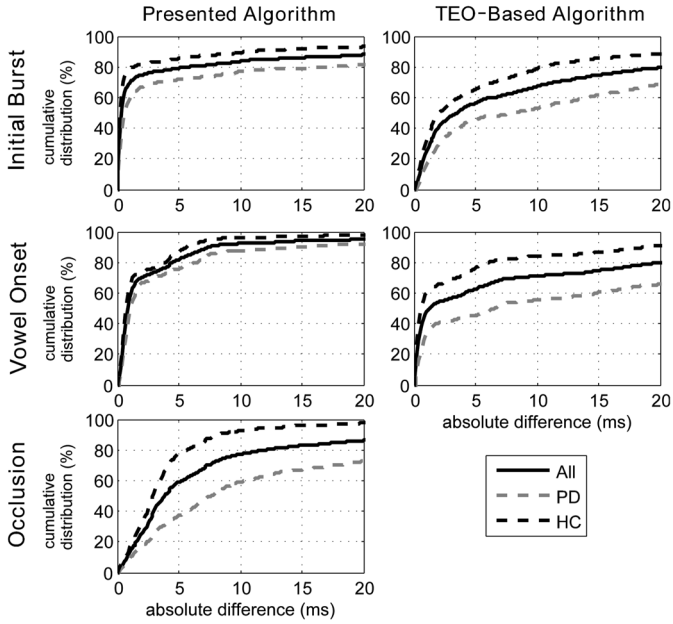


Fig. 7. Cumulative distributions of algorithm performance based on the absolute difference between automatic detection and reference labels. Performances are estimated separately for syllables in Parkinson's disease subjects (PD) and healthy controls (HC), as well as for all syllables together (All). The first column shows the performance of the algorithm presented in this study and the second column illustrates the performance achieved by the TEO-based algorithm.

III. Results

A. Algorithm Performance

Fig. 7 shows the cumulative distribution representing the absolute difference between reference manual labels and automatically detected positions for the initial burst, vowel onset and occlusion, where the left column represents the performance results of our algorithm and the right column the performance of the TEO-based algorithm proposed by Hansen *et al.* [44]. Considering a 5 ms threshold of absolute difference, performance of our algorithm for all syllables was 79.2% for the initial burst, 81.7% for vowel onset and 59.2% for occlusion. The detection performance for occlusion increased to 77.3% at a 10 ms threshold.

Considering a 5 ms threshold for initial burst of HC group, our approach achieved 85.4% in comparison to TEO-based algorithm with 65.1% accuracy [44]. In the case of vowel onset, our approach reached 86.7% compared to 76.2% by the TEO-based algorithm. In PD group, our approach achieved a score of 71.9% in comparison to 45.2% by the TEO-based algorithm for the initial burst and 5 ms threshold. Similarly, we reached a performance of 75.8% in comparison to 45.6% by the TEO-based algorithm in the detection of vowel onset. Although the TEO-based approach achieved high, and even comparable performances in HC group, its accuracy was relative low in PD group due to overall decreased speech quality.

B. Group Differences and Relationships Between Metrics

The characteristics of each measurement, including the mean and standard deviation of values in the PD and HC groups, and effect sizes are listed in Table III. Significant differences between PD and HC performances were found in each feature group. The correlations between features based on automatic detection and manual reference labels showed high reliability ($r = 0.70$ to 0.99, $p < 0.001$) for all features except for those based upon precision of consonant articulation which showed moderate reliability ($r = 0.40$ to 0.69, $p < 0.001$).

TABLE III
Overview of Results

#	Feature	HC μ	HC σ	PD μ	PD σ	Effect size†
	Voice Quality					
1	VSQ (-)	0.45	0.10	0.41	0.15	0.33
2	VSQ_{30} (-)	0.45	0.11	0.37	0.11	0.74*
3	VVQ (ms)	0.15	0.14	0.41	0.36	0.96**
	Coordination of Laryngeal and Supralaryngeal Activity					
4	VOT:all (ms)	20.33	6.14	34.50	6.17	2.30***
5	VOT:/pa/ (ms)	14.08	4.66	26.57	6.15	2.30***
6	VOT:/ta/ (ms)	22.21	7.91	36.42	10.33	1.54***
7	VOT:/ka/ (ms)	24.73	8.39	40.49	7.05	2.03***
8	VOT ratio:all (%)	28.32	6.51	35.43	6.57	1.08***
9	VOT ratio:/pa/ (%)	22.40	5.77	30.84	8.20	1.19***
10	VOT ratio:/ta/ (%)	29.90	7.98	36.41	7.81	0.83**
11	VOT ratio:/ka/ (%)	32.65	8.08	39.04	6.97	0.85**
	Precision of Consonant Articulation					
12	CST:/pa/ (rad $\times 10^{-3}$)	-3.23	1.30	-2.25	1.31	0.76*
13	CST:/ta/ (rad $\times 10^{-3}$)	-2.89	2.02	-2.00	1.15	0.54
14	CST:/ka/ (rad $\times 10^{-3}$)	-4.29	1.91	-2.02	1.34	1.38***
15	CSM:/pa/ (kHz)	4.93	0.38	4.98	0.47	0.11
16	CSM:/ta/ (kHz)	5.00	0.61	5.42	1.01	0.50
17	CSM:/ka/ (kHz)	4.81	0.41	4.87	0.49	0.13
	Tongue Movement					
18	1FT:/pa/ (rad)	0.02	0.11	-0.13	0.13	1.19***
19	1FT:/ta/ (rad)	0.03	0.14	-0.11	0.13	1.10**
20	1FT:/ka/ (rad)	0.14	0.14	-0.02	0.13	1.14**
21	2FT:/pa/ (rad)	-0.09	0.26	-0.06	0.25	0.12
22	2FT:/ta/ (rad)	0.55	0.22	0.28	0.26	1.14***
23	2FT:/ka/ (rad)	-0.53	0.21	-0.43	0.21	0.46
	Occlusive Weakening					
24	SNR (dB)	28.02	4.16	25.13	5.03	0.63*
	Speech Timing					
25	DDK rate (syll/s)	7.74	0.65	6.69	0.88	1.36***
26	DDK pace (ms)	64.34	11.26	58.29	16.66	0.42
27	DDK fluctuation (ms)	17.9	11.18	31.42	19.44	0.85**

† Measurements Reaching Significance are Denoted by Asterisks: $^{*}) p < 0.05$, $^{**}) p < 0.01$, and $^{***}) p < 0.001$.

In the voice quality dimension, the VVQ was significantly increased in PD patients when compared to controls ($t(44) = -3.13$, $CI = [-0.93E^{-4}, 4.29E^{-4}]$, $p = 0.003$). Similarly, the VSQ_{30} was decreased in PD patients ($t(44) = 2.42$, $CI = [-1.45E^{-1}, -0.13E^{-1}]$, $p = 0.02$). In the dimension considering the coordination of laryngeal and supralaryngeal articulators, both VOT (e.g. VOT:all $t(44) = -7.54$, $CI = [1.04E^{-2}, 1.80E^{-2}]$, $p = 0.003$) and VOT ratio (e.g. VOT ratio:all $t(44) = -3.57$, $CI = 0.31E^{-1}, 1.11E^{-1}]$, $p = 0.003$) features reflected a considerable increase for PD participants, with VOT generally providing superior results to VOT ratio as demonstrated by effect sizes. Considering the disrupted precision of consonant articulation, a significant difference in CST between the HC and PD groups for /pa/ ($t(44) = -2.48$,

$CI = [1.83E^{-6}, 0.18E^{-6}]$, $p = 0.02$) and /ka/ ($t(44) = -4.54$, $CI = [-1.26E^{-5}, 3.28E^{-5}]$, $p < 0.0001$) syllables was observed, whereas only a trend was detected for /ta/ ($t(44) = -1.7899$, $CI = [-1.15E^{-6}, 0.19E^{-6}]$, $p = 0.08$). However, we found no significant group differences for CSM extracted through various consonants. In the tongue movement dimension, all the 1FTs for /pa/ ($t(44) = 3.88$, $CI = [-2.23E^{-1}, -0.70E^{-1}]$, $p = 0.0004$), /ta/ ($t(44) = 3.61$, $CI = [-2.27E^{-1}, -0.64E^{-1}]$, $p = 0.0008$) and /ka/ ($t(44) = 3.75$, $CI = [-2.42E^{-1}, -0.72E^{-1}]$, $p = 0.0006$) syllables were significantly different between the PD and HC groups. In contrast, only 2FT for the /ta/ syllable ($t(44) = 3.72$, $CI = [-4.20E^{-1}, -1.25E^{-1}]$, $p = 0.0006$) was found to be impaired in PD patients. Lower SNR for the PD group provided significant distinction ($t(44) = 2.05$, $CI = [-5.74, -0.04]$, $p = 0.047$) in the occlusive weakening dimension. Finally, the speech timing dimension exhibited a considerable decrease in the DDK rate ($t(44) = 4.45$, $CI = [-1.53, -0.58]$, $p < 0.0001$), and increase in DDK fluctuation ($t(44) = -2.78$, $CI = [0.37E^{-2}, 2.34E^{-2}]$, $p = 0.0082$) in the PD group.

C. Classification Experiment

Considering relations between speech features, the Pearson's correlation revealed correlation higher than 0.9 between the VOT and VOT ratio measurements. Accordingly, distance correlation reached value higher than 0.8 only between VOT and VOT ratio measurements. Therefore, all 27 features were retained for the classification experiment. The most representative classification results are presented in Table IV, where the correct overall, true positive and true negative performance rates are listed. Interestingly, the best correct overall classification score of $87.1 \pm 5.4\%$ obtained by standard cross-validation and $88.4 \pm 26.4\%$ obtained by LOSO cross-validation was achieved for the combination of six parameters (VSQ_{30}, VOT:/pa/, CST:/ka/, 2FT:/ta/, SNR, DDK rate), each representing one different speech dimension. Fig. 8 shows probability distributions for six representative features with the best classification accuracy estimated using the Gaussian kernel density method.

IV. Discussion

In the current study we present a fully automatic approach to assess articulatory disorders in PD. In contrast to previous research that primarily focused on the assessment of dysphonic patterns, this study is the first to explore the automatic quantification of acoustic aspects of articulatory dysfunction in PD. Our designed speech features proved capable of describing parkinsonian dysarthria and even differentiating between speech in de novo PD patients and controls with a high classification accuracy of 88%. Interestingly, the strongest classification accuracy for a single articulatory feature was obtained through the VOT, suggesting consonant articulation is a very powerful PD indicator.

Automatic segmentation represented by cumulative distributions showed rapid growth of the performance in the first 5 ms of absolute difference between the detected and reference positions. Considering the 5 ms threshold for initial burst and vowel onset, our algorithm performance exceeded 85% accuracy for HC speakers and 70% accuracy for PD patients, illustrating adequate precision of the designed algorithm in the evaluation of both healthy and dysarthric speech. Since the occlusion does not provide such abrupt change in signal energy as the initial burst or vowel onset, our algorithm reached the lowest performance of 59% within 5 ms threshold for occlusion detection but its accuracy was substantially increased to 77% when considering 10 ms threshold. Moreover, the results of the majority of our features exhibit strong or even very strong correlation to the results obtained using precise manual labels, while none fell below moderate correlation. This is crucial from the clinical

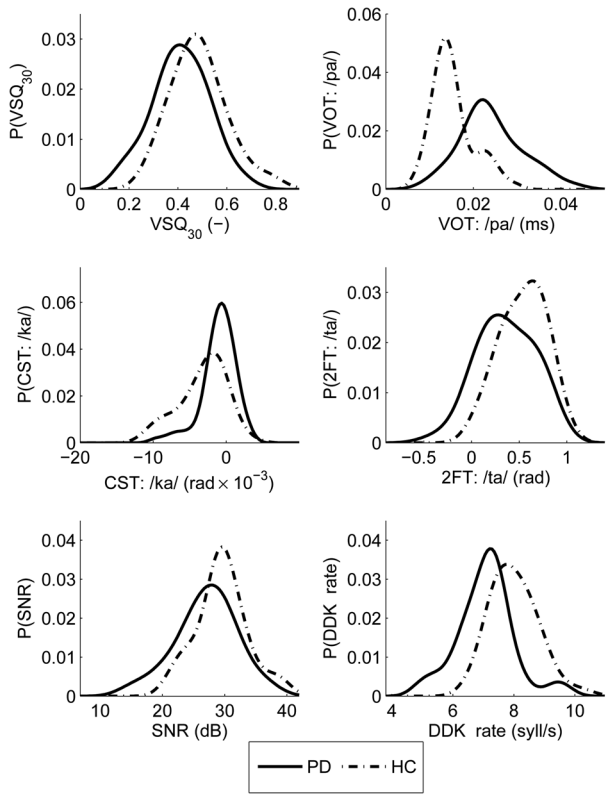


Fig. 8. Probability densities of six representative features with the best SVM classification performance. The vertical axes are the probability densities P(measure) of feature values estimated using the Gaussian kernel density method. The dashdot lines represent the HC group and solid lines the PD group.

TABLE IV
Representative Classification Results

Feature set (number of measurements)	Correct overall (%)	True positive (%)	True negative (%)
Cross-validation based on 60% training set and 40% testing set			
VSQ_{30}, VOT:/pa/, CST:/ka/, 2FT:/ta/, SNR, DDK rate (6)	87.1 ± 5.4	86.2 ± 9.6	88.0 ± 7.5
VOT:All, 2FT:/ta/, DDK rate (3)	85.2 ± 4.5	84.5 ± 9.3	86.3 ± 8.1
All measurements (27)	82.4 ± 7.0	91.4 ± 9.9	74.8 ± 10.3
VOT: all (1)	83.3 ± 5.4	87.8 ± 7.3	78.1 ± 11.5
Leave-one-subject-out cross-validation			
VSQ_{30}, VOT:/pa/, CST:/ka/, 2FT:/ta/, SNR, DDK rate (6)	88.4 ± 26.4	86.4 ± 31.6	90.5 ± 20.1
VOT:All, 2FT:/ta/, DDK rate (3)	83.7 ± 28.3	81.8 ± 29.1	85.7 ± 28.0
All measurements (27)	82.6 ± 32.5	88.6 ± 21.4	76.2 ± 40.7
VOT: all (1)	79.1 ± 34.9	90.9 ± 25.1	66.7 ± 66.7

point of view, as it is more important to achieve a correct estimation of the patient's speech performance than to obtain the precise position of individual boundaries.

Comparing our results with those obtained by the TEO-based algorithm using a 5 ms threshold [44], both algorithms showed relatively high performances, exceeding 65% for utterances in healthy speakers. However, taking into account the performance of the PD group separately, the performance of the TEO-based algorithm declined under 50% accuracy, while our algorithm still maintained sufficient accuracy, exceeding 70%. Thus, the presented comparison shows that results provided by our algorithm are less vulnerable to PD-induced signal aggravation than those obtained by the TEO-based approach. Nevertheless, it is important to note that the TEO-based algorithm was primarily designed for real-time accent analysis, whereas our algorithm is focused on reliable dysarthric speech assessment, which does not require real-time processing.

Due to pathological changes in the basal ganglia, PD disrupts the effective execution of articulatory movements leading to various phonatory, articulatory, and prosodic disturbances. Accordingly, the analysis of freely connected speech seems to be the best way to assess the impact of PD on speech [28], [47], [48]. However, the fully automatic estimation of relevant articulatory features such as VOT from free running speech is a very difficult task and to the best of our knowledge, no such algorithm has been presented to date. To provide a robust, fully automatic classifier, previous studies have primarily used speech tests with a fixed frame such as sustained phonation [13], [18], which significantly lowers the complexity of analysis and preserves as much useful information as possible. Moreover, the advantage of analyzing sustained phonation resides in fact that the speaker's native language has no or only a small effect on dysphonia parameters. Although sustained phonation measurements provide a precise estimation of dysphonic features, Parkinsonian dysphonia is only a subset of dysarthric aspects of speech, whereas dysarthria is primarily a distinctive disorder of articulation [49]. Contrary to sustained phonation measurements, our approach based upon DDK task assessment provides a wide range of articulatory aspects related to dysarthria that may be subjected to evaluation, and allows their automatic assessment; however, possible language dependency cannot be excluded.

Voice quality is represented by decreased VSQ_{30}, and by increased VVQ. Decreased VSQ_{30} in PD participants reflects increased noise caused by insufficient vocal fold adduction and phonatory instability caused by a decreased ability to keep laryngeal muscles in a fixed position [28], [42]. Increased VVQ illustrates disrupted timing of vowel gestures [23].

VOT as the most powerful PD predictor suggests the imprecise coordination of laryngeal and supralaryngeal articulation as an early, prominent sign of PD. Each VOT measurement showed considerable prolongation of consonant duration, which may indicate disrupted coordination between the laryngeal muscle group and supralaryngeal articulators (tongue, jaws, and lips). However, previous studies focused on VOT in PD have provided inconsistent results. While some researchers reported increased or unchanged VOT in PD patients [50], [51], other studies suggested decrease in VOT due to parkinsonian articulatory disorders [52], [53]. A study by Fischer and Goberman [22] suggested that this inconsistency may be related to different analysis methods used and the fact that measurements were not performed rate-independently. As PD patients may be able to willingly compensate decreased speech rates, Fischer and Goberman [22] identified the VOT ratio as an appropriate rate-independent measurement. In our study, VOT was found to be superior to VOT ratio, probably as a result of the similar length of each syllable, and partially due to the effort of repeating sequences as fast and as steady as possible, which may suppress willing compensation.

The willing compensation of speech rate is at the cost of reduced range of motion of the supralaryngeal articulators. The range of motion may also be reduced due to hypokinesia. Incomplete articulatory movements may be manifested as increased turbulent airflow leakage around the insufficiently closed obstacle, causing increased noise and alterations of the frequency spectrum. The significant difference between PD and HC groups, as captured by the CST of /pa/ and /ka/ syllables, illustrates the impact of insufficient articulatory movements during consonant enunciation.

The effect of hypokinetic dysarthria on vowels may be also described by increased noise and spectral alterations. The increased noise component in consonants is probably a result of insufficient closure of the supralaryngeal articulators, whereas the vowel noise component may be the result of insufficient vocal fold adduction [28]. On the other hand, the distorted setup of supralaryngeal articulators may evoke notable changes in formant frequencies. Therefore, the 1FT and 2FT are used to indicate disruptions of articulatory movements during voicing [28], [42]. The 1FT, which is connected with movement of the tongue in the vertical direction, illustrates impairment in all /pa/, /ta/, /ka/ syllables. The 2FT, describing advance of the tongue, shows disruption only during the /ta/ syllable, which is articulated by the tip of the tongue.

Disruption of articulatory movements leading to occlusive weakening during silent gaps between single words can be captured by decreased SNR in PD. Similar to the case of consonant articulation, this is likely caused by insufficient articulatory closure resulting in leakage of turbulent airflow [28], [37].

The general effect of dysarthria is well described by a considerable decrease of the DDK rate in PD speakers. Although the DDK pace measurement did not prove significant alterations in silent gap lengths, the DDK fluctuation revealed considerable instability of silent gaps in PD. The silent gap instability and non-significant DDK pace may suggest the effect of short rushes of speech, which can be caused by a combination of akinesia and speech hastening [16].

The presented classification experiment shows that a complex view on various aspects of Parkinsonian speech impairment using simple the task of fast syllable repetition provides great potential for fully automatic assessment of the severity of hypokinetic dysarthria in PD speakers. Using our novel DDK-based approach, we were able to predict PD group membership with a very high performance of approximately 87.1% using standard cross-validation and 88.4% using LOSO cross-validation. Since our database consists only of 80 speech samples from 46 participants, the advantage of standard cross-valida-

tion is that it provides lower variance in results due to possibility to set up larger test group. Yet, training and testing subsets may contain different utterances from the same individuals. This problem is treated by using of LOSO cross-validation, however, the result variance is increased because only 2 utterances were available per subject.

Notably, the best SVM feature subset comprises six measurements where each one represents a different aspect of speech, confirming the importance of complex speech assessment in PD. It has already been shown that the complex assessment of speech profile in PD may be essential in providing information about the effect of therapy in the course of disease progression on a particular speech apparatus [17].

Recent studies focused on the differentiation between PD and healthy speakers presented very high classification performances of 89% [13] and 98% [18] using a single sustained phonation task for the evaluation of dysphonia. However, considering that speech severity may be influenced by the severity of motor manifestations, disease duration, and specific effects of dopaminergic treatment [17], [54], [55], an exact comparison with previous results is not possible. Our PD patients were investigated immediately after the diagnosis was established and before symptomatic treatment was initiated, whereas previous datasets consisted of treated Parkinsonian patients with various disease durations after diagnosis (6.6 ± 7.3 years in [13]). In our preliminary findings [21], we achieved 85% performance in the differentiation between PD and HC participants. However, this classification score was obtained using various features estimating prosody, phonation and articulation aspects together. The classification based upon single aspects achieved classification score of 81% for prosody using monologues, 76% for phonation using sustained vowels, and only 71% for articulation using fast syllable repetitions. Therefore, in comparison to these previous results, the current approach provides a performance improvement.

Certain limitations of the present study must be considered. Due to the problematic recruitment of de novo PD patients, the current dataset consisted of only 24 Parkinsonian native Czech speakers. The small sample size of the present study may bias the performance of the classifier to a certain extent. Although newly diagnosed, the majority of our patients were already in the middle H&Y stages 2 or 2.5. However, to consider speech tests as diagnostic decision support tool for an early diagnosis of PD, we would need to differentiate between controls and untreated PD speakers in their very early disease stages. Furthermore, the language dependency of features extracted from the DDK task cannot be excluded as such patterns have never been investigated. Another limitation of the current dataset is gender imbalance, related to the greater incidence of PD in males [56], [57]. Previous studies have documented a confounding effect of sexual dimorphism on particular speech impairments [58], and we therefore cannot exclude the possibility that articulatory impairment is influenced by gender-specific aspects of speech. Finally, our algorithm was primarily designed for parkinsonian patients with mild to moderate stages of disease and thus does not need to be sufficiently sensitive to evaluation of articulatory disorders in PD patients with advanced motor stages and severe dysarthria.

The present study provides a novel extension to available technologies, allowing the automatic evaluation of speech severity in central nervous system disorders. The algorithm based on the DDK task proved to be reliable in effective separation between subjects with PD and HC. Future research could incorporate current methodology with other robust approaches such as the automatic evaluation of phonatory patterns in dysarthric speech [13], [18], which may together increase the overall performance of speech-based diagnostic support in PD.

V. Conclusion

The main purpose of the present study was to introduce a novel approach for the fully automatic evaluation of acoustic features related to articulation attributes in PD, based on DDK utterances. Our results show that the proposed approach provides excellent conditions for reliable automatic assessment, allowing the examination of a wide range of articulatory deficits connected with hypokinetic dysarthria. Moreover, the combination of the presented acoustic features accurately predicted speech impairment even in de novo PD patients, suggesting that a precise description of vocal patterns may contribute significantly to existing assessment methods for monitoring speech severity.

References

[1] O. Hornykiewicz, "Basic research on dopamine in Parkinson's disease and the discovery of the nigrostiratal dopamine pathway: The view of an eyewitness," *Neurodegener. Dis.*, vol. 5, no. 3–4, pp. 114–117, 2008.

[2] J. Jankovic, "Parkinson's disease: Clinical features and diagnosis," *J. Neurol. Neurosur. Ps.*, vol. 79, pp. 368–376, 2012.

[3] M. C. Rodriguez-Oroz, M. Jahanshahi, P. Krack, I. Litvan, R. Macias, E. Bezard, and J. A. Obeso, "Initial clinical manifestations of Parkinson's disease: Features and pathophysiological mechanisms," *Lancet Neurol.*, vol. 8, pp. 1128–1139, 2009.

[4] H. Bernheimer, W. Birkmayer, O. Hornykiewicz, K. Jellinger, and F. Seitelberger, "Brain dopamine and the syndromes of Parkinson and Huntington. Clinical, morphological and neurochemical correlations," *J. Neurol. Sci.*, vol. 20, no. 4, pp. 415–455, 1973.

[5] R. B. Postuma, A. E. Lang, J. F. Gagnon, A. Pelletier, and J. Y. Montplaisir, "How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder," *Brain*, vol. 135, pp. 1860–1870, 2012.

[6] A. L. Whone, R. L. Watts, A. J. Stoessl, M. Davis, S. Reske, C. Nahmias, A. E. Lang, O. Rascol, M. J. Ribeiro, P. Remy, W. H. Poewe, R. A. Hauser, and D. J. Brooks, "Progression of Parkinson's disease with ropinirole versus levodopa: The REAL-PET study," *Ann. Neurol.*, vol. 54, pp. 93–101, 2003.

[7] G. Becker, A. Müller, S. Braune, T. Büttner, R. Benecke, W. Greulich, W. Klein, G. Mark, J. Rieke, and R. Thümler, "Early diagnosis of Parkinson's disease," *J. Neurol.*, vol. 249, 2002, suppl. III/40–III/48.

[8] B. T. Harrel, M. S. Cannizzaro, H. Cohen, N. Reilly, and P. J. Snyder, "Acoustic characteristics of parkinsonian speech: A potential biomarker of early disease progression and treatment," *J. Neurolinguist.*, vol. 17, pp. 439–453, 2004.

[9] K. Ho, R. Iansek, C. Marigliani, J. Bradshaw, and S. Gates, "Speech impairment in large sample of patients with Parkinson's disease," *Behav. Neurol.*, vol. 11, no. 3, pp. 131–137, 1999.

[10] J. A. Logemann, H. B. Fisher, B. Boshes, and E. R. Blonsky, "Frequency and occurrence of vocal tract dysfunction in the speech of a large sample of Parkinson patients," *J. Speech. Hear. Disord.*, vol. 11, pp. 47–57, 1978.

[11] K. M. Rosen, R. D. Kent, A. L. Delaney, and J. R. Duffy, "Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers," *J. Speech Lang. Hear. R.*, vol. 49, pp. 395–341, 2006.

[12] R. D. Kent, J. F. Kent, G. Weismer, and J. R. Duffy, "What dysarthrias can tell us about the neural control of speech," *J. Phonetics*, vol. 28, pp. 273–302, 2000.

[13] M. A. Little, P. M. McSharry, E. J. Hunter, J. Speielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 4, pp. 1015–1022, Apr. 2009.
[14] G. J. Canter, "Speech characteristics of patients with Parkinson's disease: III. Articulation diadochokinesis, and overall speech adequacy," *J. Speech Hear. Disord.*, vol. 30, pp. 217–224, 1965.
[15] F. Rudzicz, "Articulatory knowledge in the recognition of dysarthric speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 4, pp. 947–960, May 2011.
[16] F. L. Darley, A. E. Aronson, and J. R. Brown, *Motor speech disorders*. Philadelphia, PA, USA: Saunders, 1975.
[17] J. Rusz, R. Čmejla, H. Růžičková, J. Klempíř, V. Majerová, J. Picmausová, J. Roth, and E. Růžička, "Evaluation of speech impairment in early stages of Parkinson's disease: A prospective study with the role of pharmacotherapy," *J. Neural. Transm.*, vol. 120, no. 2, pp. 319–329, 2013.
[18] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, May 2012.
[19] C. A. Baumgartner, S. Sapir, and L. O. Ramig, "Voice quality changes following phonatory-respiratory effort treatment (LSTVT) versus respiratory effort treatment for individuals with Parkinson disease," *J. Voice*, vol. 15, no. 1, pp. 105–114, 2001.
[20] K. Chenausky, J. Mac Auslan, and R. Goldhor, "Acoustic analysis of PD Speech," *Parkinson's Disease*, vol. 2011, p. 13, 2011, Article ID 435232.
[21] J. Rusz, R. Čmejla, H. Růžičková, and E. Růžička, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease," *J. Acoust. Soc. Amer.*, vol. 129, no. 1, pp. 350–367, 2011.
[22] E. Fischer and A. M. Goberman, "Voice onset time in Parkinson's disease," *J. Commun. Disord.*, vol. 43, pp. 21–34, 2010.
[23] A. M. Goberman and M. Blomgren, "Fundamental frequency change during offset and onset of voicing in individuals with Parkinson Disease," *J. Voice*, vol. 22, no. 2, pp. 178–191, 2006.
[24] I. Midi, M. Dogan, M. Koseoglu, G. Can, M. A. Sheitoglu, and D. I. Gunal, "Voice abnormalities and their relation with motor dysfunction in Parkinson's disease," *Acta Neurol. Scand.*, vol. 117, pp. 26–34, 2008.
[25] K. Rosen, B. Murdoch, J. Folker, A. Vogel, L. Cahill, M. Delatycki, and L. Corben, "Automatic method of pause measurement for normal and dysarthric speech," *Clin. Linguist. Phonet.*, vol. 24, no. 2, pp. 141–154, 2010.
[26] H. Ackerman, J. Koznack, and I. Hertrich, "The temporal control of repetitive articulatory movements in Parkinson's disease," *Brain Lang.*, vol. 57, pp. 312–319, 1997.
[27] S. Fletcher, "Time-by-count measurement of didochokinetic syllable rate," *J. Speech Hear. R.*, vol. 15, pp. 757–762, 1972.
[28] R. D. Kent, G. Weismer, J. F. Kent, J. K. Vorperian, and J. R. Duffy, "Acoustic studies of dysarthric speech: Methods, progress, and potential," *J. Commun. Disord.*, vol. 32, pp. 141–186, 1999.
[29] C. Oszancak, P. Auzou, M. Jan, and D. Hannequin, "Measurements of voice onset time in dysarthric patients: Methodological consideration," *Folia Phoniatr. Logo.*, vol. 53, pp. 48–57, 2001.
[30] J. Rusz, R. Čmejla, H. Růžičková, J. Klempíř, V. Majerová, J. Picmausová, J. Roth, and E. Růžička, "Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test," *Mov. Disord.*, vol. 26, no. 10, pp. 1951–1952, 2011.
[31] A. J. Hughes, S. E. Daniel, L. Kilford, and A. J. Lees, "Accuracy of clinical diagnosis of idiopathic Parkinson's disease: A clinicopathological study of 100 cases," *J. Neurol. Neurosur. Ps.*, vol. 55, pp. 181–184, 1992.
[32] M. M. Hoehn and M. D. Yahr, "Parkinsonism: Onset, progression, and mortality," *Neurology*, vol. 17, pp. 427–442, 1967.
[33] G. Stebbing and C. Goetz, "Factor structure of the unified parkinson's disease rating scale: Motor examination section," *Mov. Disord.*, vol. 13, pp. 633–636, 1998.
[34] Y. Wang, R. Kent, J. Duffy, J. Thomas, and G. Weismer, "Alternating motion rate as an index of speech motor disorder in traumatic brain injury," *Clin. Linguist. Phonet.*, vol. 18, no. 1, pp. 57–84, 2004.
[35] L. Volaitis and J. Miller, "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," *J. Acoust. Soc. Amer.*, vol. 92, pp. 723–735, 1992.
[36] J. S. Allen, J. L. Miller, and D. DeSteno, "Individual talker differences in voice-onset-time," *J. Acoust. Soc. Amer.*, vol. 113, no. 1, pp. 544–552, 2003.
[37] D. Duez, "Acoustic analysis of occlusive weakening in parkinsonian French speech," presented at the Int. Congr. Phonetic Sci., Saarbrücken, Germany, 2007.
[38] R. Titze, "Workshop on acoustic voice analysis," in *National Center for Speech and Voice*, Denver, CO, USA, 1994.
[39] J. D. Harris and D. Nelson, "Glottal pulse alignment in voiced speech for pitch determination," in *Proc. IEEE Conf. Acoust., Speech, Signal Process.*, 1993, vol. 2, pp. 519–522.
[40] J. J. O. Ruanaidh and W. J. Fitzgerald, "Numerical bayesian methods applied to signal processing," in *Series on Statistics and Computing*. Berlin, Germany: Springer-Verlag, 1996.
[41] R. Čmejla, J. Rusz, P. Bergl, and J. Vokřál, "Bayesian changepoint detection for the automatic assessment of fluency and articulatory disorders," *Speech Commun.*, vol. 55, pp. 178–189, 2013.
[42] B. E. F. Lindblom and J. E. F. Sundberg, "Acoustical consequences of lip, tongue, jaw, and larynx movement," *J. Acoust. Soc. Amer.*, vol. 50, pp. 1166–1179, 1971.
[43] S. Sapir, L. O. Ramig, J. L. Spielman, and C. Fox, "Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech," *J. Speech Lang. Hear. R.*, vol. 53, pp. 114–125, 2010.
[44] J. H. L. Hansen, S. S. Gray, and W. Kim, "Automatic voice onset time detection for unvoiced stops (/p/, /t/, /k/) with application to accent classification," *Speech Commun.*, vol. 52, pp. 777–789, 2010.
[45] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: Datamining, inference, and prediction*, 2nd ed. New York, NY, USA: Springer, 2009.
[46] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "The elements of statistical learning: Datamining, inference, and prediction," in *Technical report*. Taipei, Taiwan: National Taiwan Univ., 2010.
[47] J. Rusz, R. Čmejla, T. Tykalová, H. Růžičková, J. Klempíř, V. Majerová, J. Picmausová, J. Roth, and E. Růžička, "Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task," *J. Acoust. Soc. Amer.*, vol. 134, pp. 2171–2181, 2013.
[48] S. Zhao, F. Rudzicz, L. G. Carvalho, C. Márquez-Chin, and S. Livingstone, "Automatic detection of expressed emotion in Parkinson's disease," in *Proc. ICASSP*, Florence, Italy, 2014.
[49] J. R. Duffy, *Motor speech disorders. Substrates, differential diagnosis and management*, 2nd ed. St. Louis, MO, USA: Elsevier Mosby, 2005.
[50] K. Forrest, G. Weismer, and G. Turner, "Kinematic, acoustic and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric males," *J. Acoust. Soc. Amer.*, vol. 85, pp. 2608–2622, 1989.
[51] K. Bunton and G. Weismer, "Segmental level analysis of laryngeal function in persons with motor speech disorders," *Folia Phoniatr. Logo.*, vol. 54, pp. 223–239, 2002.
[52] A. Flint, S. Campbell-Taylor, I. Galey, and C. Levinton, "Acoustic analysis in the differentiation between Parkinson's disease and major depression," *J. Psycholinguist. Res.*, vol. 21, pp. 383–399, 1992.
[53] G. Weismer, "Articulatory characteristics of Parkinsonian dysarthria: Segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination," in *The dysarthrias: Physiology, Acoustics, Perception, Management*, M. Mcneil, J. Rosenbeck, and A. Aronson, Eds. San Diego, CA, USA: College-Hill Press, 1984, pp. 101–130.
[54] S. Skodda, W. Visser, and U. Schlegel, "Short- and long-term dopaminergic effects on dysarthria in early Parkinson's disease," *J. Neural Trans.*, vol. 117, pp. 197–205, 2010.
[55] G. M. Schlutz and M. K. Grant, "Effect of speech therapy and pharmacologic and surgical tratments on voice and speech in Parkinson's disease: A review of the literature," *J. Commun. Disord.*, vol. 33, pp. 59–88, 2000.
[56] S. K. Van Den Eeden, C. M. Tanner, A. L. Bernstein, R. D. Fross, A. Leimpeter, D. A. Bloch, and L. M. Nelson, "Incidence of parkinson's disease: Variation by age, gender, and race/ethnicity," *Amer. J. Epidemiol.*, vol. 157, pp. 1015–1022, 2003.
[57] M. Balderschi, A. di Carlo, W. A. Rocca, P. Vanni, S. Maggi, E. Perissinotto, F. Grigoletto, L. Amaducci, and D. Inyitari, "Parkinson's disease and parkinsonism in longitudinal study: Two-fold higher incidence in men," *Neurology*, vol. 55, pp. 1358–1363, 2000.
[58] I. Heitrich and H. Ackerman, "Gender/specific vocal dysfunction in parkinson's disease: Electroglottographic and acoustic analyses," *Ann. Otol. Rhinol. Laryngol.*, vol. 104, pp. 197–202, 1995.

Michal Novotný received the M.S. degree in 2012 for participation on the project aimed at nondestructive estimation of elastic constants of magnetic shape memory alloys led by the Academy of Sciences of Czech Republic. Currently, he continues his study as a Ph.D. student at the Faculty of Electrical Engineering of the Czech Technical University in Prague. As a member of the Signal Analysis, Modeling, and Interpretation group (SAMI) he aims his studies on application of digital signal processing in the field of neurologic disorder-related speech pathology.

Jan Rusz received his Ph.D. degree in 2012 and currently is an Assistant Professor at the Faculty of Electrical Engineering of the Czech Technical University in Prague and member of Signal Analysis, Modeling, and Interpretation group (SAMI). His expertise covers mainly the field of speech pathology in neurologic disorders with the interdisciplinary background of digital signal processing, machine learning, physiology, and neuroscience. Results of his work have been published in several international peer-reviewed technical as well as neurological journals and presented as invited talks in international conferences. He is also Associate Editor of the *Logopedics Phoniatrics Vocology* journal.

Roman Čmejla was born in Louny, Czechoslovakia, in 1962. He received the M.S. degree in 1986 and the Ph.D. degree in communication technology in 1993, both from the Faculty of Electrical Engineering of the Czech Technical University in Prague. Since 2002 he has been an Associated Professor in the field of Electrical Engineering Theory. Since 2010 he has been the head of the Signal Analysis, Modeling, and Interpretation Lab (SAMI) of the Czech Technical University in Prague. His research interests include digital signal processing, especially analysis and processing of biological signals, particularly in the area of pathological speech, intracranial EEG, and EMG.

Evžen Růžička is a Professor and Chairman in the Department of Neurology, Charles University, Prague, Czech Republic. After receiving his M.D. degree, he obtained his neurological training from the university departments of neurology in Prague and at La Salpetriere, Paris, France. He has edited or coedited several books including the Gait Disorders volume of Advances in Neurology, and has published original articles, textbooks, and chapters covering Parkinson's disease, tremor, dystonia, Tourette's syndrome and other movement disorders. Besides numerous national conferences, he has organized and chaired international meetings including The Movement Disorder Society's International Symposium on Gait Disorders in 1999 and the European Neurological Society's annual meeting in 2012. he is a Deputy Editor of the Czech and Slovak Neurology and Neurosurgery and serves as member of editorial boards of the Polish Neurology and Neurosurgery, Neuroendocrinology Letters, Biogenic Amines and Biological and Biomedical Reports.

Appendix A6

Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy

Rusz J, Bonnet C, Klempir J, Tykalova T, Baborova E, Novotny M, Rulseh A, Ruzicka E (2015) Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *J Neurol* **262**:992-1001.

Journal name (category): **Journal of Neurology** (clinical neurology)
Quartile in category: **Q2**
Impact factor: **3.377**
Number of WoS citations (excluding autocitations): **9**

J Neurol (2015) 262:992–1001
DOI 10.1007/s00415-015-7671-1

ORIGINAL COMMUNICATION

Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy

Jan Rusz · Cecilia Bonnet · Jiří Klempíř · Tereza Tykalová · Eva Baborová · Michal Novotný · Aaron Rulseh · Evžen Růžička

Received: 11 December 2014 / Revised: 20 January 2015 / Accepted: 6 February 2015 / Published online: 17 February 2015

Abstract Although speech disorder is frequently an early and prominent clinical feature of Parkinson's disease (PD) as well as atypical parkinsonian syndromes (APS) such as progressive supranuclear palsy (PSP) and multiple system atrophy (MSA), there is a lack of objective and quantitative evidence to verify whether any specific speech characteristics allow differentiation between PD, PSP and MSA. Speech samples were acquired from 77 subjects including 15 PD, 12 PSP, 13 MSA and 37 healthy controls. The accurate differential diagnosis of dysarthria subtypes was based on the quantitative acoustic analysis of 16 speech dimensions. Dysarthria was uniformly present in all parkinsonian patients but was more severe in PSP and MSA than in PD. Whilst PD speakers manifested pure hypokinetic dysarthria, ataxic components were more affected in MSA whilst PSP subjects demonstrated severe deficits in hypokinetic and spastic elements of dysarthria. Dysarthria in PSP was dominated by increased dysfluency, decreased slow rate, inappropriate silences, deficits in vowel articulation and harsh voice quality whereas MSA by pitch fluctuations, excess intensity variations, prolonged phonemes, vocal tremor and strained-strangled voice quality. Objective speech measurements were able to discriminate between APS and PD with 95 % accuracy and between PSP and MSA with 75 % accuracy. Dysarthria severity in APS was related to overall disease severity ($r = 0.54$, $p = 0.006$). Dysarthria with various combinations of hypokinetic, spastic and ataxic components reflects differing pathophysiology in PD, PSP and MSA. Thus, motor speech examination may provide useful information in the evaluation of these diseases with similar manifestations.

Keywords Parkinson's disease · Atypical parkinsonism · Dysarthria · Speech disorder · Acoustic analyses

Electronic supplementary material The online version of this article (doi:10.1007/s00415-015-7671-1) contains supplementary material, which is available to authorized users.

J. Rusz · T. Tykalová · M. Novotný
Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, 160 00 Prague 6, Czech Republic

J. Rusz (✉) · C. Bonnet · J. Klempíř · E. Baborová · E. Růžička
Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University in Prague, Kateřinská 30, 120 00 Prague 2, Czech Republic
e-mail: rusz.mz@gmail.com

C. Bonnet
AP HP, Neurology Department, Pitié Salpêtrière Hospital, Paris, France

J. Klempíř
Institute of Anatomy, 1st Faculty of Medicine, Charles University, U nemocnice 3, 128 00 Prague 2, Czech Republic

A. Rulseh
Department of Radiology, Na Homolce Hospital, Prague, Czech Republic

A. Rulseh
Department of Radiology, 1st Faculty of Medicine and General University Hospital, Charles University, Prague, Czech Republic

Introduction

Parkinson's disease (PD) is a neurological disorder caused by the degeneration of dopaminergic neurons, leading to clinical features characterized by bradykinesia, rigidity, tremor and postural instability. Atypical parkinsonian syndromes (APS) such as progressive supranuclear palsy (PSP) and multiple system atrophy (MSA) differ from PD by more widespread neuronal involvement, resulting in additional clinical signs, more rapid disease progression and poor response to dopamine replacement therapy [1]. The majority of PSP and MSA patients develop clinical features that overlap those of PD and thus the correct diagnosis can be challenging in early stages of the disease. However, an accurate, early diagnosis is essential not only in assessing prognosis and making decisions regarding treatment, but also for understanding the underlying pathophysiology and for the development of new therapies [2]. Currently, a variety of imaging techniques such as Magnetic Resonance Imaging, Diffusion Tensor Imaging, Positron Emission Tomography, Single-photon Emission Computed Tomography and Transcranial Sonography may be used in the assessment of various parkinsonian syndromes [3]. In particular, automatic image-based classification based on metabolic patterns is highly accurate in distinguishing between PD, PSP and MSA patients at early stages of the disease, with more than 84 % sensitivity and 94 % specificity [4]. However, metabolic imaging is burdened by the invasive application of radiopharmaceuticals, whilst technical demands and financial costs may limit the application of other imaging methods.

Speech assessment is an inexpensive, non-invasive, quick and simple technique that could potentially be used in the evaluation of subjects with initial parkinsonism [5]. Speech disorder is a common clinical manifestation occurring in 70–100 % of patients with PD, PSP and MSA [6–8], and tends to emerge at an early stage [9, 10]. Whilst the majority of PD patients develop a clear form of hypokinetic dysarthria [6], PSP and MSA patients typically evolve mixed dysarthria with various combinations of hypokinetic, spastic and ataxic components [7, 8] due to the involvement of the basal ganglia, corticobulbar pathways and the cerebellum. Analyses of motor speech disorders may thus provide important clues to the diagnosis and pathophysiology of the underlying disease. However, perceptual dysarthria assessment may be difficult in early disease stages when speech impairment is often imperceptible [11]. To this extent, acoustic analyses have the unique potential to provide objective, sensitive and quantifiable information for the precise assessment of various deviant speech dimensions [10, 12].

Previous descriptions of speech in PSP and MSA have been mainly limited to the perceptual estimation of dysarthria type [7, 8], where spastic components appear to be more dominant in PSP and hypokinetic components in MSA. Considering individual speech aspects, only the occurrence of stuttering-like behaviour was reported to be specific for PSP [7, 8]. Few studies have provided more accurate objective descriptions of dysarthria in APS [9, 13–15]. In general, these studies have shown that the impairment of specific speech dimensions is more pronounced in APS than in PD [13–15]. Speech velocity, maximum phonation time, intonation variability and articulation precision were reduced and pauses were prolonged in PSP in comparison with PD [13–15], whilst MSA patients manifested voice perturbations and slow and variable alternating motion rates (AMR) [9]. However, little effort has been put into the investigation of complex speech impairment in APS. A direct, objective comparison between individual speech patterns in PSP and MSA patients has never been performed and distinctive speech markers that would be suitable for the differentiation of various forms of parkinsonism remain generally unknown.

Therefore, the specific speech characteristics allowing discrimination between dysarthria in PD, PSP and MSA should first be determined in clinically probable patients, with the future goal of evaluating speech analysis as an instrument for early-stage differential diagnosis. In particular, we quantitatively assessed 16 key speech dimensions using objective acoustic analyses with the following aims:

1. To characterize the type and severity of dysarthria in PSP and MSA.
2. To determine specific dysarthric patterns and estimate their reliability in differentiating between PD, PSP and MSA.
3. To explore the relationship between speech and clinical manifestations to provide greater insight into the pathophysiology of dysarthria in APS.

Methods

Subjects

From 2011 to 2014, 12 consecutive patients with the clinical diagnosis of probable PSP (10 men, 2 women) and 13 patients with the diagnosis of probable MSA (6 men, 7 women) were recruited for the present study. In this series, 9 PSP patients were diagnosed with the Richardson's syndrome (PSP-RS), 2 with PSP-parkinsonism (PSP-P) and 1 with PSP-pure akinesia with gait freezing (PAGF),

whereas 10 MSA patients were diagnosed as the parkinsonian type (MSA-P) and 3 as cerebellar type (MSA-C). Additionally, 15 patients with idiopathic PD (9 men, 6 women) were investigated. The PD patients were selected in order to match PSP and MSA groups according to disease duration, which was estimated based on the self-reported occurrence of first motor symptoms. The diagnosis of PSP was established by the NINDS-PSP clinical diagnosis criteria [16], MSA according to consensus diagnostic criteria for MSA [17] and PD based on the UK Parkinson's Disease Society Bank Criteria [18]. The diagnosis was further confirmed by two neurologists (CB, JK) with experience in movement disorders. At the time of examination, all PD subjects were on stable dopaminergic medication for at least 4 weeks, consisting of levodopa and different dopamine agonists. In the PSP and MSA groups, medication consisted of various doses of levodopa alone or in combination with different dopamine agonists and/or amantadine. None of the patients received antipsychotic therapy. PSP and MSA patients were further rated by the natural history and neuroprotection in Parkinson plus syndromes–Parkinson plus scale (NNIPPS) [19] whilst PD patients were scored according to the Unified Parkinson's Disease Rating Scale motor subscore (UPDRS III). Item 18 of the UPDRS III was used for perceptual description of speech severity. Patient characteristics are summarized in Table 1.

The healthy control (HC) group consisted of 37 age-matched subjects (21 men, 16 women; mean age 63.1, SD 7.9, range 50–75 years) with no history of neurological or communication disorders. All subjects recruited were Czech native speakers.

Speech recordings

Speech recordings were performed in a quiet room with a low ambient noise level using a head-mounted condenser microphone (Bayerdynamic Opus 55, Heilbronn, Germany) situated approximately 5 cm from the mouth of each subject. Speech signals were sampled at 48 kHz with 16-bit resolution. Each participant was instructed to perform sustained phonation of the vowel/a/per one breath as long and steadily as possible, fast/pa/-/ta/-/ka/syllable repetition at least seven times per one breath and monologue on a given topic for approximately 90 s. All participants performed the sustained phonation and syllable repetition tasks twice with a relatively high test–retest reliability ($r = 0.77$–0.93, $p < 0.001$).

Dysarthria assessment

Quantitative acoustic vocal assessment was performed to investigate 16 deviant speech dimensions associated with hypokinetic, spastic or ataxic dysarthria [20, 21], which correspond to previous descriptions of speech and neuropathological findings in patients with PSP and MSA [7, 8]. The deviant speech dimensions investigated were selected considering the possibility of their objective assessment using acoustic analyses. In addition, these speech dimensions were chosen in order to be gender independent [10, 12]; there were no significant differences between male and female healthy participants across all investigated acoustic variables.

We evaluated eight dimensions widely observed in hypokinetic dysarthria of PD, including airflow insufficiency,

Table 1 Clinical characteristics of patients

	PSP Mean/SD (range)	MSA Mean/SD (range)	PD Mean/SD (range)
General			
Age (years)	65.8/5.4 (54–72)	60.8/4.9 (55–72)	61.1/6.5 (52–72)
Age of disease onset (years)	62.1/5.5 (50–68)	57.2/5.4 (50–70)	56.5/6.4 (47–67)
Symptom duration (years)	3.8/1.4 (1–6)	3.6/1.3 (2–6)	4.6/1.5 (1–6)
L-dopa equivalent (mg)	800/373 (500–1500)	899/394 (260–1480)	615/317 (300–1045)
Amantadine (mg)	200/107 (100–400)	300/89 (200–400)	
NNIPPS	66.3/28.7 (19–116)	78.5/19.9 (46–123)	
UPDRS III			15.9/7.4 (6–30)
UPDRS III speech 18 item	2.0/1.0 (0–3)	2.0/0.7 (1–3)	0.6/0.5 (0–1)
Subscores			
Tremor	2.5/2.6 (0–6)[a]	1.7/2.6 (0–9)[a]	2.1/2.5 (0–9)[b]
Rigidity	3.0/2.7 (0–7)[a]	4.7/3.2 (0–11)[a]	3.1/1.9 (1–7)[b]
Bradykinesia	20.6/11.3 (4–40)[a]	27.1/7.4 (16–39)[a]	6.1/2.7 (2–11)[b]
Bulbar/pseudobulbar	9.1/4.1 (3–17)[a]	7.9/2.3 (4–12)[a]	
Pyramidal	0.3/0.5 (0–1)[a]	0.8/1.2 (0–3)[a]	
Cerebellar	0.1/0.3 (0–1)[a]	5.6/7.1 (0–22)[a]	

NNIPPS natural history and neuroprotection on Parkinson plus syndromes-Parkinson plus scale, *UPDRS* unified Parkinson disease rating scale

[a] NNIPPS subscore

[b] UPDRS III subscore

harsh voice, rapid AMR, inappropriate silences, reduce loudness, monopitch, imprecise vowels and dysfluency. Considering elements of spastic dysarthria, we assessed strained-strangled voice quality, slow AMR and slow rate. To capture components related to ataxic dysarthria, we examined excess pitch fluctuations, vocal tremor, irregular AMR, prolonged phonemes and excess intensity variations. See Table 2 and Supplementary Material Online for comprehensive details on acoustic speech analyses.

Statistical analyses

Final values used for statistical analyses were calculated by averaging the data for each participant obtained in two vocal task runs. To assess group differences, each acoustic metric was compared across all three groups (PSP, MSA, PD) using a Kruskal–Wallis test with post hoc Bonferroni adjustment. Effect sizes were measured with Cohen's *d,* with $d > 0.5$ indicating a medium effect and $d > 0.8$ indicating a large effect. The Spearman coefficient was calculated to determine correlations between speech variables in APS and NNIPPS subscales. The level of significance was set to $p < 0.05$.

Estimation of the type and severity of dysarthria across individual patients was inspired by previous research on dysarthria in PSP and MSA [7, 8]. First, as the reference interval, the 5th and 95th percentile was calculated from the probability distribution of healthy controls across each acoustic measurement. The speech performance of each subject was then compared with the reference interval

Table 2 List of speech dimensions for hypokinetic, spastic and ataxic dysarthria

No.	Deviant speech dimension (weighting factor)[a]	Vocal task	Acoustic measure	Description
Hypokinetic				
1.	Airflow insufficiency (10 %)	Sustained phonation	Maximum phonation time (MPT)	Insufficient breath support for speech production;
2.	Harsh voice (10 %)	Sustained phonation	Jitter, Shimmer, Harmonics-to-noise ratio (HNR)	Harsh, rough and raspy voice;
3.	Rapid AMR (10 %)	Syllable repetition	Diadochokinetic (DDK) acceleration	Pace acceleration, rapid, blurred speech;
4.	Inappropriate silences (10 %)	Monologue	Percent pause time (PPT), Number of pauses (No. pauses)	Inappropriate silence intervals;
5.	Reduced loudness (20 %)	Monologue	Mean speech intensity (Mean Int)	Insufficiently loud, i.e. hypophonic voice;
6.	Monopitch (20 %)	Monologue	Pitch variability (F0 SD)	Monotone voice, lacking normal pitch and inflection changes;
7.	Imprecise vowels (10 %)	Monologue	Vowel articulation index (VAI)	Vowel sounds are distorted throughout their total duration;
8.	Dysfluency (10 %)	Monologue	Percent dysfluent words (PDW)	Involuntary repetition of speech movements, prolongation of sounds and vocal blocks;
Spastic				
9.	Strained-strangled voice (40 %)	Sustained phonation	Degree of voicelessness (DUV)	Voice (phonation) sounds strained or strangled (effortful squeezing of voice through glottis);
10.	Slow AMR (20 %)	Syllable repetition	DDK rate	Abnormally slow motion rate of articulators;
11.	Slow rate (40 %)	Monologue	Articulation rate	Abnormally slow rate of actual speech;
Ataxic				
12.	Excess pitch fluctuations (30 %)	Sustained phonation	Pitch variability (F0 SD)	Uncontrolled alterations in voice pitch;
13.	Vocal tremor (20 %)	Sustained phonation	Frequency tremor intensity index (FTRI)	Tremulous phonation;
14.	Irregular AMR (10 %)	Syllable repetition	DDK regularity	Rate alternates from slow to fast;
15.	Prolonged phonemes (10 %)	Syllable repetition	Vowel duration	Prolongation of phonemes;
16.	Excess intensity variations (30 %)	Monologue	Intensity variations (Int SD)	Sudden, uncontrolled alterations of loudness including both silence and quiet voice.

[a] The number in parentheses indicate weighting factors applied in computing severity and type of dysarthria. Higher factors are used for dimensions considered to be distinctive for each type of dysarthria

across all speech dimensions. If the subject speech performance did not match the reference interval, it was considered as affected. Weighting factors in percentages were then applied to all affected speech performances in order to enhance the impact of distinctive dimensions according to specific dysarthria type (Table 2) [7, 8, 20, 21]. A total score was obtained reflecting the degree of hypokinetic, spastic and ataxic dysarthria components; possible scores ranged from 0 to 100 % for each type of dysarthria.

We additionally introduced a classification experiment to determine the best combination of acoustic features and estimate their sensitivity and specificity in differentiating between PD, PSP and MSA groups. A support vector machine (SVM) with a Gaussian radial basis kernel was applied to search for all combinations across acoustic features. Subsequently, a cross-validation scheme was used to validate reproducibility of the SVM classifier, where the original data were randomly separated into a training subset composed of 75 % of the data and a testing subset containing 25 % of the data; this cross-validation process was repeated twenty times for each combination. The overall classification performance of the SVM-based model was computed as the average percentage of correctly classified subjects into an appropriate group through all twenty cycles. Comprehensive details on classification procedure has been published previously [22].

Results

Table 3 provides numerical data and comparison between PD, PSP and MSA across all 16 speech dimensions investigated. In comparing PSP and PD groups, statistical analyses revealed significant alterations in three hypokinetic dimensions of harsh voice, inappropriate silences and imprecise vowels, one spastic dimension of slow rate and two ataxic dimensions of excess pitch fluctuations and irregular AMR. Comparison between MSA and PD groups revealed significant differences in all five ataxic dimensions but only in one hypokinetic dimension of inappropriate silences and one spastic dimension of strained-strangled voice. Notably, only one dimension of speech dysfluency was able to significantly separate PSP and MSA groups.

At least one deviant speech dimension was found in all PD and APS speakers. The severity of dysarthria was similar in PSP and MSA patients but considerably greater than in the PD group (Fig. 1a). Eight PSP (68 %) and 12 MSA (92 %) patients exhibited dysarthria with a combination of all hypokinetic, ataxic and spastic components, whereas all PD patients (100 %) manifested pure hypokinetic dysarthria. Conversely, pure hypokinetic dysarthria was found only in one PSP patient (8 %) and was more severe than in any PD patient investigated, whereas the remaining PSP and MSA patients showed a combination of at least one affected hypokinetic and one spastic or ataxic component. Speech in PSP was primarily characterized by the occurrence of hypokinetic components (51 %) followed by spastic components (43 %), whereas speech in MSA was characterized by the occurrence of ataxic components (56 %) followed by spastic components (45 %) (Fig. 1b). The majority of PSP patients (83 %) showed predominant hypokinetic, spastic or hypokinetic-spastic dysarthria. MSA patients manifested either predominant ataxic dysarthria (46 %) or showed ataxic dysarthria with various combinations and severity of hypokinetic and spastic components (Fig. 1c). Table 4 summarizes our findings and details the percentage of affected patients across individual speech dimensions.

The combination of six acoustic features related to five deviant speech dimensions including harsh voice (jitter), inappropriate silences (percent pause time and number of pauses), slow AMR (diadochokinetic rate), excess intensity variation (intensity variation) and excess pitch fluctuation (pitch variation) were able to separate PD from APS with a very high classification accuracy of 95.3 ± 6.4 %, with a sensitivity of 93.4 ± 8.7 % and specificity of 99.5 ± 4.1 %. Furthermore, the four deviant speech dimensions including harsh voice (harmonics-to-noise ratio), fluency (percent dysfluent word), slow rate (articulation rate) and vocal tremor (frequency tremor intensity index) were able to discriminate PSP from MSA with an accuracy of 75.2 ± 13.3 (sensitivity of 74.3 ± 15.3 %, specificity 81.2 ± 17.7 %).

Acoustic assessment of the extent of dysarthria severity in APS showed significant correlation to overall NNIPPS score ($r = 0.54$, $p = 0.006$). In addition, the bulbar/pseudobulbar NNIPPS subscore correlated with the severity of spastic dysarthria components ($r = 0.42$, $p = 0.04$) and the cerebellar NNIPPS subscore showed a correlation trend with severity of ataxic dysarthria components ($r = 0.36$, $p = 0.07$). From individual speech patterns, only slow rate showed negative correlation to the bulbar/pseudobulbar NNIPPS subscore ($r = -0.47$, $p = 0.02$). There were no other significant correlations between speech parameters and NNIPPS subscores.

Discussion

The current study is the first quantitative, objective investigation attempting to broaden our knowledge concerning speech disorder in PSP and MSA. Our results show that the characteristics of speech disorder may reflect the underlying neuropathology of PD and APS. Dysarthria was uniformly present in all patients with PSP

Table 3 Results of acoustic speech analyses

No.	Deviant speech dimension acoustic measure (5th to 95th percentile of healthy control group)	Groups			Group differences	Effect size†		
		PSP Mean/SD (range)	MSA Mean/SD (range)	PD Mean/SD (range)	*p*	PSP vs. PD	MSA vs. PD	PSP vs. MSA
Hypokinetic								
1.	Airflow insufficiency							
	MPT (s) (11.7–27.1)	13.2/5.0 (7.6–23.5)	13.5/6.8 (6.4–33.6)	17.1/8.5 (7.6–43.0)	0.18	−0.56	−0.47	−0.05
2.	Harsh voice							
	Jitter (%) (0.26–1.77)	1.60/1.27 (0.43–4.40)	1.62/1.21 (0.29–4.27)	0.73/0.36 (0.35–1.83)	0.22	0.93	1.00	0.02
	Shimmer (%) (2.05–9.83)	8.48/3.12 (2.33–12.65)	8.58/3.92 (2.31–16.18)	5.40/2.76 (2.59–11.74)	0.03	1.05*	0.94	−0.02
	HNR (dB) (15.3–25.2)	15.0/3.9 (10.0–23.6)	16.4/5.3 (11.0–25.2)	20.4/2.6 (15.0–24.4)	0.008	−1.62*	−0.95	−0.29
3.	Rapid AMR							
	DDK acceleration (0.87–1.02)	1.06/0.47 (0.75–2.53)	1.00/0.16 (0.84–1.41)	0.95/0.13 (0.77–1.36)	0.85	0.32	0.32	0.18
4.	Inappropriate silences							
	PPT (%) (29.2–38.7)	37.1/2.8 (32.0–40.1)	36.9/4.8 (28.7–43.2)	33.8/2.5 (27.6–37.2)	0.005	1.59**	1.06*	0.05
	No. pauses (pauses/s) (3.11–5.03)	2.83/0.56 (1.63–3.53)	3.12/0.53 (2.16–4.11)	3.67/0.63 (2.88–4.09)	0.006	−1.39**	−0.93	−0.54
5.	Reduced loudness							
	Mean Int (dB) (59.7–67.9)	61.9/3.6 (55.7–66.6)	61.9/3.1 (57.9–68.1)	63.7/3.2 (58.1–69.4)	0.24	−0.52	−0.56	−0.01
6.	Monopitch							
	F0 SD (st) (1.71–3.29)	1.79/0.36 (1.05–2.20)	1.94/0.44 (1.32–2.67)	1.81/0.31 (1.40–2.49)	0.68	−0.08	0.32	−0.37
7.	Imprecise vowels							
	VAI (0.83–1.05)	0.81/0.06 (0.69–0.89)	0.83/0.06 (0.72–0.90)	0.88/0.06 (0.79–0.99)	0.02	−1.28**	−0.93	−0.27
8.	Dysfluency							
	PDW (%) (0.83–6.86)	7.83/5.08 (2.34–18.95)	3.60/2.45 (0.79–7.79)	4.75/2.76 (1.21–10.84)	0.03	0.75	−0.44	1.06*
Spastic								
9.	Strained-strangled voice							
	DUV (%) (0–2.23)	3.57/5.75 (0–18.05)	11.21/22.09 (0–81.35)	0.20/0.52 (0–1.55)	0.005	0.83	0.71**	−0.47
10.	Slow AMR							
	DDK rate (syll/s) (5.49–8.03)	5.72/1.32 (3.60–8.03)	5.45/1.32 (3.42–7.61)	6.82/1.12 (5.51–9.69)	0.03	−0.90	−1.12	0.20
11.	Slow rate							
	Articulation rate (word/s) (1.93–3.52)	2.34/0.50 (1.72–3.16)	2.49/0.59 (1.52–3.28)	3.03/0.60 (2.07–3.91)	0.02	−1.23*	−0.91	−0.26
Ataxic								
12.	Excess pitch fluctuations							
	F0 SD (st) (0.16–0.80)	0.71/0.41 (0.28–1.43)	1.02/0.61 (0.16–2.32)	0.34/0.14 (0.18–0.68)	<0.001	1.19*	1.54***	−0.61
13.	Vocal tremor							
	FTRI (%) (0.16–1.11)	0.86/0.53 (0.29–2.19)	1.81/1.58(0.22–5.39)	0.51/0.23 (0.22–1.02)	0.02	0.86	1.15**	−0.81

Table 3 continued

No.	Deviant speech dimension acoustic measure (5th to 95th percentile of healthy control group)	Groups			Group differences	Effect size†		
		PSP	MSA	PD				
		Mean/SD (range)	Mean/SD (range)	Mean/SD (range)	*p*	PSP vs. PD	MSA vs. PD	PSP vs. MSA
14.	Irregular AMR							
	DDK regularity (ms) (9.7–35.4)	51.1/40.3 (9.8–131.0)	43.7/27.1 (12.9–94.9)	18.6/8.8 (5.4–35.4)	0.009	1.12*	1.25*	0.22
15.	Prolonged phonemes							
	Vowel duration (%) (30.0–48.5)	46.4/18.7 (27.3–95.0)	49.5/10.8 (34.7–69.1)	37.9/5.8 (30.1–47.1)	0.04	0.61	1.34*	−0.20
16.	Excess intensity variations							
	Int SD (dB) (6.13–8.59)	7.96/0.79 (6.34–9.25)	8.50/1.05 (6.73–10.0)	7.11/0.67 (5.90–8.37)	0.009	1.17	1.58**	−0.58

MPT maximum phonation time, maximum duration of prolonged vowel; *Jitter* frequency instability of the vocal folds; *Shimmer* amplitude instability of the vocal folds; *HNR* harmonics-to-noise ratio, addition of noise in speech; *DDK acceleration* diadochokinetic acceleration, increased rate of syllable vocalization per second through time; *PPT* percent pause time, pause time percentage relative to total speech time; *No. pauses* number of pauses relative to total speech time; *Mean Int* average level of speech loudness; *F0 SD* variations of vibration rate of the vocal folds; *VAI* vowel articulation index, centralization of formant frequencies; *PDW* percent dysfluent words, number of dysfluencies relative to number of words; *DUV* degree of voicelessness, aperiodicity of voice; *DDK rate* diadochokinetic rate, rate of syllable vocalizations per second; *Articulation rate* words per seconds related to total speech time after removal of pauses; *FTRI* frequency tremor intensity index, indicating vocal tremor; *DDK regularity* diadochokinetic regularity, ability to maintain a constant rate of consonant–vowel combinations; *Vowel duration* duration of vowel related to duration of syllable; *Int SD* variation of loudness; *AMR* alternating motion rates

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

† Effect size 0.8 is considered large, 0.5 is considered medium, and 0.2 is considered small

Fig. 1 Characteristics of dysarthria: **a** percentage of affected patients according to dysarthria severity; **b** percentage of deviant speech dimensions according to individual dysarthric components; **c** percentage of patients according to predominant type of their dysarthria

and MSA and generally consisted of a combination of hypokinetic, spastic and ataxic components, whereas PD patients manifested pure hypokinetic elements. Therefore, using objective speech measurements, we were able to discriminate between APS and PD with 95 % accuracy. Moreover, the speech of PSP patients was characterized by the predominant occurrence of hypokinetic-spastic dysarthria whereas MSA patients manifested predominantly ataxic dysarthria, resulting in a discrimination accuracy of 75 % in the differentiation between PSP and MSA groups.

In contrast to previous perceptual examinations suggesting predominant spastic components in PSP and hypokinetic in MSA [7, 8], we objectively detected

Table 4 Characteristics of deviant speech dimensions

No.	Deviant speech dimension[a]	PSP	MSA	PD
Hypokinetic				
1.	Airflow insufficiency	Common (42 %)	Common (31 %)	Common (27 %)
2.	Harsh voice	Abundant (75 %)	Frequent (69 %)	Occasional (13 %)
3.	Rapid AMR	Occasional (25 %)	Common (31 %)	Occasional (13 %)
4.	Inappropriate silences	Abundant (83 %)	Frequent (69 %)	Common (27 %)
5.	Reduced loudness	Occasional (25 %)	Occasional (23 %)	Occasional (13 %)
6.	Monopitch	Frequent (50 %)	Common (31 %)	Frequent (53 %)
7.	Imprecise vowels	Abundant (75 %)	Frequent (62 %)	Common (33 %)
8.	Dysfluency	Frequent (58 %)	Occasional (15 %)	Occasional (13 %)
Spastic				
9.	Strained-strangled voice	Common (42 %)	Frequent (62 %)	Rare (0 %)
10.	Slow AMR	Frequent (50 %)	Frequent (54 %)	Rare (0 %)
11.	Slow rate	Common (42 %)	Occasional (23 %)	Rare (0 %)
Ataxic				
12.	Excess pitch fluctuations	Common (33 %)	Frequent (69 %)	Rare (0 %)
13.	Vocal tremor	Common (33 %)	Frequent (54 %)	Rare (0 %)
14.	Irregular AMR	Common (33 %)	Common (31 %)	Rare (0 %)
15.	Prolonged phonemes	Occasional (25 %)	Frequent (54 %)	Rare (0 %)
16.	Excess intensity variations	Occasional (25 %)	Frequent (54 %)	Rare (0 %)

AMR alternating motion rates

[a] The parentheses represent percentage of affected persons according to specific speech dimension: 0–10 % subjects affected are considered rare, 11–25 % occasional, 26–45 % common, 46–70 % frequent, and 71–100 % abundant

predominant hypokinetic components in PSP and ataxic in MSA. Interestingly, ataxic components were predominant even though the majority of our patients were MSA-P, probably reflecting great sensitivity of speech to minor cerebellar deficits. Furthermore, dysarthria was perceptually estimated to be less severe in MSA than PSP [23], whereas dysarthria was more severe in our MSA patients, probably as a result of greater disease disability. On the other hand, we may hypothesize that predominant ataxic dysarthria in MSA is perceptually more intelligible than hypokinetic dysarthria in PSP. Indeed, listeners who heard and subsequently transcribed ataxic speech benefited more from its exposure than did listeners who heard and then transcribed hypokinetic speech [24].

Recognizing characteristic deviant speech dimensions may have important implications in improving the accuracy of early clinical diagnosis [7, 8]. Dysarthria in PSP and MSA differed from that in PD due to greater severity and the presence of spastic and ataxic components. In the present study, at least one spastic or ataxic deviant speech dimension was detected in almost every PSP and MSA patient, including those with short disease duration. In comparing PSP and MSA, in addition to hypophonic monotony of parkinsonian speech, dysarthria in our PSP patients was dominated by increased dysfluency, decreased slow rate, inappropriate silences, deficits in vowel articulation and harsh voice quality, whereas patients with MSA more frequently manifested pitch fluctuations, excess intensity variations, prolonged phonemes, vocal tremor and strained-strangled voice quality.

Dysfluency was the only single speech aspect distinctive for PSP but was rarely observed in MSA. In particular, only two of our MSA patients showed increased dysfluencies, which were rather associated with cluttering in one case and poor working memory in the second case, as opposed to the stuttering-like behaviour typically observed in PSP and later stages of PD [7, 25]. The occurrence of stuttering-like behaviour may be due to involvement of the globus pallidus and primary motor cortex, which represent regions of the brain commonly affected in PSP [26]. In fact, stuttering was reported as a consequence of pallidal deep brain stimulation in patients with dystonia [27] and was widely present in manganese-induced ephedrone parkinsonism associated with toxic and neurodegenerative damage to globus pallidus [12]. In addition, motor planning responsible for control of fluency has recently been suggested to be coded in the left primary motor cortex whereas this speech motor-related asymmetry was missing in stuttering [28]. Yet, it has been shown that increased dopamine levels in PD may lead to the emergence of stuttering [29, 30], where the motor cortex may play a similar role as in the case of levodopa-induced dyskinesia [31].

In addition, MSA patients showed overall poorer voice control in comparison with PSP. The strained-strangled voice quality, excess pitch fluctuation and vocal tremor observed in MSA patients may together give the perceptual impression of quivery-croaky strained speech with increased pitch, whereas severe harshness in the voice of PSP subjects may resemble growling dysarthria. These aspects contributing to decreased quality of voice probably arise

due to uncontrolled movements of the laryngeal muscles, fluctuation of vocal fold tension and incomplete vocal fold closure, representing a rather non-specific marker of neuronal dysfunction. Speech in PSP may be further characterized by a slower rate accompanied by inappropriate silence intervals, which was also noted in patients with Huntington's disease [32] and thus it may be hypothesized as a result of damage to the striatum and generally more widespread neuronal atrophy. Furthermore, PSP patients manifested more affected vowel articulation than MSA, which may also contribute to a perceived reduction in intelligibility in PSP in comparison with MSA [23]. Conversely, speech in MSA exhibited more prolonged phonemes and excess intensity variations that substantially contributed to the perceptual impression of scanning dysarthria.

Predominant hypokinetic-spastic dysarthria with fewer ataxic components in our PSP group is consistent with observed widespread neurodegeneration involving the midbrain as well as the globus pallidus, striatum, hypothalamic nucleus, pons, superior cerebellar peduncle and cerebella dentate nucleus [26]. The clinical features of the dysarthria in our MSA patients showing predominant ataxic dysarthria with fewer spastic and hypokinetic components conform to the known neuropathological changes which include degeneration of cerebellum, middle cerebellar peduncle, striatum, substantia nigra, inferior olivary nucleus and pons [33]. However, only one previous neuropathological study identified relationship between the severity of hypokinetic components in PSP and the degree of neuronal loss and gliosis in the substantia nigra [34]. Our current findings support the role of corticobulbar pathways and the cerebellum in the development of mixed dysarthria in APS as we observed relationships between the severity of spastic components and bulbar/pseudobulbar manifestations, as well as between the severity of ataxic components and cerebellar signs.

The results of the present study indicate the potential of speech analyses in the differentiation of PD from APS, with 93 % sensitivity and 100 % specificity in patients with an average symptom duration longer than 2 years. These results are similar to recent neuroimaging studies reporting comparable sensitivity and specificity in metabolic pattern analysis or Diffusion Tensor Imaging in the differential diagnosis of parkinsonism [4, 35]. In addition, our classification results between PD and APS seem to be superior to very recently introduced breath analysis, which showed 88 % sensitivity and 88 % accuracy [36]. However, it is noteworthy to point out that our speech-based classification between PSP and MSA provided only 74 % sensitivity and 81 % specificity, whilst previous neuroimaging studies have reported 90 % sensitivity and 100 % specificity [4, 35].

Certain limitations of the present study should be noted. As our PD patients were investigated in their ON condition, we cannot exclude that some differences between PD and APS were more pronounced due to the beneficial effect of dopaminergic therapy. However, it is assumed that short-term dopaminergic therapy has no or very little effect on speech in PD [37]. We did not differentiate between speech in the various subtypes of PSP and MSA due to the limited opportunity in recruiting a larger number of participants. Nevertheless, at least in PSP patients, different subtypes of disease seem to have no substantial effect on global speech performance [14].

Objective identification of deviant speech dimensions can be diagnostically helpful in a number of neurological disorders and may provide measures of treatment response and disease progression. Future studies should further elaborate and extend our findings as well as show the sensitivity of speech in the differentiation between PD, PSP and MSA in very early disease stages.

Acknowledgments This study was supported by the Czech Ministry of Health (NT 12288-5).

Conflicts of interest The authors report no conflicts of interest.

Ethical standard Each participant provided written, informed consent. The study was approved by the Ethics Committee of the General University Hospital in Prague, Czech Republic, and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

References

1. Schrag A, Ben-Shlomo Y, Quinn NP (1999) Prevalence of progressive supranuclear palsy and multiple system atrophy: a cross-sectional study. Lancet 354:1771–1775
2. Wenning GK, Litvan I, Tolosa E (2011) Milestones in atypical and secondary Parkinsonisms. Mov Disord 26:1083–1095
3. Berti V, Pupi A, Monsconi L (2011) PET/CT in diagnosis of movement disorders. Ann N Y Acad Sci 1228:93–108
4. Tang CC, Poston KL, Eckert T, Feigin A, Frucht S et al (2010) Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis. Lancet Neurol 9:149–158
5. Rusz J, Cmejla R, Ruzickova H, Klempir J, Majerova V et al (2011) Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test. Mov Disord 26:1951–1952
6. Ho AK, Iansek R, Marigliani C, Bradshaw J, Gates S (1998) Speech impairment in large sample of patients with Parkinson's disease. Behav Neurol 11:131–137
7. Kluin KJ, Foster NL, Berent S, Gilman S (1993) Perceptual analysis of speech disorders in progressive supranuclear palsy. Neurology 43:563–566
8. Kluin KJ, Gilman S, Lohman M, Junck L (1996) Characteristics of the dysarthria in multiple system atrophy. Arch Neurol 53:545–548
9. Kim Y, Kent RD, Kent J, Duffy JR (2010) Perceptual and acoustic features of dysarthria associated with multiple system atrophy. J Med Speech Lang Pathol 18:66–70

10. Rusz J, Cmejla R, Ruzickova H, Ruzicka E (2011) Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. J Acoust Soc Am 129:350–369
11. Rusz J, Cmejla R, Tykalova T, Ruzickova H, Klempir J et al (2013) Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task. J Acoust Soc Am 134:2171–2181
12. Rusz J, Megrelishvili M, Bonnet C, Okujava M, Brozova H et al (2014) A distinct variant of mixed dysarthria reflects parkinsonism and dystonia due to ephedrone abuse. J Neural Transm 121:655–664
13. Sachin S, Shukla G, Goyal V, Singh S, Aggarwal V et al (2008) Clinical speech impairment in Parkinson's disease, progressive supranuclear palsy, and multiple system atrophy. Neurol India 56:122–126
14. Skodda S, Visser W, Schlegel U (2011) Acoustical analysis of speech in progressive supranuclear palsy. J Voice 25:725–731
15. Saxena M, Behari M, Kumaran SS, Goyal V, Narang V (2014) Assessing speech dysfunction using BOLD and acoustic analysis in parkinsonism. Parkinsonism Relat D 20:855–861
16. Litvan I, Agid Y, Calne D, Campbell G, Dubois B et al (1996) Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele–Richardson–Olszewski syndrome): report of the NINDS-SPSS international workshop. Neurology 47:1–9
17. Gilman S, Wenning GK, Low PA, Brooks DJ, Mathias CJ et al (2008) Second consensus statement on the diagnosis of multiple system atrophy. Neurology 71:670–676
18. Hughes AJ, Daniel SE, Kilford L, Lees AJ (1992) Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 55:181–184
19. Payan CA, Viallet F, Landwerhrmeyer BG, Bonnet AM, Borg M et al (2011) Disease severity and progression in progressive supranuclear palsy and multiple system atrophy: validation of the NNIPPS-Parkinson plus scale. PLoS One 6:e22293
20. Duffy JR (2013) Motor Speech Disorders: Substrates, Differential Diagnosis and Management, 3rd edn. Mosby, St. Louis
21. Darley FL, Aronson AE, Brown JR (1969) Differential diagnostic patterns of dysarthria. J Speech Hear Res 12:246–269
22. Novotny M, Rusz J, Cmejla R, Ruzicka E (2014) Automatic evaluation of articulatory disorders in Parkinson's disease. IEEE/ACM T Audio Speech Lang Process 22:1366–1378
23. Hartelius L, Gustavsson H, Astrand M, Holmberg B (2006) Perceptual analysis of speech in multiple system atrophy and progressive supranuclear palsy. J Med Speech Lang Pathol 14:241–247
24. Liss JM, Spitzer SM, Caviness JN, Adler C (2002) The effect of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. J Acoust Soc Am 112:3022–3030
25. Benke T, Hohenstein C, Poewe W, Butterworth B (2000) Repetitive speech phenomena in Parkinson's disease. J Neurol Neurosurg Psychiatry 69:319–325
26. Nath U, Ben-Shlomo Y, Thomson RG, Lees AJ, Burn DJ (2003) Clinical features and natural history of progressive supranuclear palsy: a clinical cohort study. Neurology 69:910–916
27. Nebel A, Reese R, Deuschl G, Mehdorn HM, Volkmann J (2009) Acquired stuttering after pallidal deep brain stimulation for dystonia. J Neural Transm 116:167–169
28. Neef NE, Hoang TNL, Neef A, Paulus W, Sommer M (2015) Speech dynamics are coded in the left motor cortex in fluent speakers but not in adults who stutter. Brain. doi:10.1093/brain/awu390 (in press)
29. Tykalova T, Rusz J, Cmejla R, Klempir J, Ruzickova H, et al (2015) Effect of dopaminergic medication on speech dysfluency in Parkinson's disease: a longitudinal study. J Neural Transm: in press. doi:10.1007/s00702-015-1363-y
30. Louis ED, Winfield L, Fahn S, Ford B (2001) Speech dysfluency exacerbated by levodopa in Parkinson's disease. Mov Disord 16:562–565
31. Ostock CY, Dupre KB, Jaunarajs KL, Walters H, George J et al (2011) Role of the primary motor cortex in L-Dopa-induced dyskinesia and its modulation by 5-HT1A receptor stimulation. Neuropharmacology 61:753–760
32. Rusz J, Klempir J, Tykalova T, Baborova E, Cmejla R et al (2014) Characteristics and occurrence of speech impairment in Huntington's disease: possible influence of antipsychotic medication. J Neural Transm 121:655–664
33. Gilman S, Wenning GK, Low PA, Brooks DJ, Mathias CJ et al (2001) Second consensus statement on the diagnosis of multiple system atrophy. Neurology 71:670–676
34. Kluin KJ, Gilman S, Foster NL, Sima AAF, D'Amato CJ et al (2001) Neuropathological correlates of dysarthria in progressive supranuclear palsy. Arch Neurol 58:265–269
35. Prodoehl J, Li H, Planetta JP, Goetz CG, Shannon KM, Tangonan R et al (2013) Diffusion Tensor Imaging of Parkinson's disease, atypical parkinsonism, and essential tremor. Mov Disord 28:1816–1822
36. Nakleh MK, Badarny S, Winer R, Jeries R, Finberg J, Haick H. (2015) Distinguishing idiopathic Parkinson's disease from other parkinsonian syndromes by breath test. Parkinsonism Relat D. doi:10.1016/j.parkreldis.2014.11.023 (in press)
37. Ho AK, Bradshaw R, Iansek R (2008) For better or worse: the effect of levodopa on speech in Parkinson's disease. Mov Disord 23:575–580

Appendix A7

Quantitative assessment of motor speech abnormalities in idiopathic REM sleep behaviour disorder

Rusz J, Hlavnicka J, Tykalova T, Buskova J, Ulmanova O, Ruzicka E, Sonka K (2016) Quantitative assessment of motor speech abnormalities in idiopathic REM sleep behaviour disorder. *Sleep Med* **19**:141-147.

Journal name (category): **Sleep Medicine** (clinical neurology)
Quartile in category: **Q2**
Impact factor: **3.339**
Number of WoS citations (excluding autocitations): **4**

Sleep Medicine 19 (2016) 141–147

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Original Article

Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behaviour disorder

Jan Rusz [a,b,*], Jan Hlavnička [a], Tereza Tykalová [a], Jitka Bušková [b,c], Olga Ulmanová [b], Evžen Růžička [b], Karel Šonka [b]

[a] Faculty of Electrical Engineering, Department of Circuit Theory, Czech Technical University in Prague, Prague, Czech Republic
[b] Department of Neurology and Centre of Clinical Neuroscience, Charles University in Prague, First Faculty of Medicine, Prague, Czech Republic
[c] National Institute of Mental Health, Czech Republic

ARTICLE INFO

Article history:
Received 7 May 2015
Received in revised form 8 July 2015
Accepted 17 July 2015
Available online 14 September 2015

Keywords:
Parkinson's disease
Parkinsonism
Speech and voice disorders
Dysarthria
Acoustic analyses

ABSTRACT

Objective: Patients with idiopathic rapid eye movement sleep behaviour disorder (RBD) are at substantial risk for developing Parkinson's disease (PD) or related neurodegenerative disorders. Speech is an important indicator of motor function and movement coordination, and therefore may be an extremely sensitive early marker of changes due to prodromal neurodegeneration.
Methods: Speech data were acquired from 16 RBD subjects and 16 age- and sex-matched healthy control subjects. Objective acoustic assessment of 15 speech dimensions representing various phonatory, articulatory, and prosodic deviations was performed. Statistical models were applied to characterise speech disorders in RBD and to estimate sensitivity and specificity in differentiating between RBD and control subjects.
Results: Some form of speech impairment was revealed in 88% of RBD subjects. Articulatory deficits were the most prominent findings in RBD. In comparison to controls, the RBD group showed significant alterations in irregular alternating motion rates ($p = 0.009$) and articulatory decay ($p = 0.01$). The combination of four distinctive speech dimensions, including aperiodicity, irregular alternating motion rates, articulatory decay, and dysfluency, led to 96% sensitivity and 79% specificity in discriminating between RBD and control subjects. Speech impairment was significantly more pronounced in RBD subjects with the motor score of the Unified Parkinson's Disease Rating Scale greater than 4 points when compared to other RBD individuals.
Conclusion: Simple quantitative speech motor measures may be suitable for the reliable detection of prodromal neurodegeneration in subjects with RBD, and therefore may provide important outcomes for future therapy trials.

1. Introduction

Idiopathic rapid eye movement sleep behaviour disorder (RBD) is a parasomnia characterised by dream-enactment behaviours associated with REM sleep without muscle atonia [1]. Recent studies have shown that patients diagnosed with RBD are at increased risk for developing α-synucleinopathy, particularly Parkinson's disease (PD) or dementia with Lewy bodies (DLB), and less frequently multiple system atrophy (MSA) [2–4], with risk estimates of 33.1% at five years, 75.7% at 10 years, and 90.9% at 14 years after onset [5]. This high conversion rate to neurodegenerative disease provides a unique opportunity to observe the clinical development of parkinsonism or cognitive impairment [6]. Identifying predictive markers of neurodegeneration is essential [4,6] as they could provide invaluable information for future trials and disease-modifying therapies before the onset of motor and cognitive symptoms [7].

Motor speech disorder is a common clinical manifestation occurring in 70%–100% of patients with PD, DLB, and MSA, and typically appears early in the course of disease [8–11]. Hypokinetic dysarthria tends to be the dominant subtype in PD and DLB, whereas ataxic–hypokinetic dysarthria prevails in MSA [8–11]. Hypokinetic dysarthria affects primarily phonatory, articulatory, and prosodic speech subsystems, and may be related to numerous deviant dimensions such as reduced vocal loudness, poor voice quality, harshness, articulatory undershoot of vowels and consonants, dysrhythmia, articulatory decay, monopitch, monoloudness, variability of speech rate, and dysfluency [12]. Ataxic dysarthria is characterised by distorted articulation, reduced speech rate, and deviant prosodic modulations, particularly rhythmical irregularities during fast repetitive productions of syllables [12].

* Corresponding author. Department of Neurology and Centre of Clinical Neuroscience, Charles University in Prague, First Faculty of Medicine, Kateřinská 30, 12000 Prague, Czech Republic. Tel.: +420 224 352 287; fax: +420 224 311 081.
E-mail address: rusz.mz@gmail.com (J. Rusz).

http://dx.doi.org/10.1016/j.sleep.2015.07.030
1389-9457/

Speech disorders may be a prodromal sign of PD, as speech dysfunction is present in up to 90% of de novo PD patients [13–15]. Family members of PD patients have perceptually noted changes in speech before the diagnosis was established [16], and previous studies have reported cases in which reduced intonation variability was observed several years before the onset of the first motor symptoms [17]. Furthermore, ultrasonic vocalisation deficits were among the first prodromal markers of motor dysfunction in a murine model of PD [18]. Based on the Unified Parkinson's Disease Rating Scale, Postuma et al. [6] estimated that vocal and facial akinesia is the earliest indicator of parkinsonism in RBD patients, followed by rigidity, gait abnormalities, limb bradykinesia, and tremor.

An objective, quantitative assessment of speech in RBD is currently lacking. Speech evaluation is inexpensive, noninvasive, and simple to administer, and acoustic analyses provide objective, sensitive, and quantifiable information for the precise assessment of various deviant speech dimensions [13]. In addition, current advances in information and communication technologies have provided speech assessment the unique opportunity to be considered a simple screening test for the development of parkinsonism [13,14]. However, speech abnormalities in RBD should first be well explored.

Therefore, the aims of the current study were as follows: (1) to propose an acoustic methodology that would be sensitive to potential motor speech deficits in RBD; (2) to quantitatively characterise speech disorders in RBD; (3) to determine the most salient features of speech dysfunction in RBD, and to estimate their specificity and sensitivity in differentiating between RBD and healthy control subjects; and (4) to explore the relationship between speech and clinical findings to provide deeper insight into the pathophysiology of speech dysfunction in RBD.

2. Methods

2.1. Patients

A total of 16 consecutive Czech patients (10 men, 6 women), mean age 65.6 years [standard deviation (SD) 7.0 years], diagnosed with idiopathic RBD according to the International Classification of Sleep disorders diagnostic criteria, second edition [19], were included in the study. The examination consisted of detailed medical and pharmacological history, neurological assessment, and night polysomnography from 10 PM to 6 AM during a 1-day hospitalisation. Polysomnographic features of RBD were analysed from the chin and tibialis superficialis muscles according to the American Association of Sleep Medicine scoring rules [20]. Five subjects were treated with antidepressants before the diagnosis of RBD was established (Table 1), but only two subjects were receiving antidepressants at the time of the diagnostic polysomnography (RBD07, RBD15). Diagnostic investigation revealed that seven subjects fulfilled the criteria of obstructive sleep apnoea [19]. The mean apnoea/hypopnoea index (AHI) was 9.1 (SD 6.9); the AHI did not exceed the value of 20 in any RBD patient. The average number of periodic limb movements in sleep (PLMS) per one hour was 16.1 (SD 34.5); a value greater than 15 was found in four RBD patients (RBD05, RBD12, RBD14, RBD16).

At the time of speech investigation, nine of 16 patients were treated by clonazepam at bedtime to alleviate symptoms of RBD. None of the patients complained of motor or cognitive difficulties or had a history of treatment with antiparkinsonian medication or any other therapy influencing sleep, cognition, or motor features. All patients were examined by a movement disorders specialist (O.U.) and scored according to the Unified Parkinson's Disease Rating Scale motor subscore (UPDRS III). The clinical characteristics of the RBD subjects are summarised in Table 1.

The healthy control group consisted of 16 sex- and age-matched subjects (10 men, 6 women), mean age 65.6 years (SD 7.0 years), with no history of neurological or communication disorders or abnormalities of sleep. Each participant provided written informed consent. The study was approved by the Ethics Committee of the General University Hospital in Prague, Czech Republic, and has therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

2.2. Speech examination

Speech recordings were performed in a quiet room with a low ambient noise level, in the afternoon, using a head-mounted condenser microphone (Beyerdynamic Opus 55, Heilbronn, Germany) placed approximately 5 cm from the subject's mouth. Speech signals were sampled at 48 kHz with 16-bit resolution. Recording was performed in each subject during a single session with a speech specialist (J.R.). All participants were instructed to perform three

Table 1
Clinical characteristics of RBD patients.

Patient no.	Sex	Age (y)	RBD symptoms duration (y)	Antidepressant therapy before RBD diagnosis	UPDRS III motor score	UPDRS III speech item 18	Clonazepam (mg/day)
RBD01	F	73	1	None	5	1	0.125
RBD02	M	54	8	None	1	0	0
RBD03	M	66	11	None	7	0	0.5
RBD04	M	59	3	None	2	0	0
RBD05	M	75	16	None	10	0	0.5
RBD06	M	70	10	None	7	1	0.5
RBD07	F	64	3	SSRI, NaSSA	9	1	0.5
RBD08	F	71	5	SNRI	12	0	0.5
RBD09	M	69	5	None	2	0	0
RBD10	F	57	1	None	3	0	0
RBD11	F	68	1	SSRI	7	0	1
RBD12	F	67	5	SSRI	2	0	0.125
RBD13	M	68	11	None	4	0	2
RBD14	M	65	12	None	1	0	0
RBD15	M	51	10	SSRI, SARI	0	0	0
RBD16	M	73	5	None	3	0	0
Mean (SD)		65.6 (7.0)	6.7 (4.6)		4.7 (3.6)	0.19 (0.40)	0.36 (0.53)

Abbreviations: F, female; M, male; NaSSA, noradrenergic and specific serotonergic antidepressant; RBD, rapid eye movement sleep behaviour disorder; SARI, serotonin antagonist reuptake inhibitor; SD, standard deviation; SNRI, serotonin–noradrenaline reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor; UPDRS III, Unified Parkinson's Disease Rating Scale motor subscore.

vocal tasks as follows: (1) sustained phonation of the vowel /a/ per one breath for as long and as steadily as possible; (2) fast /pa/-/ta/-/ka/ syllable repetition at least seven times per one breath; and (3) monologue for approximately 90 seconds on a given topic. These three speaking tasks were chosen because they can provide most of the information necessary for the objective description and interpretation of motor speech disorders. The sustained phonation and syllable repetition tasks were repeated twice for every subject per session, with a satisfactory test–retest reliability (Pearson: $r = 0.69–0.86$, $p < 0.001$).

2.3. Acoustic analyses

We performed quantitative vocal assessment, investigating 15 deviant speech dimensions connected with phonatory, articulatory, and prosodic dimensions [12]. The selected speech dimensions allowed objective assessment using acoustic analyses, which provide a noninvasive, precise, inexpensive, and reliable method to describe speech abnormalities. In addition, all speech dimensions investigated in the current study were chosen to be gender independent [13]. The definitions of all acoustic parameters are summarised in Table 2.

To assess phonatory characteristics, we examined airflow insufficiency (maximum phonation time, MPT), irregular pitch fluctuations (standard deviation of pitch sequence, F0 SD), signal perturbations [frequency microinstability (jitter)], increased noise [harmonics-to-noise ratio (HNR)], and aperiodicity [degree of voicelessness (DUV)] [13,21].

To explore articulatory characteristics, we evaluated imprecise vowels [vowel articulation index (VAI)] [22]. We further extracted imprecise consonants [voice onset time (VOT)], slow alternating motion rates (AMR) using the diadochokinetic (DDK) task (DDK rate), and irregular AMR [diadochokinetic regularity (DDK reg)] [15]. In addition, resonant frequency attenuation (RFA) represents a novel measurement sensitive to articulatory decay of natural spontaneous speech (Appendix S1).

To examine prosodic characteristics, we calculated reduced loudness [average level of speech intensity (Mean Int)], monoloudness [standard deviation of intensity contour (Int SD)], monopitch [standard deviation of pitch contour (F0 SD)], and inappropriate silences [number of pauses (NoP)] [13]. In addition, we investigated dysfluency [percent dysfluent words (PDW)] [23]. The more detailed descriptions of individual traditional acoustic measures may be found in a previous study [11].

Table 2
Overview of applied speech measurements.

Deviant speech dimension (derived from vocal task)	Acoustic feature	Definition of acoustic feature	Feature extraction/method (studies with detailed description)
Phonation			
Airflow insufficiency (sustained phonation)	MPT	Maximum phonation time, aerodynamic efficiency of the vocal tract measured as the maximum duration of the prolonged vowel.	Automatic/algorithm (time duration of sustained phonation)
Irregular pitch fluctuations (sustained phonation)	F0 SD	Standard deviation of fundamental frequency (F0 SD), variation in frequency of vocal fold vibration. The F0 sequence (pitch) was converted to a semitone scale to avoid differences in gender.	Automatic/PRAAT (Boersma and Weenink [21], Rusz et al. [13])
Signal perturbations (sustained phonation)	Jitter	Frequency perturbation, extent of variation of the voice range. Jitter is defined as the variability of the F0 of speech from one cycle to the next.	Automatic/PRAAT (Boersma and Weenink [21], Rusz et al. [13])
Increased noise (sustained phonation)	HNR	Harmonics-to-noise ratio, the amount of noise in the speech signal, mainly due to incomplete vocal fold closure. HNR is defined as the amplitude of noise relative to tonal components in speech.	Automatic/PRAAT (Boersma and Weenink [21], Rusz et al. [13])
Aperiodicity (sustained phonation)	DUV	Degree of unvoiced segments, the fraction of pitch frames marked as unvoiced. A frame was considered unvoiced if it had voicing strength below the voicing threshold of 0.45.	Automatic/PRAAT (Boersma and Weenink [21])
Articulation			
Imprecise vowels (monologue)	VAI	Vowel articulation index, based on formant centralisation, defined as VAI = (F1a + F2i)/(F1i + F1u + F2a + F2u). F1 and F2 for each vowel were averaged by the extraction of ten defined corner vowels.	User-controlled analyses/PRAAT (Rusz et al. [22])
Imprecise consonants (syllable repetition)	VOT	Voice onset time, defined as the length of the entire consonant from initial burst to vowel onset.	Automatic/algorithm (Novotny et al. [15])
Slow AMR (syllable repetition)	DDK rate	Diadochokinetic rate, representing the number of syllable vocalisations per second.	Automatic/algorithm (Novotny et al. [15])
Irregular AMR (syllable repetition)	DDK reg	Diadochokinetic regularity, defined as the standard deviation of distances between following syllables nuclei.	Automatic/algorithm (Novotny et al. [15])
Articulatory decay (monologue)	RFA	Resonant frequency attenuation, representing decrease of spectral energy as a result of decayed articulatory movements.	Automatic/algorithm (Appendix S1)
Prosody			
Reduced loudness (monologue)	Mean Int	Mean speech loudness, representing average squared amplitude within a predefined time–energy segment.	Automatic/algorithm (Rusz et al. [13])
Monoloudness (monologue)	Int SD	Speech loudness variation, defined as standard deviation of intensity contour after removing a period of silence exceeding 60 ms.	Automatic/algorithm (Rusz et al. [13])
Monopitch (monologue)	F0 SD	Pitch variation, defined as standard deviation of F0 contour converted to semitone scale.	Automatic/algorithm (Rusz et al. [13])
Inappropriate silences (monologue)	NoP	Number of pauses relative to total speech time after removing periods of silence lasting less than 60 ms.	Automatic/algorithm (Rusz et al. [13])
Dysfluency (monologue)	PDW	Percentage of dysfluent words, defined as number of dysfluent events normalised by the total number of words.	Auditory-perceptual/PRAAT (Tykalova et al. [23])

Numbers in square brackets are reference numbers.

2.4. Statistical analyses

Final values of all speech parameter across sustained phonation and syllable repetition tasks were calculated for each participant by averaging the data obtained in two vocal task runs. The Kolmogorov–Smirnov test was used to evaluate the normality of the distribution. To assess group differences, the independent-samples *t* test was used for normally distributed data and the Mann–Whitney U test for non-normally distributed data. Pearson and Spearman correlations were applied to test for significant relationships between normally and non-normally distributed data, respectively. With respect to the explorative nature of study and the fact that each acoustic variable represents a unique speech aspect, adjustment for multiple comparisons was not performed, and the level of significance was maintained at $p < 0.05$. Effect sizes were determined by Cohen's *d*, with $d > 0.8$ indicating a large effect and $d > 0.5$ indicating a medium effect.

Estimation of the severity of affected speech dimensions was performed using Wald decision theory, which was applied to compare probability distributions between RBD and healthy subjects. As a result, each subject's performance was evaluated according to the specific speech dimension and was considered affected, normal, or intact. An affected speech dimension was ranked as 1 point, whereas an intact speech dimension was ranked as –1 point. A total score was obtained reflecting the overall degree of affected phonatory, articulatory, and prosodic speech components. Comprehensive details on Wald analysis as well as the estimation of severity of affected speech dimensions have been published previously [13].

Finally, we introduced a classification experiment to determine the best combination of acoustic features and to estimate their sensitivity and specificity in differentiating between RBD and HC groups. A support vector machine (SVM) with a Gaussian radial basis kernel was applied to search for all combinations across acoustic features. A cross-validation scheme was used to validate reproducibility of the SVM classifier, where the original data were randomly separated into a training subset composed of 75% of the data and a testing subset containing 25% of the data. This cross-validation process was repeated 20 times for each combination, and the overall classification performance of the SVM-based model was computed as the average percentage of correctly classified subjects into an appropriate group through all 20 cycles. Comprehensive details on SVM-based classification procedure have recently been published [15].

3. Results

A subgroup of seven RBD subjects showed slight motor impairment (hereafter symptomatic subgroup RBD-S with UPDRS III > 4), based on evidence that the 95% confidence interval of the UPDRS III score estimated in healthy subjects ranges from approximately 1 to 4 [4], whereas a subgroup of nine RBD participants did not show substantial motor deficits (hereafter asymptomatic subgroup RBD-AS with UPDRS III < 5). In particular, three of seven subjects from the RBD-S subgroup manifested perceptually mild speech impairment (UPDRS III 18 speech = 1), whereas no speech impairment was observed in the RBD-AS subgroup (UPDRS III 18 speech = 0).

Table 3 provides numerical data and a comparison between the RBD and control groups across all 15 speech dimensions investigated. In comparison to controls, the RBD group showed significant alterations in the speech dimensions of irregular AMR (DDK reg, $p = 0.009$) and articulatory decay (RFA, $p = 0.01$).

Fig. 1 displays the results of quantitative speech analysis. According to the individual speech profile related to the number of deviant phonatory, articulatory, and prosodic dimensions, 14 of 16 RBD subjects (88%) showed abnormalities in two or more speech dimensions, whereas only one control subject (6%) manifested two affected speech dimensions (Fig. 1A). Three or more affected speech dimensions were found in ten RBD subjects (63%) and none of the control subjects (Fig. 1A). Observed speech abnormalities were primarily related to articulatory decay, irregular pitch fluctuations, irregular and slow AMR, and signal perturbations, which were affected in more than 40% of RBD subjects (Fig. 1B). The severity of speech impairment in RBD was most related to articulatory problems, followed by phonatory deficits (Fig. 1B). Appendix S2 provides a comparison between speech disorders in RBD subjects and de novo PD patients using the same methodology designed in the current study.

Table 3
Results of acoustic speech analyses.

Acoustic variable	Group		Statistics	
	RBD	Controls	RBD vs controls	
	Mean (SD)	Mean (SD)	*p*	Effect size[a]
Phonation				
MPT (s)	18.8 (8.4)	15.5 (3.9)	0.17	0.5
F0 SD (st)	0.50 (0.24)	0.43 (0.40)	0.52	0.23
Jitter (%)	0.86 (0.53)	0.59 (0.28)	0.09	0.63
HNR (dB)	19.5 (2.6)	19.3 (3.4)	0.81	0.09
DUV (%)	1.71 (4.19)	0.07 (0.18)	0.13	0.55
Articulation				
VAI (–)	0.92 (0.06)	0.93 (0.07)	0.51	–0.24
VOT (ms)	21.6 (5.1)	20.4 (3.8)	0.48	0.25
DDK rate (syll/s)	6.56 (0.87)	7.10 (0.65)	0.07	–0.65
DDK reg (ms)	25.4 (11.8)	16.0 (6.6)	0.009	0.98
RFA (dB)	0.98 (0.16)	1.13 (0.15)	0.01	–0.95
Prosody				
Mean Int (dB)	63.7 (3.8)	63.8 (3.1)	0.96	–0.01
Int SD (dB)	7.07 (0.81)	7.03 (0.71)	0.89	0.05
F0 SD (st)	2.44 (0.82)	2.33 (0.58)	0.66	0.16
NoP (pauses/s)	3.84 (0.53)	4.02 (0.57)	0.36	–0.32
PDW (%)	3.97 (2.77)	3.67 (1.99)	0.73	0.13

Abbreviations: MPT, maximum phonation time, maximum duration of prolonged vowel; F0 SD, variability of fundamental frequency (ie, vibration rate of the vocal folds); Jitter, frequency instability; HNR, harmonics-to-noise ratio, addition of noise in speech; DUV, degree of unvoiced segments, aperiodicity of voice; VAI, vowel articulation index, centralisation of formant frequencies; VOT, voice onset time, duration of consonant articulation; DDK rate, diadochokinetic rate, rate of syllable vocalisations per second; DDK regularity, diadochokinetic regularity, ability to maintain a constant rate of consonant–vowel combinations; RFA, resonant frequency attenuation, decrease of spectral energy during articulation; Mean Int, average level of speech loudness; Int SD, variability of loudness variation; NoP, number of pauses relative to total speech time; PDW, percentage of dysfluent words, number of dysfluencies relative to number of words; SD, standard deviation.

[a] Cohen's d: effect size 0.8 is considered large, 0.5 is considered medium, and 0.2 is considered small.

Acoustic features related to aperiodicity, irregular AMR, articulatory decay, and dysfluency represent the most salient features of speech dysfunction in RBD. The combination of these four acoustics was able to discriminate RBD from HC with a sensitivity of 95.6% and specificity of 78.7%, and with an area under the curve (AUC) of 0.828. Appendix S3 lists classification results of quantitative motor speech measures.

We further observed a significant relationship between the duration of RBD symptoms and dysfluency ($r = 0.58$, $p = 0.02$). No relationships were found between speech metrics and UPDRS III score. Nevertheless, we observed significant differences or trends between RBD-S and RBD-AS in aperiodicity, slow AMR, reduced loudness, and overall severity of speech impairment (Fig. 2).

4. Discussion

The current study revealed the presence of speech impairment in individuals with RBD. Since RBD is currently considered as a prodromal stage of PD and related neurodegenerative disorders, we may assume that these observed speech abnormalities also represent

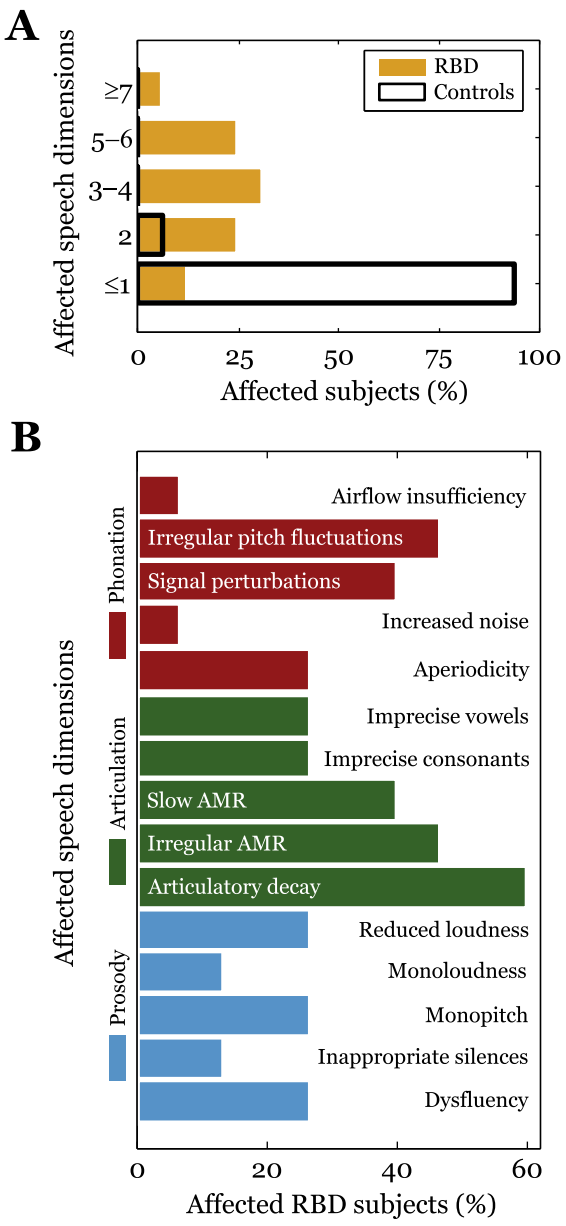


Fig. 1. Characteristics of speech. (A) Number of affected speech dimensions across participants. (B) Percentage of affected RBD subjects according to the specific speech dimension. RBD = rapid eye movement sleep behaviour disorder; AMR = alternating motion rates.

prodromal markers of neurodegeneration. Based on our analysis, 88% of RBD subjects manifested at least two affected speech dimensions, whereas 63% of RBD individuals showed a greater extent of speech impairment than any of the control speakers. Although RBD subjects demonstrated various combinations of phonation, articulation, and prosodic deficits, articulatory impairment was the most prominent. The combination of four acoustic measures representing aperiodicity, irregular AMR, articulatory decay, and dysfluency led to 96% sensitivity and 79% specificity in discriminating between RBD and healthy control subjects. This classification score was achieved using a newly designed measure of articulatory decay, resulting in a significant performance increase.

In general, our subjects with RBD manifested various combinations of phonatory, articulatory, and prosodic speech deficits. Similar vocal abnormalities have also been observed in newly diagnosed PD patients before the introduction of pharmacotherapy [13]. However, the predominant patterns of speech deficits in de novo PD were characterised by prosodic–articulatory abnormalities primarily represented by monopitch and imprecise consonants (Appendix S2) [13,15]. Conversely, patients with RBD primarily manifested articulatory–phonatory speech involvement and less frequently prosodic deviations. We may therefore hypothesise that RBD patients may present a distinct phenotype of PD-related speech disorder due to a differing underlying pattern of neurodegeneration in PD with RBD, in comparison to PD without RBD [24,25].

Previous research has suggested that PD with RBD presents more akinetic-rigid disease, with more autonomic dysfunction and more cognitive impairment, and with a higher risk of dementia [26–29]. Indeed, certain relationships between bradykinesia or rigidity and dysphonia patterns have been previously reported in PD [14], supporting the increased phonatory deviations observed in our RBD subjects. Another explanation of different speech patterns in RBD may be a high conversion rate to DLB [5]. Little is known about dysarthria in DLB; perceptual investigation has revealed only hypokinetic and monotonic patterns of dysarthria, which appear to be similar to PD [10]. Nevertheless, dysarthria typically occurs earlier and tends to be more severe in atypical parkinsonian syndromes than in PD [10,11]. Therefore, one might expect that speech patterns would differ between initial stages of DLB and PD.

Speech abnormalities observed in at least 25% of our RBD cases such as tendency towards decay of articulatory movements (mumbling), signal perturbations, imprecise vowels and consonants, reduced loudness, monopitch, and dysfluency represent common patterns of hypokinetic dysarthria in PD, as confirmed by a number of previous perceptual and acoustic studies [12,30]. It has been suggested that decreased dopamine levels in the brain may lead to the development of stuttering-like behaviour in PD [23,31]. This

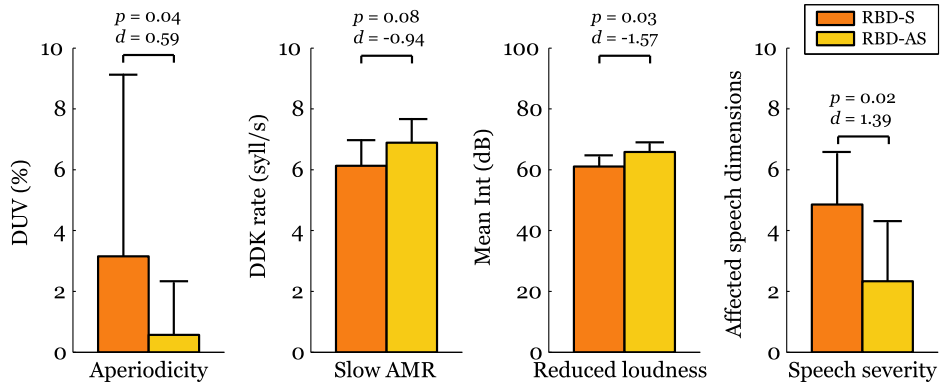


Fig. 2. Differences between RBD subgroups. RBD = rapid eye movement sleep behaviour disorder; RBD-S = symptomatic RBD subgroup; RBD-AS = asymptomatic RBD subgroup; DUV = degree of unvoiced segments; DDK rate = diadochokinetic rate; Mean Int = average level of loudness; AMR = alternating motion rates.

hypothesis may be further supported by the relationship that we observed between dysfluency severity and RBD symptom duration, indicating a greater reduction of dopamine levels, possibly due to a longer duration of prodromal neurodegeneration.

The poor vocal control in our RBD subjects was mirrored by aperiodicity, irregular pitch fluctuations, and abnormal AMR, which are not specific patterns of hypokinetic dysarthria in PD, and seem to be more distinctive for atypical parkinsonian syndromes [11]. However, aperiodicity and pitch fluctuations represent rather nonspecific markers of neuronal dysfunction, probably due to uncontrolled movement of laryngeal muscles or fluctuation of vocal fold tension. Increased variability of regular syllable repetitions can be considered as a specific feature mainly in patients with cerebellar ataxia, including MSA [11,12].

In a previous report, a UPDRS motor score cut-off value of >4 was a very good indicator of parkinsonism two years before onset [6]. Our RBD-S subgroup showed substantially more pronounced speech impairment than the RBD-AS subgroup, indicating that the severity of speech disorders to a certain extent parallels motor disability due to an underlying degenerative process. Especially, aperiodicity and slow AMR reflected increased motor disability in the RBD-S subgroup. Interestingly, slow AMR and aperiodicity were also observed in untreated PD patients (Appendix S2) [13] but were not present in PD speakers with short disease duration on stable dopaminergic medication [11].

One potential advantage of the current approach is that acoustic measures and computational analyses of the speech abnormalities are noninvasive, cost-effective, valid, precise, and reliable methods to detect, characterise, and monitor the progression of the disease [32]. In particular, increasing computational power has enabled a higher level of automation in speech assessment. In the present study, investigation of the majority of vocal patterns was based on a fully automatic process with a minimum user control of the analysis procedure. However, analysis of certain deviant speech dimensions such as imprecise vowel articulation and disfluency involves hand-labeling and auditory perceptual evaluation, which are time consuming and require an experienced investigator. Therefore, future research on elaboration and extension of currently available technologies for assessment of various dysarthria patterns is warranted.

There are some limitations to the present study. The findings of our pilot analyses are based on a small number of patients. Our results should therefore be confirmed in a larger population sample, including follow-up evaluation to verify diagnosis of PD or related neurodegenerative disorders. We acknowledge that we did not perform specific testing for neuropsychological and other nonmotor involvement, as the primary aim was the investigation of motor speech deviations. Furthermore, nine of our RBD subjects were treated by clonazepam, which may influence speech, coordination, and cognition. Nevertheless, speech abnormalities were observed both in RBD individuals who were and in those who were not on clonazepam. In fact, all but one RBD subject received low clonazepam doses, and the testing was done in the afternoon, when the effect of a dose of clonazepam given the previous evening should no longer influence performance.

5. Conclusion

Our findings reveal significant patterns of motor speech dysfunction in idiopathic RBD subjects with sensitivity and specificity that are comparable to results based on olfactory dysfunction [33] and superior to those found in previous studies investigating preclinical motor markers [6]. Therefore, we believe that speech disorder may be a promising marker indicating prodromal neurodegeneration in RBD, and the results of this pilot study are worth following up in future larger studies. As complex speech assessment represents a quantitative and easily performed motor test, vocal analysis appears to be a promising screening test, particularly in view of the anticipated advent of neuroprotective treatment.

Conflict of interest

The authors report no conflicts of interest.

The ICMJE Uniform Disclosure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: http://dx.doi.org/10.1016/j.sleep.2015.07.030.

Acknowledgements

This study was supported by the Ministry of Health of the Czech Republic (MZ ČR 15-28038A).

Appendix: Supplementary material

Supplementary data to this article can be found online at doi:10.1016/j.sleep.2015.07.030.

References

[1] Boeve B. REM sleep behaviour disorder. Ann N Y Acad Sci 2010;1184:15–54.
[2] Schenck CH, Bundlie SR, Mahowald MW. Delayed emergence of a parkinsonian disorder in 38% of older men initially diagnosed with idiopathic rapid eye movement sleep disorder. Neurology 1996;46:388–93.
[3] Iranzo A, Molinuevo JL, Santamaria J, et al. Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study. Lancet Neurol 2006;5:572–7.
[4] Postuma RB, Gagnon JF, Vendette M, et al. Markers of neurodegeneration in idiopathic REM sleep behaviour disorder and Parkinson disease. Brain 2009;132:2298–307.
[5] Iranzo A, Fernandez-Arcos A, Tolosa E, et al. Neurodegenerative disorder risk in idiopathic REM sleep behaviour disorder: study in 174 patients. PLoS ONE 2014;9:e89741.
[6] Postuma RB, Lang AE, Gagnon JF, et al. How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder. Brain 2012;135:1860–70.
[7] Schenck CH, Montplaisir JY, Frauscher B, et al. Rapid eye movement sleep behaviour disorder: devising controlled active treatment studies for symptomatic and neuroprotective therapy – a consensus statement from the International Rapid Eye Movement Sleep Behaviour Disorder Study Group. Sleep Med 2013;14:795–806.
[8] Kluin KJ, Gilman S, Lohman M, et al. Characteristics of the dysarthria in multiple system atrophy. Arch Neurol 1996;53:545–8.
[9] Ho AK, Iansek R, Marigliani C, et al. Speech impairment in large sample of patients with Parkinson's disease. Behav Neurol 1998;11:131–7.
[10] Muller J, Wenning GK, Verny M, et al. Progression of dysarthria and dysphagia in postmortem-confirmed parkinsonian disorders. Arch Neurol 2001;58:259–64.
[11] Rusz J, Bonnet C, Klempir J, et al. Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. J Neurol 2015;262:992–1001.
[12] Duffy JR. Motor speech disorders: substrates, differential diagnosis and management. 3rd ed. St Louis: Mosby; 2013.
[13] Rusz J, Cmejla R, Ruzickova H, et al. Quantitative acoustic assessment of speech and voice disorders in early untreated Parkinson's disease. J Acoust Soc Am 2011;129:350–67.
[14] Rusz J, Cmejla R, Ruzickova H, et al. Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test. Mov Disord 2011;26:1951–2.
[15] Novotny M, Rusz J, Cmejla R, et al. Automatic evaluation of articulatory disorders in Parkinson's disease. IEEE Trans Audio Speech Lang Process 2014;22:1366–78.
[16] Tetrud JW. Preclinical Parkinson's disease: detection of motor and nonmotor manifestations. Neurology 1991;41(Suppl. 2):69–71.
[17] Harel BT, Cannizzaro MS, Cohen H, et al. Acoustic characteristics of Parkinsonian speech: a potential biomarker of early disease progression and treatment. J Neurolinguistics 2004;17:439–53.
[18] Grant LM, Richter F, Miller JE, et al. Vocalization deficits in mice overexpressing alpha-synuclein, a model of pre-manifest Parkinson's disease. Behav Neurosci 2014;128:110–21.
[19] American Academy of Sleep Medicine. The international classification of sleep disorders: diagnostic and coding manual. 2nd ed. Westchester, IL: American Academy of Sleep Medicine; 2005.
[20] Iber C, Ancoli-Israel S, Chesson A, et al. for the American Academy of Sleep Medicine. The AASM manual for the scoring of sleep and associated events:

rules, terminology and technical specifications. 1st ed. Westchester, IL: American Academy of Sleep Medicine; 2007.
[21] Boersma P, Weenink D. PRAAT, a system for doing phonetics by computer. Glot Int 2001;5:341–5.
[22] Rusz J, Cmejla R, Tykalova T, et al. Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task. J Acoust Soc Am 2013;134:2171–81.
[23] Tykalova T, Rusz J, Cmejla R, et al. Effect of dopaminergic medication on speech dysfluency in Parkinson's disease: a longitudinal study. J Neural Transm 2015;122:1135–42.
[24] Postuma RB, Gagnon JF, Vendette M, et al. REM sleep behaviour disorder in Parkinson's disease is associated with specific motor features. J Neurol Neurosurg Psychiatry 2008;79:1117–21.
[25] Postuma RB, Bertrand JA, Montplaisir J, et al. Rapid eye movement sleep disorder and risk of dementia in Parkinson's disease: a prospective study. Mov Disord 2012;27:720–6.
[26] Romenets SR, Gagnon JF, Latreille V, et al. Rapid eye movement sleep behaviour disorder and subtypes of Parkinson's disease. Mov Disord 2012;27:996–1003.
[27] Anang JB, Gagnon JF, Bertrand JA, et al. Predictors of dementia in Parkinson disease: a prospective cohort study. Neurology 2014;83:1253–60.
[28] Gong Y, Xiong KP, Mao CJ, et al. Clinical manifestations of Parkinson disease and the onset of rapid eye movement sleep behaviour disorder. Sleep Med 2014;15:647–53.
[29] Neikrug AB, Avanzino JA, Liu L, et al. Parkinson's disease and REM sleep behaviour disorder result in increased non-motor symptoms. Sleep Med 2014;15:959–66.
[30] 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;57:1330–43.
[31] Civier O, Bullock D, Max L, et al. Computational modelling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation. Brain Lang 2013;169:263–78.
[32] Skodda S. Analysis of voice and speech performance in Parkinson's disease: a promising tool for the monitoring of disease progression and differential diagnosis. Neurodegener Dis Manag 2012;2:535–45.
[33] Miyamoto T, Miyamoto M, Iwanami M, et al. Olfactory dysfunction in idiopathic REM sleep behaviour disorder. Sleep Med 2010;11:458–61.

Appendix A8

Distinct patterns of consonant articulation among Parkinson's disease, progressive supranuclear palsy and multiple system atrophy

Tykalova T, Rusz J, Klempir J, Cmejla R, Ruzicka E (2017) Distinct patterns of consonant articulation among Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Brain Lang* **165**:1-9.

Journal name (category): **Brain & Language** (linguistics/psychology, experimental/audiology & speech-language pathology/neurosciences)
Quartile in category: **Q1/Q2/Q1/Q3**
Impact factor: **2.493**
Number of WoS citations (excluding autocitations): **3**

Brain & Language 165 (2017) 1–9

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Distinct patterns of imprecise consonant articulation among Parkinson's disease, progressive supranuclear palsy and multiple system atrophy

Tereza Tykalova [a], Jan Rusz [a,b,*], Jiri Klempir [b,c], Roman Cmejla [a], Evzen Ruzicka [b]

[a] Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 166 27 Prague 6, Czech Republic
[b] Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University in Prague, Katerinska 30, 120 00 Prague 2, Czech Republic
[c] Institute of Anatomy, First Faculty of Medicine, Charles University in Prague, U nemocnice 3, 128 00 Prague 2, Czech Republic

ARTICLE INFO

Article history:
Received 14 July 2016
Revised 15 November 2016
Accepted 18 November 2016

Keywords:
Parkinson's disease
Atypical Parkinsonian syndromes
Voice onset time
Dysarthria
Speech disorders
Acoustic analysis
Perceptual assessment

ABSTRACT

Distinct speech characteristics that may aid in differentiation between Parkinson's disease (PD), progressive supranuclear palsy (PSP) and multiple system atrophy (MSA) remain tremendously under-explored. Here, the patterns and degree of consonant articulation deficits across voiced and voiceless stop plosives in 16 PD, 16 PSP, 16 MSA and 16 healthy control speakers were evaluated using acoustic and perceptual methods. Imprecise consonant articulation was observed across all Parkinsonian groups. Voice onset time of voiceless plosives was more prolonged in both PSP and MSA compared to PD, presumably due to greater severity of dysarthria and slower articulation rate. Voice onset time of voiced plosives was significantly shorter only in MSA, likely as a consequence of damage to cerebellar structures. In agreement with the reduction of pre-voicing, MSA manifested increased number of voiced plosives misclassified as voiceless at perceptual evaluation. Timing of articulatory movements may provide important clues about the pathophysiology of underlying disease.

1. Introduction

Idiopathic Parkinson's disease (PD) is a common neurological disorder characterized by the progressive loss of dopaminergic neurons in the substantia nigra pars compacta, affecting 1.6% of persons over the age of 65 years (deRijk et al., 1997). Dopamine concentrations have been shown to be significantly reduced before distinct motor deficits become apparent (Hornykiewicz, 1998). The cardinal signs of PD, often referred to as Parkinsonism, include resting tremor, bradykinesia, muscular rigidity and postural instability. Other neurodegenerative diseases that go beyond the signs and symptoms of Parkinsonism are known as atypical Parkinsonian syndromes (APS). Progressive supranuclear palsy (PSP) and multiple system atrophy (MSA) are the most common APS, with an estimated prevalence of 30–40 per 100,000 among persons older than 65 years (Schrag, Ben-Shlomo, & Quinn, 1999). Characteristic clinical features of PSP include supranuclear gaze palsy, frequent falls, bradykinesia, axial rigidity, cognitive decline and communication disorders (Nath, Ben-Shlomo, Thomson, Lees, & Burn, 2003; Steele, Olszewski, & Richardson, 1964), reflecting widespread neurodegeneration involving the midbrain as well as the globus pallidus, striatum, hypothalamic nucleus, pons, superior cerebellar peduncle and cerebellar dentate nucleus (Nath et al., 2003). Conversely, MSA manifests by various combinations of autonomic, cerebellar and Parkinsonian features (Wenning, Colosimo, Geser, & Poewe, 2004), corresponding to degeneration of the cerebellum, middle cerebellar peduncle, striatum, substantia nigra, inferior olivary nucleus and pons (Gilman et al., 2008). APS differ from PD by poor levodopa response and more rapid disease progression resulting in shorter life expectancy (O'Sullivan et al., 2008; Wenning, Litvan, & Tolosa, 2011). Furthermore, the underlying pathophysiology differs as PD and MSA are α-synucleinopathies while PSP is a tauopathy. However, the differentiation between PD and both PSP and MSA can be challenging as the initial signs are frequently nonspecific and overlap those of PD (Osaki et al., 2004; Schrag et al., 1999).

1.1. Speech impairment in PD, PSP and MSA

Dysarthria is a well-recognized clinical manifestation of Parkinsonian disorders, developing in 90–100% of patients with PD, PSP and MSA during the course of the disease (Ho, Iansek, Marigliani, Bradshaw, & Gates, 1998; Kluin, Foster, Berent, & Gilman, 1993;

* Corresponding author at: Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 166 27 Praha 6, Czech Republic and Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University in Prague, Katerinska 30, 120 00 Prague 2, Czech Republic.
E-mail address: rusz.mz@gmail.com (J. Rusz).

http://dx.doi.org/10.1016/j.bandl.2016.11.005
0093-934X/

Kluin, Gilman, Lohman, & Junck, 1996; Muller et al., 2001; Rusz et al., 2015). Speech impairment is an early and prominent manifestation that can contribute primarily to the diagnosis of PSP (Goetz, Leurgans, Lang, & Litvan, 2003; Kim & McCann, 2015; Wenning et al., 2011), but has also been largely documented in the early stages of PD and MSA (Huh et al., 2015; Kim, Kent, Kent, & Duffy, 2010; Rusz, Cmejla, Ruzickova, & Ruzicka, 2011).

Due to dysfunction of the basal ganglia, the majority of PD patients manifest hypokinetic dysarthria characterized by monopitch, monoloudness, reduced stress, variable rate, imprecise articulation, harsh voice quality, speech dysfluencies and inappropriate silence (Darley, Aronson, & Brown, 1969b; Ho et al., 1998). Conversely, PSP and MSA patients typically evolve mixed dysarthria with a combination of hypokinesia, ataxia and spasticity as a result of more widespread neuronal atrophy (Kluin et al., 1993, 1996; Rusz et al., 2015). Indeed, previous studies (Kluin et al., 1993, 1996) investigating 46 MSA and 44 PSP patients using oral motor and perceptual speech analysis have reported mixed dysarthria with combinations of all hypokinetic, spastic and ataxic components in two-thirds of APS patients. Hypokinetic components followed by ataxic components were predominant in MSA patients, while spastic components were mostly present in PSP patients (Kluin et al., 1993, 1996).

Considering individual speech aspects, only the occurrence of stuttering-like behaviour has been reported to be distinctive for PSP as compared to MSA (Kluin et al., 1993, 1996; Rusz et al., 2015). A small number of studies have also focused on an objective description of the dysarthria profile in APS in comparison to PD (Huh et al., 2015; Kim et al., 2010; Rusz et al., 2015; Sachin et al., 2008; Skodda, Visser, & Schlegel, 2011). In general, these studies have shown that the impairment of specific speech dimensions is more pronounced in APS than in PD (Huh et al., 2015; Rusz et al., 2015; Skodda et al., 2011). Dysarthria in PSP has been reported to be characterized by stuttering-like behaviour, reduced speech rate, decreased intonation variability, prolonged pauses, articulation imprecision and poor quality of voice (Rusz et al., 2015; Skodda et al., 2011), whereas MSA patients have been said to manifest with excess pitch fluctuations, excess intensity variations, increased voice pitch, reduced speech rate, prolonged phonemes, vocal tremor, voice perturbations and slow variable alternating motion rates (Huh et al., 2015; Kim et al., 2010; Rusz et al., 2015; Saxena, Behari, Kumaran, Goyal, & Narang, 2014). However, little effort has been made to investigate consonant articulation in APS.

1.2. Consonant articulation in PD, PSP and MSA

The description of disturbed consonant articulation in various diseases has typically been based on perceptual assessment in subgroups of patients defined by dysarthria subtype such as spastic, ataxic or hypokinetic, rather than by disease aetiology (i.e., PD, PSP or MSA; (Chakraborty, Roy, Hazra, Biswas, & Bhattacharya, 2008; Darley, Aronson, & Brown, 1969a; Hartelius, Gustavsson, Astrand, & Holmberg, 2006; Logemann & Fisher, 1981; Weismer, 1984). Furthermore, previous studies were limited primarily to documenting the occurrence of articulation deficits and did not describe specific features characterizing imprecise consonants (Chakraborty et al., 2008; Darley et al., 1969a; Hartelius et al., 2006). In particular, in the classic study by Darley et al. (1969b), imprecise consonant articulation was perceptually found to be one of the most deviant speech dimensions in PD. The presence of imprecise consonant articulation has also been perceptually revealed in a cohort of MSA and PSP patients (Hartelius et al., 2006). Interestingly, although in general speech deviation of greater severity was found in PSP, consonant articulation was more severely affected in MSA (Hartelius et al., 2006).

With regard to acoustic analyses, several measurements can be used for description of consonants including various measures of duration, formant transitions, spectral moments or energy-based measures (Kent & Read, 1992). Among them, voice onset time (VOT) determined for stop plosives is perhaps the most frequently used parameter and a relatively large amount of data has been published on VOT in PD patients. Unfortunately, previous studies have provided rather contradictory findings. While some researchers have reported increased VOT duration (Forrest, Weismer, & Turner, 1989; Novotny, Rusz, Cmejla, & Ruzicka, 2014), others have observed unchanged (Fischer & Goberman, 2010; Ravizza, 2003) or even decreased VOT (Flint, Black, Campbelltaylor, Gailey, & Levinton, 1992) in PD subjects. It has been suggested that these discrepancies may be due to the fact that the measurement of VOT is dependent on speaking rate (Volaitis & Miller, 1992); however, VOT ratio, a rate-independent variation of VOT, did not clarify these ambiguous findings (Fischer & Goberman, 2010; Novotny et al., 2014).

Only one previous study has focused on the acoustic investigation of consonant characteristics for five categories of plosives in PD, PSP, and MSA in comparison to controls (Saxena et al., 2014). However, this study provided rather inconsistent findings across various consonant categories and speaker groups (Saxena et al., 2014). In particular, the authors revealed no significant alterations of VOT duration in dentals across all groups, but observed increased VOT duration of velars in PD, palatals in PSP, bilabials in MSA, PSP and PD, and of retroflexes in PSP and MSA (Saxena et al., 2014). However, a direct comparison of consonant articulation between PD, PSP and MSA has never been performed.

1.3. Aim of the present study

The aim of the current study was therefore to investigate the patterns and degree of consonant articulation deficits across different voiceless and voiced stop plosives in PD, PSP, MSA and healthy speakers using objective acoustic measures to help elucidate distinct speech characteristics that could aid in the differentiation between various forms of Parkinsonism. In addition, perceptual examination of phonetic contrast between voiceless and voiced plosives was performed to determine if consonant imprecision was a notable feature of dysarthria in PD, PSP and MSA. Additionally, the relationships between speech performances and clinical manifestations were explored to provide greater insight into the pathophysiology of speech production in PD, PSP and MSA.

2. Methods

2.1. Participants

From 2011 to 2015, a total of 48 consecutive patients including 16 fulfilling the diagnostic criteria for idiopathic PD (5 men and 11 women), 16 with a diagnosis of probable PSP (11 men and 5 women) and 16 with a diagnosis of probable MSA (5 men and 11 women) were recruited. Among APS, hereafter hypernym for the MSA and PSP subgroups, 13 PSP patients were diagnosed with PSP-Richardson syndrome, 2 with PSP-Parkinsonism and 1 with PSP-pure akinesia with gait freezing, whereas MSA patients were diagnosed with the MSA-Parkinsonian subtype in 14 cases and the MSA-cerebellar subtype in 2 cases. The clinical diagnoses of all patients were established by a specialist in movement disorders (JK) according to the UK Parkinson's Disease Society Bank Criteria for PD (Hughes, Daniel, Kilford, & Lees, 1992), the NINDS-PSP clinical diagnostic criteria for PSP (Litvan et al., 1996) or the consensus diagnostic criteria for MSA (Gilman et al., 2008). At the time of the examination, all patients treated pharmacologically were on stable

medication for at least 4 weeks, consisting of various doses of levodopa alone or in combination with different dopamine agonists and/or amantadine. None of the patients received antipsychotic drugs. Disease duration was estimated based on the self-reported occurrence of first motor symptoms. APS patients were further scored according to the natural history and neuroprotection in Parkinson plus syndromes–Parkinson plus scale (NNIPPS; range from 0 to 346, where a higher score indicates more severe disability (Payan et al., 2011), while PD patients were rated according to the Unified Parkinson's Disease Rating Scale motor subscore (UPDRS III; range from 0 to 108, where a higher score indicates more severe motor disability). Item 18 of the UPDRS III was used for the perceptual description of overall dysarthria severity. No patients reported a history of speech, language or hearing disorders unrelated to their Parkinsonism symptoms. No statistically significant differences between the PSP and MSA groups were found for symptom duration, medication dose, motor severity or speech severity (Mann-Whitney U test: $p = 0.14–0.98$). Patient demographic characteristics are summarized in Table 1.

The majority of investigated PD, PSP, and MSA subjects had also participated in a former study focused on the detailed assessment of severity and patterns of dysarthria (Rusz et al., 2015). In general, dysarthria was uniformly present in all Parkinsonian patients and ranged from mild to severe. The PD group manifested mild hypokinetic dysarthria with dominant monopitch, imprecise vowels and inappropriate silences. The PSP group showed mild to moderate hypokinetic-spastic dysarthria with dominant dysfluency, slow speech rate, inappropriate silences, imprecise vowel articulation, monopitch and harsh voice quality. The MSA group presented mild to moderate ataxic-hypokinetic dysarthria dominated by excessive pitch and loudness variations, prolonged phonemes, vocal tremor and strained-strangled voice quality. However, characteristics related to imprecise consonant articulation were not previously reported.

The healthy control (HC) group consisted of 16 subjects (5 men and 11 women) of comparable age (mean age 62.8, SD 7.3, range 53–74 years). No significant differences in age distribution were detected between the HC, PD, PSP and MSA groups (analysis of variance: $p = 0.30$). Healthy individuals reported no history of neurological disease or any disorder that may affect speech, language or hearing.

All participants were Czech native speakers and were able to fully cooperate during the procedure. No subjects had signs of major depression or cognitive deficits that could interfere with the measurements. The study was approved by the Ethics Committee of the General University Hospital in Prague, Czech Republic, and all participants provided written, informed consent for the neurological examination and recording procedure.

2.2. Speech material

A series of tokens designed as "CVtka" were used for the assessment of consonant articulation, where C represented a consonant and V corresponded to a corner vowel. Specifically, six stop plosives were covered including /p/, /t/, /k/, /b/, /d/ and /g/. The vowels consisted of /a/, /ɪ/ and /u/. These 3 vowels were chosen as they are a representative sample with respect to vowel height, as vowel height has been reported to have an effect on some consonant characteristics such as VOT (Fischer & Goberman, 2010). From a phonetic point of view, /b/, /d/ and /g/ in the CV context are usually pronounced as prevoiced in Czech (i.e., voiced during closure), while /p/, /t/ and /k/ are pronounced as voiceless and unaspirated (Simackova, Podlipsky, & Chladkova, 2012). Furthermore, the stress is always on the first syllable in two-syllable words used without prepositions. The suffix /tka/ was added to evoke more naturally sounding tokens. Indeed, from 18 created stimuli, 7 were existing Czech words whereas the remaining 11 were meaningless. No analyses were performed on the suffix.

2.3. Recording procedure

The audio data were recorded in a quiet room with a low level of ambient noise using a head-mounted condenser microphone (Bayerdynamic Opus 55, Heilbronn, Germany) placed approximately 5 cm from the subject's lips. The speech signals were sampled at 48 kHz with 16-bit resolution. The recordings were collected during one session with a speech specialist who conveyed instructions to the subjects. Each participant had to complete a series of speaking tasks as part of a larger protocol lasting approximately 20 min. There were no time limits during the recordings. All participants were asked to repeat their performance at any time if they or the examiner were not fully satisfied with their initial attempt. The performance of the task focused on consonant articulation was selected for further investigation. During the task, the participants were instructed to read the words presented by the examiner on paper cards. The subjects were further warned not to be surprised as some of the words would be meaningless and to simply read what they saw. As Czech is a language

Table 1
Patient demographics.

	PD (n = 16; 5 men, 11 women) Mean/SD (range)	PSP (n = 16; 11 men, 5 women) Mean/SD (range)	MSA (n = 16; 5 men, 11 women) Mean/SD (range)
Age (years)	63.3/7.3 (49–74)	66.1/5.1 (54–72)	61.9/5.9 (52–72)
Symptom duration (years)	6.3/4.1 (2–14)	4.3/2.4 (1–11)	3.8/1.3 (2–6)
Levodopa equivalent (mg/day)	781/471 (0–1680)	447/538 (0–1500)	567/573 (0–1700)
Amantadine (mg/day)	100/136 (0–300)	169/178 (0–400)	131/145 (0–400)
NNIPPS total	–	71.9/28.5 (19–132)	71.8/22.5 (43–123)
UPDRS III total	17.3/9.6 (7–38)	–	–
UPDRS III speech item	0.7/0.8 (0–2)	2.1/0.7 (0–3)	1.7/0.7 (0–3)
Tremor subscore	2.1/1.4 (0–5)[a]	2.2/2.6 (0–7)[b]	1.4/2.4 (0–9)[b]
Rigidity subscore	3.1/1.9 (1–7)[a]	3.6/3.1 (0–11)[b]	3.9/3.3 (0–11)[b]
Bradykinesia subscore	7.8/5.0 (2–18)[a]	22.1/10.9 (4–45)[b]	25.1/8.3 (12–39)[b]
Bulbar/pseudobulbar subscore	–	9.2/4.2 (3–18)[b]	7.5/2.2 (4–12)[b]
Cerebellar subscore	–	0.0/0.0 (0–0)[b]	4.1/6.4 (0–22)[b,c]

PD, Parkinson's disease; PSP, progressive supranuclear palsy; MSA, multiple system atrophy; SD, standard deviation; NNIPPS, natural history and neuroprotection on Parkinson plus syndromes-Parkinson plus scale; UPDRS, unified Parkinson disease rating scale.

[a] UPDRS III subscore.

[b] NNIPPS subscore.

[c] 7 of 16 MSA patients manifested non-zero cerebellar subscore.

with fixed unambiguous pronunciation rules, no training of reading was performed. Tokens were printed in a large font on separate cards to provide optimal breathing patterns during reading and to minimize the effect of individual speech rate. Each of the 18 cards included one stimulus. The cards were presented at a stable pace, approximately one card per two seconds in quasi-randomized order. The entire task was repeated twice. As a result, a total of 36 tokens were obtained from each participant.

2.4. Acoustic analysis

Audio samples were analyzed using specialized speech software PRAAT® (Boersma, 2002). As previous research has shown that altered VOT is among the most powerful indicators of speech disorder in PD (Novotny et al., 2014) and subjects with cerebellar atrophy present with a reduced categorical separation of the VOT of voiceless and voiced stop consonants (Ackermann, Graber, Hertrich, & Daum, 1999; Ackermann & Hertrich, 1997), three acoustic variables including VOT, VOT ratio and vowel duration were investigated. These acoustic variables were measured from the first syllable of each token by hand using both wide-band spectrogram and oscillographic sound pressure signal displayed on screen. VOT was determined as the interval between the articulatory release of stop and the onset of vocal fold vibration (Kent & Read, 1992). Both positive and negative values of VOT were allowed. The negative values of VOT refer to voicing lead or prevoicing (voicing begins before the stop release) characteristic for voiced plosives (Kent & Read, 1992). If multiple bursts occurred, the initial burst was used to measure VOT (Fischer & Goberman, 2010). The VOT ratio was defined as VOT divided by the duration of whole syllable (Fischer & Goberman, 2010; Novotny et al., 2014).

All data were analyzed by one investigator (TT) as the selection of CV boundaries is a simple and well-defined task, as also documented by previous researchers reporting high intra- and inter-rater reliability (Fischer & Goberman, 2010; Flint et al., 1992). The assigned values for each participant obtained from two vocal task runs were averaged. Subsequently, the values for each stop plosive were averaged across all three corner vowels. For statistical comparisons, the acoustic parameters were further assessed for a subset of voiceless (defined as the average of /p/, /t/ and /k/) and voiced (defined as the average of /b/, /d/ and /g/) consonants separately. No statistically significant differences between the speech performances of male and female participants across any acoustic variable or speakers group were found (Mann-Whitney *U* test: $p = 0.12–0.86$).

2.5. Non-measurable data

The acoustic variables were not obtained from a small proportion of data due to methodological constraints. In particular, the amount of omitted tokens was 6.1% in MSA, 3.3% in PSP and less than 1% in PD and HC speakers. The main reason for classification of stimuli as non-measurable was the absence of burst in the PSP group (52% of omitted tokens) and the absence of voicing, i.e., missing vowel or vowel lasting less than 4 pitch periods in the MSA group (29% of omitted tokens). As the missing data were likely to be distributed randomly, at least 3 tokens for each stop consonant were always available for further analysis.

2.6. Perceptual assessment

A total of 2240 recorded tokens from all four groups of participants were used for perceptual analyses of phonetic contrast between voiceless and voiced plosives. The remaining 64 (2.8%) tokens were discarded as they were not used in acoustic analyses due to methodological constraints such as incomplete stop consonant production, devoicing of vowels or the presence of dysfluencies. All tokens were randomized using a computer algorithm and subsequently separated to sequences containing a maximum of 100 tokens; each lasting approximately 10 min. An interval of at least 5 s duration with no speech production was included between two contiguous tokens to ensure enough time for listeners to note the perceived item. Three speech specialists (TT, JR, RC) performed the perceptual assessment using the same standard over-the-ear headphones (ATH-T500, Audio-technica) adjusted to comfortable volume. The listeners were instructed to replay each sequence of recordings and note the initial consonant of each token. At least 10 min break was required between two sequences. As a measure of perceived voicing contrast the percentage of /p/, /t/ or /k/ misassigned as /b/, /d/ or /g/ (Err/voiceless/) and vice versa (Err/voiced/) was determined. The final score was calculated by averaging perceptual ratings obtained across three raters. Re-analysis of 20% of all data was performed by the same investigator (TT) that performed the original set of measures. Based upon a two-way random single measures intra-class correlation, the estimated inter-rater reliability across three raters was 0.95 ($p < 0.001$) whereas intra-rater reliability across the same rater was 0.97 ($p < 0.001$).

2.7. Statistical analysis

All acoustic variables were normally distributed whereas perceptual metrics were not found to be normally distributed (Kolmogorov-Smirnov test). Group differences were calculated using analysis of variance for acoustic parameters and the Kruskal-Wallis test for perceptual measures. Post-hoc significance was assessed by the Fisher least-squares difference. Pearson and Spearman correlations were applied to test for significant relationships between normally and non-normally distributed data, respectively. Due to the exploratory nature of the study, adjustment for multiple comparisons was not performed and the level of significance was set to $p < 0.05$.

3. Results

Table 2 provides acoustic data for consonant articulation across all groups through six plosives including /p/, /b/, /t/, /d/, /k/ and /g/. Fig. 1 shows a representative example of stop consonant duration for the voiceless plosive /t/ and voiced plosive /d/ depicted for HC, PD, PSP and MSA subjects.

Fig. 2A depicts comparison of articulation performances among PD, PSP, MSA and HC subjects for subsets of voiceless plosives. VOT was found to be the best parameter for differentiating between groups [$F(3,60) = 16.7$, $p < 0.001$, $\eta^2 = 0.45$]. Post hoc comparisons revealed significantly longer VOT in both PSP and MSA compared to PD or HC individuals (both $p < 0.001$). VOT ratio significantly discriminated between speaker groups [$F(3,60) = 5.8$, $p = 0.002$, $\eta^2 = 0.22$], as the HC group manifested significantly smaller VOT ratio than all patient groups including PD ($p < 0.05$), PSP ($p < 0.01$) and MSA ($p < 0.001$). The vowel duration slightly varied among groups [$F(3,60) = 2.9$, $p = 0.04$, $\eta^2 = 0.13$] due to the PSP group, which showed significantly longer vowel length compared to both HC ($p < 0.05$) and PD ($p < 0.01$).

Fig. 2B illustrates comparison of articulation performances among PD, PSP, MSA and HC subjects for subsets of voiced plosives. Group differences were found for both VOT [$F(3,60) = 9.3$, $p < 0.001$, $\eta^2 = 0.32$] and VOT ratio [$F(3,60) = 14.5$, $p < 0.001$, $\eta^2 = 0.42$]. Post hoc comparisons revealed significantly shorter negative VOT in MSA as compared to all groups including HC ($p < 0.001$), PD ($p < 0.01$) and PSP ($p < 0.001$). Accordingly, a smaller negative VOT ratio was found in MSA as compared to HC, PD and

Table 2
Results of acoustic speech analyses for six plosives including /p/, /b/, /t/, /d/, /k/ and /g/. Bold numbers indicate patient values that significantly differ from the HC group ($p < 0.05$).

	Bilabial plosives		Alveolar plosives		Velar plosives	
	/p/ Mean/SD	/b/ Mean/SD	/t/ Mean/SD	/d/ Mean/SD	/k/ Mean/SD	/g/ Mean/SD
VOT (ms)						
HC	14/4	−100/37	20/5	−102/42	29/10	−109/32
PD	18/6	−**69**/23	24/10	−71/30	31/10	−**76**/30
PSP	**28**/8	−90/54	**40**/16	−94/53	**51**/19	−83/49
MSA	**28**/11	−**43**/33	**39**/10	−**41**/50	**43**/11	−**22**/60
VOT ratio (-)						
HC	0.17/0.04	−0.54/0.09	0.22/0.05	−0.53/0.09	0.30/0.07	−0.54/0.08
PD	**0.22**/0.06	−0.46/0.11	0.27/0.09	−0.44/0.11	0.34/0.08	−0.44/0.12
PSP	**0.26**/0.06	−**0.42**/0.17	**0.31**/0.09	−0.43/0.16	0.36/0.07	−**0.37**/0.14
MSA	**0.26**/0.08	−**0.25**/0.22	**0.34**/0.07	−**0.19**/0.29	0.35/0.12	−**0.10**/0.34
Vowel duration (ms)						
HC	69/13	78/17	73/15	84/18	67/16	88/22
PD	63/11	73/12	65/12	77/11	60/10	82/13
PSP	84/42	**99**/47	90/38	**104**/41	**89**/36	**107**/40
MSA	77/25	84/23	78/35	91/33	77/28	90/24

VOT, voice onset time; HC, healthy controls; PD, Parkinson's disease; PSP, progressive supranuclear palsy; MSA, multiple system atrophy; SD, standard deviation.

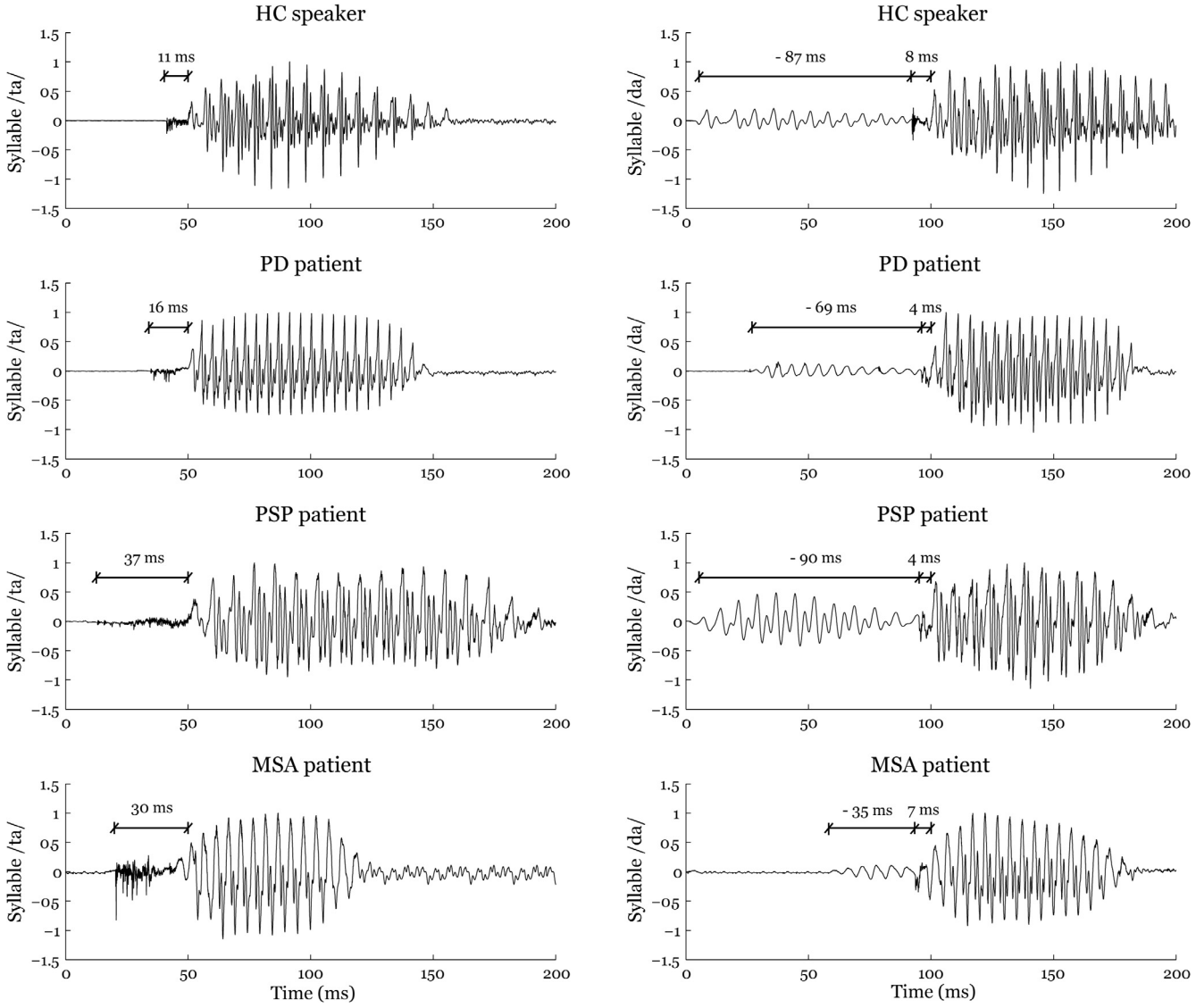


Fig. 1. Representative example of stop consonant duration of voiceless plosive /t/ and voiced plosive /d/ depicted for HC, PD, PSP and MSA subjects. *HC*, healthy controls; *PD*, Parkinson's disease; *PSP*, progressive supranuclear palsy; *MSA*, multiple system atrophy.

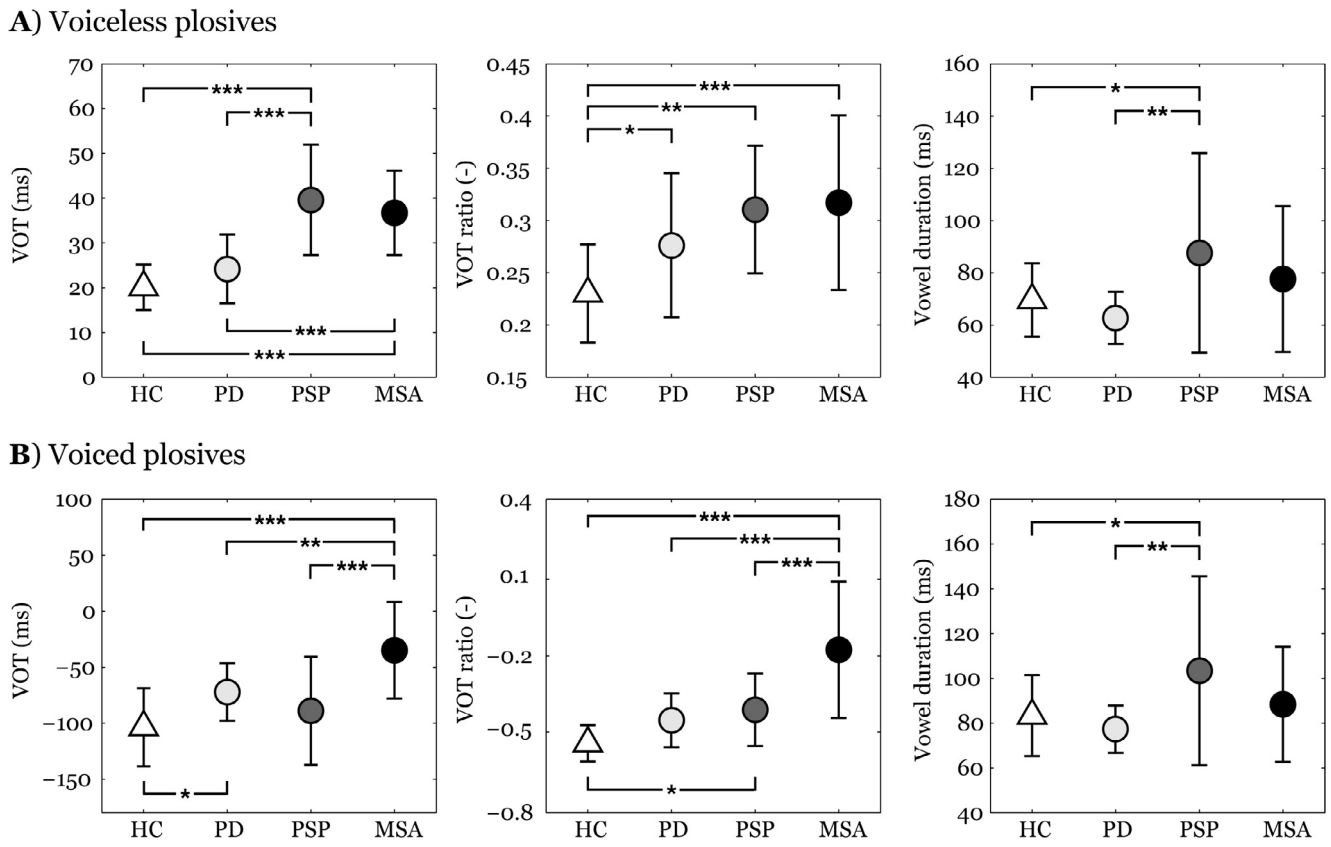


Fig. 2. Acoustic measures of VOT, VOT ratio, and vowel duration across HC, PD, PSP and MSA speakers for a subset of (A) voiceless plosives and (B) voiced plosives. Symbols represent mean values and error bars represent SD values. Statistical comparison between groups: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. *VOT*, voice onset time; *HC*, healthy controls; *PD*, Parkinson's disease; *PSP*, progressive supranuclear palsy; *MSA*, multiple system atrophy.

PSP ($p < 0.001$). Additionally, when compared to HC speakers, PD individuals exhibited shorter negative VOT ($p < 0.05$) and PSP patients smaller negative VOT ratio ($p < 0.05$). The vowel duration also varied among groups [$F(3,60) = 2.8$, $p < 0.05$, $\eta^2 = 0.12$] due to the PSP group, which exhibited slightly longer vowel length compared to both HC ($p < 0.05$) and PD ($p < 0.01$).

Fig. 3 displays results of perceptual analyses of phonetic contrast between voiceless and voiced plosives among PD, PSP, MSA and HC groups. No statistically significant difference between groups for Err/voiceless/ was observed [$\chi^2(3,60) = 3.9$, $p = 0.27$, $\eta^2 = 0.06$]. In contrast, Err/voiced/ varied significantly among groups [$\chi^2(3,60) = 17.1$, $p < 0.001$, $\eta^2 = 0.27$] due to the MSA group, which showed more misassignment errors than HC ($p < 0.001$), PD ($p < 0.01$), or PSP ($p < 0.05$).

The cerebellar motor impairment in APS patients correlated with a reduction of pre-voicing (cerebellar NNIPPS subscore *vs.* voiced VOT: $r = 0.59$, $p < 0.001$; cerebellar NNIPPS subscore *vs.* voiced VOT ratio: $r = 0.56$, $p = 0.001$) as well as with the percentage of voiced plosives perceptually misassigned as voiceless plosives (cerebellar NNIPPS subscore *vs.* Err/voiced/: $r = 0.58$, $p < 0.001$). No other significant correlations were detected between speech parameters and non-speech motor subscores in Parkinsonian groups. Additionally, overall dysarthria severity in all patients subgroups correlated with a prolonged duration of VOT of voiceless plosives (UPDRS III speech item *vs.* voiceless VOT: $r = 0.50$, $p < 0.001$). Significant correlation was also observed between the reduction of pre-voicing and percentage of voiced plosives perceptually misassigned as voiceless plosives (voiced VOT *vs.* Err/voiced/: $r = 0.65$, $p < 0.001$; voiced VOT ratio *vs.* Err/voiced/: $r = 0.73$, $p < 0.001$).

4. Discussion

Although disturbed consonant articulation is considered to be among the most deviant speech dimensions in all dysarthria subtypes (Chakraborty et al., 2008; Darley et al., 1969b), the present study represents the first attempt to clarify potential differences in consonant articulation deficits of both voiced and voiceless plosives in PD, PSP and MSA. We observed divergent patterns of articulation abnormalities for a subset of voiceless and voiced stop consonants in PD, PSP and MSA groups. Voice onset time of voiceless plosives was found to be more prolonged in both PSP and MSA compared to PD, likely due to the greater severity of dysarthria and slower articulation rate in APS. Voice onset time of voiced plosives was revealed to be significantly shorter only in MSA presumably as a consequence of damage to cerebellar structures. This finding was further supported by perceptual evaluation where only MSA patient target words with initial voiced plosive were misassigned as those with initial voiceless plosive. Acoustic analysis demonstrated that a slight deterioration of consonant articulation for both voiced and voiceless plosives was observed in PD as compared to HC. The observed trends with respect to imprecise articulation in PD, PSP and MSA were consistent among individual consonants as well as for averaged groups of voiceless and voiced consonants. The current method based on reading isolated two-syllabic words appears to be clinically feasible for assessing consonant articulation in Parkinsonism as the mean rate of measurable tokens was found to be greater than 93% for APS and 99% for PD or HC, and in this regard seems to be even better suited than the method based upon rhythmic syllable repetitions, where the mean rate of measurable VOT in dysarthrias was 85% (Ozsancak, Auzou, Jan, & Hannequin, 2001).

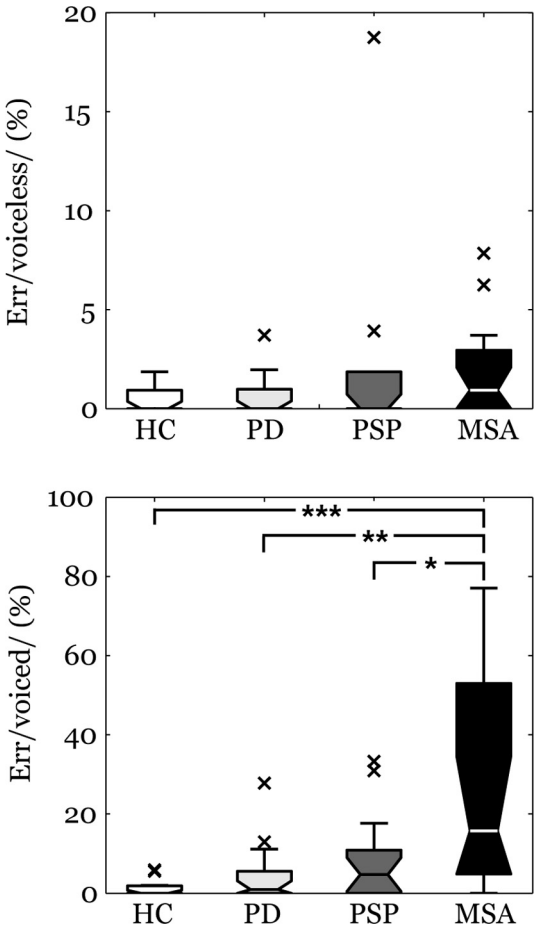


Fig. 3. Perceptual evaluation of voicing contrast based on Err/voiceless/ and Err/voiced/ measures across HC, PD, PSP and MSA speakers depicted using boxplots. Statistical comparison between groups: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. *Err/voiceless/*, percentage of voiceless plosives perceptually misassigned as voiced plosives; *Err/voiced/*, percentage of voiced plosives perceptually misassigned as voiceless plosives; *HC*, healthy controls; *PD*, Parkinson's disease; *PSP*, progressive supranuclear palsy; *MSA*, multiple system atrophy.

Regarding voiceless plosives, significantly longer VOT was observed in both PSP and MSA subjects in comparison to PD, likely as a result of greater speech impairment in APS. Indeed, we observed significant relationship between overall dysarthria severity and voiceless VOT. The observed differences between PD and APS also appear to be at least partially attributed to alterations in speaking rate as they were not preserved using the measurement of VOT ratio. Slow speaking rate is well documented in APS (Huh et al., 2015; Rusz et al., 2015; Skodda et al., 2011) and is primarily due to the presence of spastic components of dysarthria in APS patients (Rusz et al., 2015). In line with this observation, prolonged vowel duration associated, to a certain extent, with decreased speech rate was observed only in PSP patients, which commonly manifest spastic elements of dysarthria (Kluin et al., 1993).

On the contrary, the articulation of voiced plosives was markedly impaired only in MSA patients while nearly intact in PD and PSP subjects. In particular, the voiced plosives in MSA were characterized by the shortening of negative VOT duration until the voicing lead completely vanished and only short burst remained. In other words, the voiced plosives were occasionally pronounced as their voiceless cognates. Notably, contrary to voiceless plosives, these trends in voiced plosives were not suppressed using VOT ratio and thus cannot be interpreted as a simple effect of decreased speaking rate. Moreover, voiced VOT was not affected in PSP despite a similar overall severity of motor and speech impairment as compared to MSA. Consequently, we hypothesize that the disruption of voiced plosives in MSA reflects underlying cerebellar neurodegeneration. This hypothesis is further supported by the correlation between the severity of cerebellar involvement in APS and the disruption of voiced plosives as reflected by VOT and VOT ratio.

In agreement with the reduction of pre-voicing, the present study revealed a reduced perceptual contrast between voiced and voiceless plosives in MSA. This observation is in accordance with previous studies where reduced phonological contrast between /t/ and /d/ was documented in patients suffering from cerebellar atrophy (Ackermann & Hertrich, 1997; Ackermann et al., 1999), and generally confirms that VOT plays a primary role in perceptual determination between voiceless and voiced plosive cognates (Auzou et al., 2000). As perceptual judgment appears to be sufficient to reveal effects of cerebellar patterns on production of the voicing contrast in mild to moderate dysarthria, the current findings may have wider clinical implications in the evaluation of patients who may have MSA or other cerebellar degenerations.

Our further results indicate disturbed coordination between laryngeal and supralaryngeal articulators in PD patients that manifested a slightly reduced negative VOT as well as greater positive VOT ratio compared to HC. These findings are in agreement with previous research showing affected consonant articulation in PD (Auzou et al., 2000; Flint et al., 1992; Novotny et al., 2014; Saxena et al., 2014). However, previous results related to altered voice onset time in hypokinetic dysarthria secondary to PD are rather contradictory (Auzou et al., 2000; Fischer & Goberman, 2010; Saxena et al., 2014). Notably, our PD subjects had a tendency to prolong the duration of VOT in plosives with a naturally short VOT length, while shorten the duration of VOT in plosives with a naturally long VOT length. Although these trends may appear inconsistent, they can be interpreted with respect to knowledge about VOT categories in normal healthy speakers (Auzou et al., 2000). In most languages, VOT values for voiced and voiceless stops are in discrete duration ranges that correspond to one of three voicing categories including long negative VOT, short VOT and long positive VOT (Auzou et al., 2000). Thus, it can be assumed that in PD subjects, the plosives with short VOT duration are likely to be unchanged or extended beyond normal, while the plosives with both positive and negative long VOT are biased to be reduced. Indeed, previous studies in dysarthrias investigating stops with obviously long VOT duration reported shortening (Flint et al., 1992; Morris, 1989) while those examining naturally short VOT found increased or unchanged duration (Fischer & Goberman, 2010; Forrest et al., 1989; Novotny et al., 2014; Ravizza, 2003). However, other factors such as the effect of different speaking tasks or speaking rates cannot be excluded.

It thus remains to be elucidated by what mechanism the impaired programming of movements due to basal ganglia and cerebellar control circuit involvement contributes to imprecise articulation of voiced and voiceless stop consonants. Converging evidence from neuroimaging, limb control and neuropsychological studies suggests that patients with PD are unable to maintain a programmed response or to rapidly switch between responses, whereas patients with cerebellar disease have a reduced ability to program movement sequences in advance of movement onset (Spencer & Rogers, 2005). The accurate production of stop consonants requires close coordination between the larynx and the articulators (i.e., lips, tongue and jaw). Production of word-initial voiceless plosives involves a period of articulatory closure during which the vocal folds are maintained in a relatively open position without glottal pulsing, whereas word-initial voiced plosives are characterized by voicing lead at the beginning followed by a period of articulatory closure. To achieve precise coordination of the glottal opening gesture and articulatory closure release, it has been

suggested that the control of speech movement timing occurs over aggregates rather than individual articulators (Lofqvist & Lindblom, 1994; Munhall, Lofqvist, & Kelso, 1994). Difficulty in initiating articulation due to a reduced ability to maintain the speech motor program, which is characteristic for speakers with hypokinetic dysarthria (Gurd, Bessell, Watson, & Coleman, 1998; Svensson, Henningson, & Karlsson, 1993), appears to contribute to prolonged positive VOT in voiceless plosives and slightly shortened negative VOT in voiced plosives in PD, PSP as well as MSA patients, as observed in the present study. In MSA, among deficits aggravated due to the presence of hypokinetic dysarthria, insufficient programming in advance of speech onset due to ataxic dysarthria may lead to additional disruption of coordination between the larynx and the articulators, presumably resulting in substantially shortened or completely missing negative VOT in voiced plosives.

There are certain limitations to the present study. Due to limited opportunities in recruiting a larger number of participants we were not able to balance patient groups for gender, particularly in the PSP group where male participants predominated. We therefore cannot exclude the possibility that the observed changes in consonant articulation are partially affected by gender-specific aspects of speech. Nevertheless, no evident gender-specific differences for VOT have been reported in healthy adults in other languages (Lundeborg, Larsson, Wiman, & McAllister, 2012; Morris, McCrea, & Herring, 2008). Furthermore, as articulation of voiced consonants is not characterized by voicing lead in all languages (Auzou et al., 2000), our findings related to voiced stops are likely to be language-specific and may not be easily generalized to all languages. One potential limitation is that the test material was comprised of two-syllabic single-words only and thus the token-initial position of the target consonants precluded the evaluation of the consonantal closing gesture. In addition, our speech material was designed to minimize the effect of speaking rate and thus current findings may differ from those obtained from more complex speech utterances. Indeed, previous research assumed that the motor control of sentence utterances differs from that of syllable repetition (Ackermann & Hertrich, 1993). In particular, while there is clear evidence that VOT extracted from single syllables is sensitive to changes due to cerebellar atrophy (Ackermann & Hertrich, 1997; Ackermann et al., 1999), no alterations were observed in VOT obtained from sentence utterances in Friedreich ataxia (Ackermann & Hertrich, 1993), which can be considered a model of afferent cerebellar dysfunction. Thus, the cerebellum might not be a prerequisite for the initiation of articulatory gestures within the framework of sentence utterances (Ackermann & Hertrich, 1993), and single-word material may be a more feasible task than longer sentence utterances to observe distinctive patterns of imprecise articulation among PD and APS.

5. Conclusions

In conclusion, our results confirm the distinctive critical role of basal ganglia and cerebellar control circuit involvement in articulatory undershoot of voiced and voiceless stop consonants. In particular, duration of VOT in voiced plosives was revealed to be shorter only in MSA, while nearly intact in PD and PSP subjects. Considering that the only distinctive speech feature currently known for APS is the occurrence of stuttering-like behaviour in PSP (Kluin et al., 1993, 1996; Rusz et al., 2015), the alterations in VOT of voiced plosives may represent a novel marker of cerebellar dysfunction in MSA. Further studies are needed to elaborate our findings in other languages, through various kind of speech material, in larger populations of PD and APS, as well as across different types of neurodegeneration with cerebellar atrophy.

Acknowledgement

This study was supported by the Ministry of Health of the Czech Republic, grant nr. 15-28038A. All rights reserved.

References

Ackermann, H., Graber, S., Hertrich, I., & Daum, I. (1999). Phonemic vowel length contrasts in cerebellar disorders. *Brain and Language, 67*(2), 95–109. http://dx.doi.org/10.1006/brln.1998.2044.

Ackermann, H., & Hertrich, I. (1993). Dysarthria in Friedreich's ataxia: Timing of speech segments. *Clinical Linguistics & Phonetics, 7*(1), 75–91.

Ackermann, H., & Hertrich, I. (1997). Voice onset time in ataxic dysarthria. *Brain and Language, 56*(3), 321–333. http://dx.doi.org/10.1006/brln.1997.1740.

Auzou, P., Ozsancak, C., Morris, R. J., Jan, M., Eustache, F., & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clinical Linguistics & Phonetics, 14*(2), 131–150.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

Chakraborty, N., Roy, T., Hazra, A., Biswas, A., & Bhattacharya, K. (2008). Dysarthric Bengali speech: A neurolinguistic study. *Journal of Postgraduate Medicine, 54*(4), 268–272.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research, 12*(3), 462–496.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12*(2), 246.

deRijk, M. C., Tzourio, C., Breteler, M. M. B., Dartigues, J. F., Amaducci, L., LopezPousa, S., ... Rocca, W. A. (1997). Prevalence of Parkinsonism and Parkinson's disease in Europe: The EUROPARKINSON collaborative study. *Journal of Neurology Neurosurgery and Psychiatry, 62*(1), 10–15. http://dx.doi.org/10.1136/Jnnp.62.1.10.

Fischer, E., & Goberman, A. M. (2010). Voice onset time in Parkinson disease. *Journal of Communication Disorders, 43*(1), 21–34. http://dx.doi.org/10.1016/j.jcomdis.2009.07.004.

Flint, A. J., Black, S. E., Campbelltaylor, I., Gailey, G. F., & Levinton, C. (1992). Acoustic analysis in the differentiation of Parkinsons-disease and major depression. *Journal of Psycholinguistic Research, 21*(5), 383–399. http://dx.doi.org/10.1007/Bf01067922.

Forrest, K., Weismer, G., & Turner, G. S. (1989). Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. *Journal of the Acoustical Society of America, 85*(6), 2608–2622. http://dx.doi.org/10.1121/1.397755.

Gilman, S., Wenning, G. K., Low, P. A., Brooks, D. J., Mathias, C. J., Trojanowski, J. Q., ... Vidailhet, M. (2008). Second consensus statement on the diagnosis of multiple system atrophy. *Neurology, 71*(9), 670–676. http://dx.doi.org/10.1212/01.wnl.0000324625.00404.15.

Goetz, C. G., Leurgans, S., Lang, A. E., & Litvan, I. (2003). Progression of gait, speech and swallowing deficits in progressive supranuclear palsy. *Neurology, 60*(6), 917–922.

Gurd, J. M., Bessell, N., Watson, I., & Coleman, J. (1998). Motor speech versus digit control in Parkinson's disease: A cognitive neuropsychology investigation. *Clinical Linguistics & Phonetics, 12*(5), 357–378. http://dx.doi.org/10.3109/02699209808985231.

Hartelius, L., Gustavsson, H., Astrand, M., & Holmberg, B. (2006). Perceptual analysis of speech in multiple system atrophy and progressive supranuclear palsy. *Journal of Medical Speech-Language Pathology, 14*(4), 241–247.

Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1998). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology, 11*(3), 131–137.

Hornykiewicz, O. (1998). Biochemical aspects of Parkinson's disease. *Neurology, 51*(2), S2–S9.

Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical-diagnosis of idiopathic Parkinson's-disease – A clinicopathological study of 100 cases. *Journal of Neurology Neurosurgery and Psychiatry, 55*(3), 181–184. http://dx.doi.org/10.1136/jnnp.55.3.181.

Huh, Y. E., Park, J., Suh, M. K., Lee, S. E., Kim, J., Jeong, Y., ... Cho, J. W. (2015). Differences in early speech patterns between Parkinson variant of multiple system atrophy and Parkinson's disease. *Brain and Language, 147*, 14–20. http://dx.doi.org/10.1016/j.bandl.2015.04.007.

Kent, R. D., & Read, C. (1992). *The acoustic analysis of speech*. San Diego, Calif.: Singular Pub. Group.

Kim, Y., Kent, R. D., Kent, J. F., & Duffy, J. R. (2010). Perceptual and acoustic features of dysarthria in multiple system atrophy. *Journal of Medical Speech-Language Pathology, 18*(4), 66–70.

Kim, J. H., & McCann, C. M. (2015). Communication impairments in people with progressive supranuclear palsy: A tutorial. *Journal of Communication Disorders, 56*, 76–87. http://dx.doi.org/10.1016/j.jcomdis.2015.06.002.

Kluin, K. J., Foster, N. L., Berent, S., & Gilman, S. (1993). Perceptual analysis of speech disorders in progressive supranuclear palsy. *Neurology, 43*(3), 563–566.

Kluin, K. J., Gilman, S., Lohman, M., & Junck, L. (1996). Characteristics of the dysarthria of multiple system atrophy. *Archives of Neurology, 53*(6), 545–548.

Litvan, I., Agid, Y., Calne, D., Campbell, G., Dubois, B., Duvoisin, R. C., ... Verny, M. (1996). Clinical research criteria for the diagnosis of progressive supranuclear

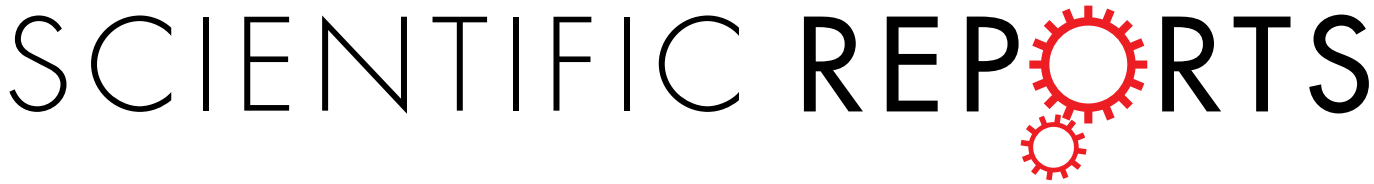
palsy (Steele-Richardson-Olszewski syndrome): Report of the NINDS-SPSP international workshop. *Neurology, 47*(1), 1–9.
Lofqvist, A., & Lindblom, B. (1994). Speech motor control. *Current Opinion in Neurobiology, 4*(6), 823–826.
Logemann, J. A., & Fisher, H. B. (1981). Vocal-tract control in Parkinsons-disease – Phonetic feature analysis of misarticulations. *Journal of Speech and Hearing Disorders, 46*(4), 348–352.
Lundeborg, I., Larsson, M., Wiman, S., & McAllister, A. M. (2012). Voice onset time in Swedish children and adults. *Logopedics Phoniatrics Vocology, 37*(3), 117–122. http://dx.doi.org/10.3109/14015439.2012.664654.
Morris, R. J. (1989). VOT and dysarthria: A descriptive study. *Journal of Communication Disorders, 22*(1), 23–33.
Morris, R. J., McCrea, C. R., & Herring, K. D. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics, 36*(2), 308–317.
Muller, J., Wenning, G. K., Verny, M., McKee, A., Chaudhuri, K. R., Jellinger, K., ... Litvan, I. (2001). Progression of dysarthria and dysphagia in postmortem-confirmed Parkinsonian disorders. *Archives of Neurology, 58*(2), 259–264. http://dx.doi.org/10.1001/archneur.58.2.259.
Munhall, K. G., Lofqvist, A., & Kelso, J. A. S. (1994). Lip-larynx coordination in speech – Effects of mechanical perturbations to the lower lip. *Journal of the Acoustical Society of America, 95*(6), 3605–3616. http://dx.doi.org/10.1121/1.409929.
Nath, U., Ben-Shlomo, Y., Thomson, R. G., Lees, A. J., & Burn, D. J. (2003). Clinical features and natural history of progressive supranuclear palsy – A clinical cohort study. *Neurology, 60*(6), 910–916.
Novotny, M., Rusz, J., Cmejla, R., & Ruzicka, E. (2014). Automatic evaluation of articulatory disorders in Parkinson's disease. *IEEE-ACM Transactions on Audio Speech and Language Processing, 22*(9), 1366–1378. http://dx.doi.org/10.1109/Taslp.2014.2329734.
Osaki, Y., Ben-Shlomo, Y., Lees, A. J., Daniel, S. E., Colosimo, C., Wenning, G. K., & Quinn, N. (2004). Accuracy of clinical diagnosis of progressive supranuclear palsy. *Movement Disorders, 19*(2), 181–189. http://dx.doi.org/10.1002/mds.10680.
O'Sullivan, S. S., Massey, L. A., Williams, D. R., Silveira-Moriyama, L., Kempster, P. A., Holton, J. L., ... Lees, A. J. (2008). Clinical outcomes of progressive supranuclear palsy and multiple system atrophy. *Brain, 131*, 1362–1372. http://dx.doi.org/10.1093/brain/awn065.
Ozsancak, C., Auzou, P., Jan, M., & Hannequin, E. (2001). Measurement of voice onset time in dysarthric patients: Methodological considerations. *Folia Phoniatrica Et Logopaedica, 53*(1), 48–57. http://dx.doi.org/10.1159/000052653.
Payan, C. A. M., Viallet, F., Landwehrmeyer, B. G., Bonnet, A. M., Borg, M., Durif, F., ... Grp, N. S. (2011). Disease severity and progression in progressive supranuclear palsy and multiple system atrophy: Validation of the NNIPPS - PARKINSON PLUS SCALE. *PLoS ONE, 6*(8). http://dx.doi.org/10.1371/journal.pone.0022293. Doi: ARTN e22293.
Ravizza, S. M. (2003). Dissociating the performance of cortical and subcortical patients on phonemic tasks. *Brain and Cognition, 53*(2), 301–310. http://dx.doi.org/10.1016/S0278-2626(03)00131-3.
Rusz, J., Bonnet, C., Klempir, J., Tykalova, T., Baborova, E., Novotny, M., ... Ruzicka, E. (2015). Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Journal of Neurology, 262*(4), 992–1001. http://dx.doi.org/10.1007/s00415-015-7671-1.
Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *Journal of the Acoustical Society of America, 129*(1), 350–367. http://dx.doi.org/10.1121/1.3514381.
Sachin, S., Shukla, G., Goyal, V., Singh, S., Aggarwal, V., Gureshkumar & Behari, M. (2008). Clinical speech impairment in Parkinson's disease, progressive supranuclear palsy, and multiple system atrophy. *Neurology India, 56*(2), 122–126.
Saxena, M., Behari, M., Kumaran, S. S., Goyal, V., & Narang, V. (2014). Assessing speech dysfunction using BOLD and acoustic analysis in Parkinsonism. *Parkinsonism & Related Disorders, 20*(8), 855–861. http://dx.doi.org/10.1016/j.parkreldis.2014.04.024.
Schrag, A., Ben-Shlomo, Y., & Quinn, N. P. (1999). Prevalence of progressive supranuclear palsy and multiple system atrophy: A cross-sectional study. *Lancet, 354*(9192), 1771–1775.
Simackova, S., Podlipsky, V. J., & Chladkova, K. (2012). Czech spoken in Bohemia and Moravia. *Journal of the International Phonetic Association, 42*(2), 225–232. http://dx.doi.org/10.1017/S0025100312000102.
Skodda, S., Visser, W., & Schlegel, U. (2011). Acoustical analysis of speech in progressive supranuclear palsy. *Journal of Voice, 25*(6), 725–731. http://dx.doi.org/10.1016/j.jvoice.2010.01.002.
Spencer, K. A., & Rogers, M. A. (2005). Speech motor programming in hypokinetic and ataxic dysarthria. *Brain and Language, 94*(3), 347–366. http://dx.doi.org/10.1016/j.bandl.2005.01.008.
Steele, J. C., Olszewski, J., & Richardson, J. C. (1964). Progressive supranuclear palsy – Heterogeneous degeneration involving brain stem basal ganglia + cerebellum with vertical gaze + pseudobulbar palsy nuchal dystonia + dementia. *Archives of Neurology, 10*(4), 333.
Svensson, P., Henningson, C., & Karlsson, S. (1993). Speech motor control in Parkinson's disease: A comparison between a clinical assessment protocol and a quantitative analysis of mandibular movements. *Folia Phoniatrica et Logopaedica (Basel), 45*(4), 157–164.
Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes – Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America, 92*(2), 723–735. http://dx.doi.org/10.1121/1.403997.
Weismer, G. (1984). Articulatory characteristics of Parkinsonian dysarthria: Segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination. In M. R. McNeil, J. C. Rosenbek, & A. E. Aronson (Eds.), *The dysarthrias*. Austin, Texas: Pro-Ed.
Wenning, G. K., Colosimo, C., Geser, F., & Poewe, W. (2004). Multiple system atrophy. *Lancet Neurology, 3*(2), 93–103. http://dx.doi.org/10.1016/S1474-4422(03)00662-8.
Wenning, G. K., Litvan, I., & Tolosa, E. (2011). Milestones in atypical and secondary Parkinsonisms. *Movement Disorders, 26*(6), 1083–1095. http://dx.doi.org/10.1002/mds.23713.

Appendix A9

Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder

Hlavnicka J, Cmejla R, Tykalova T, Sonka K, Ruzicka E, Rusz J (2017) Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. *Sci Rep* **7**:12.

Journal name (category): **Scientific Reports** (multidisciplinary sciences)
Quartile in category: **Q1**
Impact factor: **4.259**
Number of WoS citations (excluding autocitations): **3**



OPEN

Received: 29 September 2016
Accepted: 20 December 2016
Published online: 02 February 2017

Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder

Jan Hlavnička[1], Roman Čmejla[1], Tereza Tykalová[1], Karel Šonka[2], Evžen Růžička[2] & Jan Rusz[1,2]

For generations, the evaluation of speech abnormalities in neurodegenerative disorders such as Parkinson's disease (PD) has been limited to perceptual tests or user-controlled laboratory analysis based upon rather small samples of human vocalizations. Our study introduces a fully automated method that yields significant features related to respiratory deficits, dysphonia, imprecise articulation and dysrhythmia from acoustic microphone data of natural connected speech for predicting early and distinctive patterns of neurodegeneration. We compared speech recordings of 50 subjects with rapid eye movement sleep behaviour disorder (RBD), 30 newly diagnosed, untreated PD patients and 50 healthy controls, and showed that subliminal parkinsonian speech deficits can be reliably captured even in RBD patients, which are at high risk of developing PD or other synucleinopathies. Thus, automated vocal analysis should soon be able to contribute to screening and diagnostic procedures for prodromal parkinsonian neurodegeneration in natural environments.

As the most complex acquired human motor skill, speech is extremely sensitive to disturbances in the basal ganglia, which are involved in the planning, programming and execution of motor tasks[1–3]. Thus, speech changes are among the most robust motor abnormalities in Parkinson's disease (PD); a common neurological disorder associated with basal ganglia dysfunction. Up to 90% of PD patients develop perceptually distinctive speech and voice abnormalities, collectively termed hypokinetic dysarthria, characterized by decreased quality of voice, hypokinetic articulation, hypophonia, monopitch, monoloudness and deficits in timing[4,5]. However, previous research has mainly focused on the later stages of PD[5], while identification of different patterns of vocal disorders in the preclinical course of PD neurodegeneration has been severely restricted[6]. Identifying biomarkers related to neurodegeneration is essential as they could provide invaluable information not only related to prognosis and treatment, but also in the setting of clinical trials and disease modifying therapies before the onset of motor manifestations[7,8]. In this regard, vocal assessment has a potential advantage as an inexpensive, non-invasive and simple-to-administer method, scalable to large populations with the potential to be performed remotely, even by smartphone from the patient's home.

The development of a fully-automated method to detect early, distinctive patterns of neurodegeneration using only acoustic data from connected speech such as reading the short passage or monologue would have the potential to revolutionize the diagnostic process in neurodegenerative diseases manifesting motor speech disorders. The investigation of prodromal speech changes in subjects with rapid eye movement sleep behaviour disorder (RBD) provides a unique opportunity to evaluate the reliability and utility of such a tool. It is well known that people with RBD are at extremely high risk (>80%) for developing PD and related neurodegenerative disorders[7,8], and no prodromal disease marker has a predictive value near RBD[9].

[1]Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Circuit Theory, Technická 2, 166 27, Prague 6, Czech Republic. [2]Charles University in Prague, First Faculty of Medicine, Department of Neurology and Centre of Clinical Neuroscience, Kateřinská 30, 120 00, Prague 2, Czech Republic. Correspondence and requests for materials should be addressed to J.R. (email: rusz.mz@gmail.com)

	PD (n = 30)	RBD (n = 50)
Mean Age (years)	64.9 (SD 10.9, range 34–79)	64.9 (SD 9.1, range 40–83)
Men	70% (n = 21)	82% (n = 41)
Women	30% (n = 9)	18% (n = 9)
Positive history of Parkinson's disease in family	7% (n = 2)	2% (n = 1)
Mean age of disease onset (years)	63.4 (SD 11.9, range 30–78)	59.2 (SD 9.8, range 33–81)
Age of disease onset <40 years	10% (n = 3)	4% (n = 2)
Mean symptoms duration (years)	1.6 (SD 1.3, range 0.5–6)	5.8 (SD 4.4, range 1–17)
Mean Hoehn & Yahr score	2.1 (SD 0.3, range 1.5–2.5)	n/a
Mean UPDRS III total	20.2 (SD 12.4, range 6–54)	5.2 (SD 4.1, range 0–21)
Mean UPDRS III 18 speech item	0.4 (SD 0.5, range 0–1)	0.06 (SD 0.24, range 0–1)
Antidepressant therapy	10% (n = 3)	14% (n = 7)
Antiparkinsonian therapy	0 (n = 0)	0 (n = 0)
Levodopa equivalent (mg/day)	0	0
Clonazepam therapy	10% (n = 3)	12% (n = 6)
Clonazepam (mg/day)	0.07 (SD 0.22, range 0–1)	0.08 (SD 0.30, range 0–2)

Table 1. Clinical characteristics of newly diagnosed, untreated PD patients and RBD subjects. Captions: PD = Parkinson's disease, RBD = rapid eye movement sleep behaviour disorder, SD = standard deviation, n/a = not applicable.

The value of speech assessment in the differential diagnosis of motor speech disorders was suggested half a century ago in the landmark work by Darley *et al.*[10], noting that the constellation of specifically affected speech dimensions typically reflects the presumed underlying pathophysiology. Although there has been a substantial increase in computational power in recent years, most of the current methods for the evaluation of dysarthria patterns still rely either on perceptual tests, hand-labelling speech signal or manual control of the analysis procedure[11,12]. Nevertheless, several automated and quantitative approaches based on speech signal processing and machine learning have emerged for the evaluation of speech performance in PD[13,14]. However, these previous methods were designed for highly functional vocal paradigms such as sustained phonation or syllable repetition, and tested on only small samples of PD speakers[13,14]. Currently, no automatic, algorithm-based system is available that would allow robust and sensitive evaluation of different natural, connected speech patterns across a wide range of disease severity, from non-perceptible preclinical speech changes to the dysarthria in PD patients.

We developed a fully automated speech monitoring system that uses a segmentation method for the precise estimation of voiced and unvoiced segments of speech, respirations and pauses. We further proposed a set of acoustic speech features based on the segmentation algorithm applicable to connected speech, allowing the description of complex vocal disturbances due to neurodegeneration including respiratory deficits, dysphonia, imprecise articulation and dysrhythmia. We show that subliminal speech abnormalities can be reliably captured even in RBD patients and thus consider automated analysis suitable for research on vocal development in PD with potential clinical implications.

Results

Data collection. Typical vocal deficits associated with PD were examined using sample of 30 patients with newly diagnosed, untreated PD as compared to 50 healthy subjects without any history of neurological or communication disorders. Subsequently, we collected sample of 50 subjects with idiopathic RBD in order to reveal speech alterations that are typical for prodromal PD (see Table 1 and Methods). All participants were asked to perform two speaking tasks that represent natural speech and reflect motor speech disorders comprehensively[5]. First, speakers read a standardized, phonetically-balanced text of 80 words twice (Supplementary Fig. S1). Second, participants were instructed to provide monologue about their interests, job, family or current activities for approximately 90 seconds. Both PD and RBD subjects were scored according to the Unified Parkinson's Disease Rating Scale motor subscore (UPDRS III[15], ranging from 0 to 108, with 0 for no motor manifestations and 108 representing severe motor distortion). The severity of dysarthria in PD and RBD individuals was perceptually described by speech item 18 of the UPDRS III.

Automatic segmentation of connected speech. The main challenge of the proposed algorithm consisted in the class-by-class recognition of four basic physiological sources of connected speech including voiced speech, unvoiced speech, pause and respiration (see Methods, Fig. 1, Supplementary Fig. S2, and Supplementary Movie S1). To make the segmentation adaptive, the speech signal was processed inside a recognition window, where the signal was further parameterized by a sliding window into zero-crossing rate (ZCR), variance of autocorrelation function (ACR), power (PWR) and linear-frequency cepstral coefficients (LFCC). Voiced speech was determined using a cluster analysis in the space of ZCR, ACR and PWR. Subsequently, unvoiced speech was recognized using a cluster analysis in the space of the first five LFCC using voiceless intervals shorter than 300 ms in order to avoid misclassification with respirations. Consecutively, respirations were determined by a cluster analysis in the space of the first five LFCC using speechless intervals longer than 200 ms associated with the

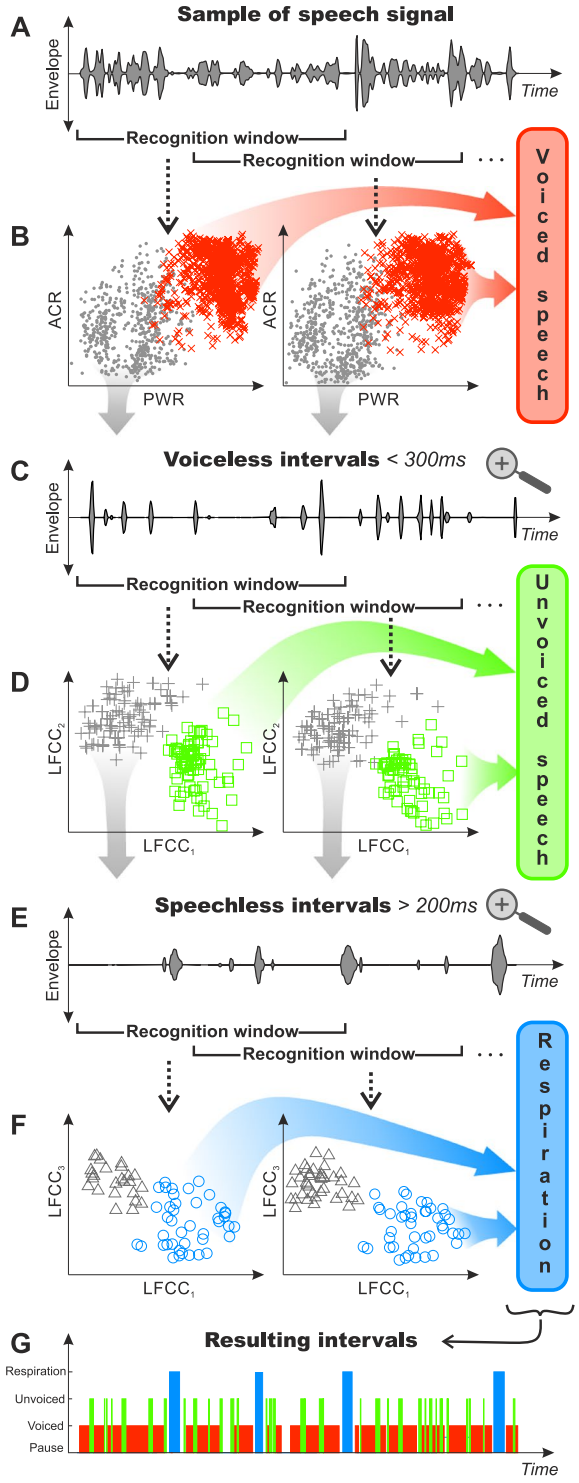


Figure 1. Process diagram illustrating the principle of the automated segmentation algorithm on a speech signal sample. (**A**) Original speech sample depicted using oscillographic sound pressure plot. (**B**) Red 'x' marks illustrate recognized voiced speech depicted in dimensions of PWR and ACR in two consecutive recognition windows. (**C**) Speech sample representing voiceless intervals shorter than 300 ms after removal of voiced intervals from original speech sample. (**D**) Green 'o' marks illustrate recognized unvoiced speech depicted in dimensions of the first linear-frequency cepstral coefficient ($LFCC_1$) and the second linear-frequency cepstral coefficient ($LFCC_2$) in two consecutive recognition windows. (**E**) Speech sample representing respiration intervals longer than 200 ms after removal of voiced and unvoiced intervals from original speech sample. (**F**) Blue 'o' marks illustrate recognized respirations depicted in dimensions of the first linear-frequency cepstral coefficient ($LFCC_1$) and the third linear-frequency cepstral coefficient ($LFCC_3$) in two consecutive recognition windows. (**G**) Resulting intervals of the segmentation plotted in time. PWR = power, ACR = variance of autocorrelation function.

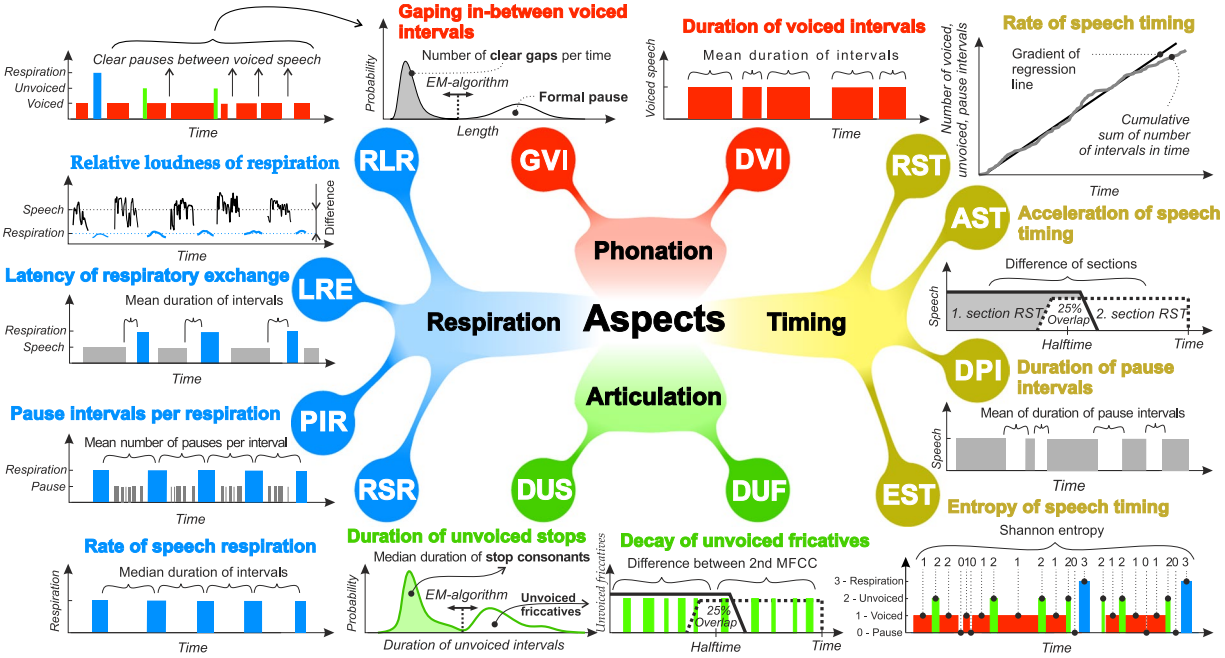


Figure 2. Mind map illustrating basic principles of individual acoustic speech features. MFCC = Mel-frequency cepstral coefficients.

approximate minimal duration required for inspiration. Finally, the resulting intervals were described by time labels for voiced speech, unvoiced speech, pause and respiration, where pauses include all respirations.

Tracking method performance. To compare and evaluate the reliability of the proposed segmentation method, manual reference labels were introduced for pause and respiration intervals. A total of 200 randomly chosen recordings including both speaking tasks across each group of participants was labelled blindly without awareness of segmentation output using speech analysis software[16]. Both pauses and respirations were labelled with respect to speech context using predefined criteria based on visual inspection of the spectrogram (see Methods).

The performance of the segmentation algorithm was evaluated for pause and respiration detection independently using hand labels. In addition, the performance of pause detection was compared with a previously designed pause detector for dysarthric speech[17] as well as a voice activity detector of ITU-T G.729B[18] commonly used in telecommunications for the reduction of transmission rate during silent periods of speech (see Methods).

The proposed segmentation method showed superior performance of $86.2 \pm 7.5\%$ efficiency across all pause lengths, in comparison to $55.4 \pm 9.6\%$ obtained using the previously designed pause detector for dysarthric speech[17], and $33.9 \pm 6.2\%$ by the voice activity detector[18]. As short pauses are less likely to be determined correctly by hand-labelling and long pauses play an important role in speech production, final performances of all detectors are described by the cumulative distribution of the mean detection efficiency depending on the duration of pause (Supplementary Fig. S3A). The pauses longer than approximately 100 ms are difficult to detect due to occurrence of respiratory signals that share similar characteristics with certain unvoiced fricatives (e.g. velar fricatives). Pauses shorter than approximately 100 ms are challenging to detect because non-speech turbulent airflow occurring during a pause can excite a spectrum similar to that of the preceding phoneme from articulators. In addition, insufficiently articulated unvoiced consonants can be hidden in natural noise background.

The proposed segmentation of respiration achieved $81.6 \pm 15.2\%$ efficiency through all respiration durations. Efficiency of respiration is expressed as the cumulative distribution of mean detection efficiency as a function of the duration of respiration (Supplementary Fig. S3B).

Acoustic speech features. Based on the outcome of adaptive segmentation, we designed a set of 12 acoustic features that were utilized with respect to speech disturbances associated with PD[5], allowing the assessment of all basic subsystems of connected speech including timing, articulation, phonation and respiration (see Methods, Fig. 2, and Supplementary Table S1).

Timing features involve information about the rhythmic organization of speech. We evaluated the speech rate with respect to quality of speech timing as the rate of speech timing (RST) including voiced, unvoiced and pause intervals. Acceleration of speech associated with parkinsonism was computed by acceleration of speech timing (AST) including voiced, unvoiced and pause intervals. Duration of pause intervals (DPI) describes the quality of speech timing, as pauses can be heavily influenced by the ability to properly initiate speech. Heterogeneity of speech in terms of the occurrence of voiced, unvoiced, pause and respiratory intervals was measured by entropy of speech timing (EST).

The assessment of articulation was quantified on intervals of unvoiced speech that represent pure involvement of the supra-laryngeal muscles. Performance of the most challenging articulatory movements represented by stop consonants was directly measured by the duration of unvoiced stops (DUS). In addition, the temporal quality of articulation was determined from unvoiced fricatives using the decay of unvoiced fricatives (DUF).

Phonatory characteristics were evaluated using intervals of voiced speech. The fundamental phonatory measurement was then the mean duration of voiced intervals (DVI). Deficits of phonatory onset and offset control were measured by gaping in-between voiced speech (GVI).

The respiratory apparatus was evaluated using data from detected respiratory intervals principally representing inspirations. The rate of speech respiration (RSR) robustly estimates respiratory rate during speech. Breath groups were described using pause intervals per respiration (PIR). The relative loudness of respiration (RLR) evaluates audibility of respiration relative to loudness of speech, eliminating dependence on microphone gain. The latency of respiratory exchange (LRE) measures the pause between expiration represented by the time speech ends and respective inspiration.

Forty-three percent of PD patients and only 6% of RBD subjects perceptually demonstrated mildly affected speech (score of 1) according to the UPDRS III speech item; 57% of PD patients and 94% of RBD subjects showed normal speech (score of 0). However, significant group differences were found for RST and DPI speech timing features, primarily associated with both differences between PD and control groups as well as between RBD and control groups (Fig. 3). Articulatory feature of DUS also significantly discriminated investigated groups, mainly due to the significant differences between RBD and control groups but also due to observed trend between PD and control groups ($p = 0.03$, uncorrected). All features except EST, DUS, RSR and LRE showed significant differences between reading passage and monologue (Fig. 3). No significant correlations were observed between any of the speech features and UPDRS III.

Index test. A blinded experiment was designed using UPDRS III with removal of action tremor to improve its predictive value[7] (hereafter UPDRS III*) (Fig. 4A), which is common in non-parkinsonian conditions[7]. UPDRS III* at a cut-off > 3 was previously revealed to be a very good indicator of initial parkinsonism[7], and thus RBD subjects with UPDRS III* ≤ 3 (hereafter asymptomatic RBD subgroup) were labelled as *motor negatives* and RBD subjects with UPDRS III* > 3 (hereafter symptomatic RBD subgroup) were labelled as *motor positives* (Fig. 4B). The speech performances of all RBD speakers were analysed automatically without any supervision (Fig. 4C). As PD patients show dysarthria profiles with unequal severity and different types of predominant speech manifestations[4,19], the speech pattern associated with PD was determined by sequential selection through all extracted speech features (Fig. 4D). This experiment was executed by quadratic discriminant function analysis, where each combination of features was evaluated using a leave-one-subject-out cross-validation scheme as follows: One individual speaker was excluded from the dataset iteratively, and the multivariate normal distribution characterized by the mean and covariance matrix was estimated using an Expectation Maximization algorithm for PD and controls independently. The excluded speaker was assigned to the distribution of PD or controls using Bayes discriminant rule. The best combination of acoustic features representing the most salient speech pattern of PD was determined as the one with highest accuracy (Fig. 4E). Finally, all RBD subjects were classified using the resulting speech pattern learned from the dataset of PD and control subjects (Fig. 4F). Speech performances of all RBD subjects assigned to the distribution of PD were labelled as *speech positives*, otherwise *speech negatives*. Finally, we obtained the true positive score as the number of *motor positives* equal to *speech positives*, the true negative score as the number of *motor negatives* equal to *speech negatives*, the false positive score as the number of *motor negatives* equal to *speech positives*, and the false negative score as the number of *motor positives* equal to *speech negatives*. Discriminative efficiency was then described by accuracy, sensitivity, and specificity.

As a result, a subgroup of 27 asymptomatic RBD subjects was classified using *motor negatives* labels, whereas a subgroup of 23 symptomatic RBD subjects was classified using *motor positives* labels. The most distinctive disturbed speech patterns between the PD and control groups were found for a combination of RST in reading passage, DVI in monologue, DPI in reading passage, DPI in monologue, DUS in reading passage, DUS in monologue and PIR in monologue with 71.3% accuracy (56.7% sensitivity and 80.0% specificity). Based upon the predictive model obtained through comparison between PD and controls, the results of the speech test indicate that *motor positives* and *motor negatives* from the RBD group were determined with 70.0% accuracy (73.9% sensitivity and 66.7% specificity) (Fig. 5).

Discussion

The results of our work represent the first step toward the development of a fully automated tool for the large-scale evaluation of prodromal vocal impairment due to neurodegeneration. Our findings indicate that the time for basic research through automated quantitative vocal analysis of natural speech is already upon us. The newly designed algorithm has proven sufficiently transparent to provide suggestions on typical patterns of parkinsonian prodromal vocalization deficits in RBD. Thus, the present study demonstrates the potential benefit of adding objective acoustic evaluation to the standard test battery used to identify those at high risk of developing neurodegeneration[20].

Interestingly, although the vast majority of RBD subjects were without perceptible speech impairment as indicated by the UPDRS III speech item, we objectively captured similar speech timing abnormalities in both de novo PD and RBD subjects. Indeed, converging evidence from neuroimaging, limb control and neuropsychological studies has suggested the presence of timing deficits in PD due to the inability to maintain a programmed response or to rapidly switch between responses[21,22]. Dysrhythmia patterns in our PD and RBD cohort included prolongation of pauses and decreased rate of speech intervals. While the prolongation of pauses in PD has been well documented[23], the decreased rate of speech intervals revealed in the present work provides new insight into the production of speech in PD. In particular, decreased rate of speech intervals indicates less diversity between

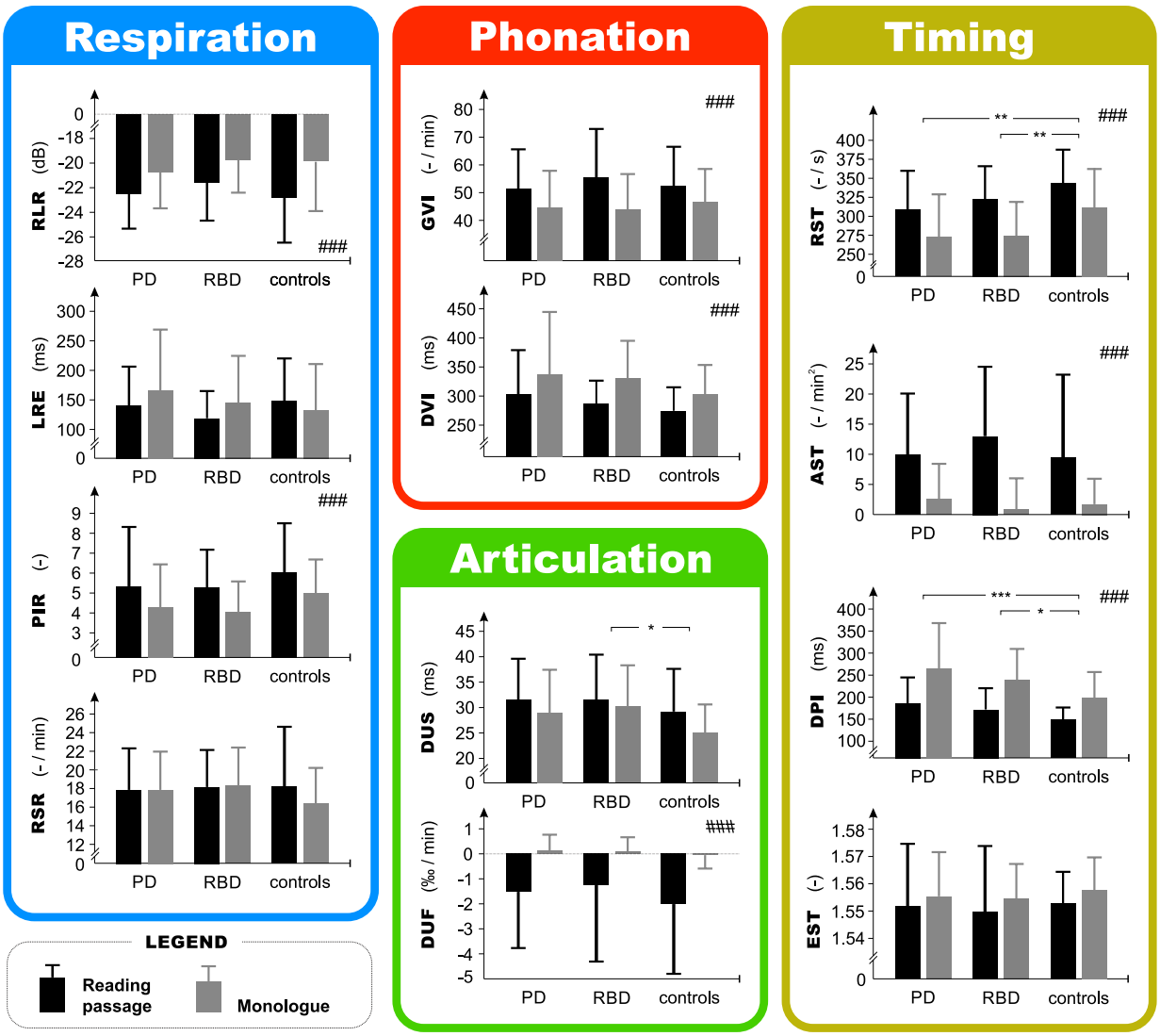


Figure 3. Results of acoustic speech analyses. Bars represent mean values and error bars represent SD values. Repeated measures analysis of variance (RM-ANOVA) was used to test for group differences: GROUP (PD vs. RBD vs. controls): corrected $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ after Bonferroni adjustment; TASK (reading passage vs. monologue): corrected $^{\#}p < 0.05$, $^{\#\#}p < 0.01$, $^{\#\#\#}p < 0.001$ after Bonferroni adjustment. None of the features showed significant GROUP × TASK interaction. RST = rate of speech timing, AST = acceleration of speech timing, DPI = duration of pause intervals, EST = entropy of speech timing, DUS = duration of unvoiced stops, DUF = decay of unvoiced fricatives, DVI = duration of voiced intervals, GVI = gaping in-between voiced intervals, RSR = rate of speech respiration, PIR = pause intervals per respiration, RLR = relative loudness of respiration, LRE = latency of respiratory exchange, PD = Parkinson's disease, RBD = rapid eye movement sleep behaviour disorder.

follow-up speech segments, likely as a result of decreased range of motion of the speech apparatus. On closer examination there were also trends toward changes in measurement of voiced intervals duration, representing voicing leakage through pauses, with mean values for the RBD group intermediate between PD patients and controls. Thus, the decreased rate of speech intervals and prolonged pauses appear to also be, at least partially, influenced by a decreased ability to stop voicing properly, which may reflect weak adduction of the vocal folds due to bradykinesia and rigidity of laryngeal muscles. In addition to the decreased ability to stop voicing properly, RBD subjects tend to extensively pronounce unvoiced stop consonants to acoustically resemble fricatives due to insufficiently closed articulators, which can be considered a precursor of the phenomenon called spirantization[24]. These observations are in agreement with previous research suggesting that PD with RBD may represent a distinct phenotype manifesting more as an akinetic-rigid syndrome, in comparison to PD without RBD[25,26].

In accordance with the majority of previous studies[27], we did not observe direct correlations between speech and motor assessment in PD and RBD. Nevertheless, we were able to blindly predict symptomatic motor RBD group membership with 70% accuracy based on speech assessment. In general, we may thus assume that speech impairment partially parallels increasing limb motor disability due to the underlying neurodegenerative process.

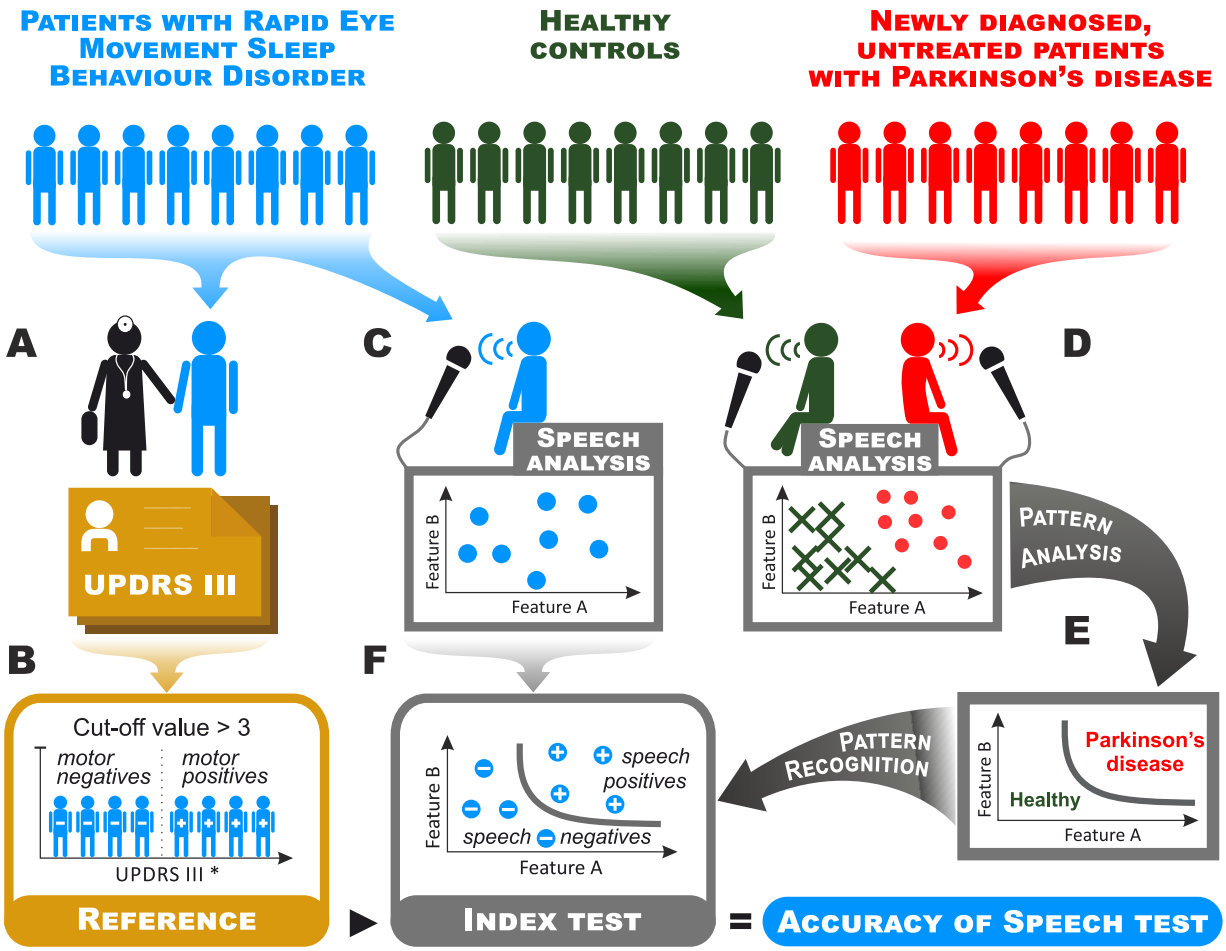


Figure 4. Flowchart describing the procedure of speech test experiment. (**A**) The motor score of UPDRS was examined in each RBD subject by a well-trained neurologist. (**B**) RBD subjects were separated into a subgroup of *motor negatives* (asymptomatic, UPDRS III* ≤ 3) and a subgroup of *motor positives* (symptomatic, UPDRS III* > 3). (**C,D**) Speech performances of all newly diagnosed, untreated PD patients, RBD subjects, and controls were analyzed using the set of designed speech features. (**E**) The most distinctive parkinsonian speech patterns were determined as the best resulting combination of speech features for differentiating between newly diagnosed, untreated PD patients and controls. (**F**) All RBD subjects were separated into a subgroup of *speech positives* (subjects with unchanged speech performance) and a subgroup of *speech negatives* (subjects with speech performance closer to PD speakers) based on the most distinctive parkinsonian speech pattern obtained through comparison between newly diagnosed, untreated PD and controls. PD = Parkinson's disease, RBD = rapid eye movement sleep behaviour disorder, UPDRS = Unified Parkinson's Disease Rating Scale; UPDRS III* = motor part of the UPDRS III score after removal of action tremor.

Considering that motor dysfunction strongly predicts disease onset, regardless of primary diagnosis of parkinsonism or dementia[7,20], the screening of motor speech changes may improve stratification for future neuroprotective therapy against PD and other synucleinopathies.

Automated segmentation methods for connected speech in dysarthria such as the speech-pause detector are very rare. One might assume that the precise identification of pauses is an easy task as it is used on a daily basis for voice activity detection in telecommunication transmission. However, the efficiency of detection was only 33.9% using a traditional voice activity detector[18], as it requires detection in the condition of environmental noise, and thus most non-speech sounds including respiration are accepted as voice activity. In particular, efficiency of detection was not substantially improved even using a previously designed pause detector for dysarthric speech[17], which reached an accuracy of only 55.4%. However, this method is based on thresholding the power envelope and thus will always struggle with any presence of non-speech sounds and the turbulent airflow of insufficiently closed articulators in general[17]. The pause detection of our algorithm reached a very high accuracy of 86.2% and substantially outperformed conventional methods. Additionally, although no respiration detector suitable for dysarthria evaluation is currently available, the proposed segmentation method also achieved high efficiency of 81.6% for the detection of respirations.

Our proposed algorithm is applicable for the assessment of complex parkinsonian vocal deficits and will likely be able to track speech progression as we found significant differences between controls and mild to moderate PD with more extended treatment periods across all speech subsystems including timing, articulation, phonation

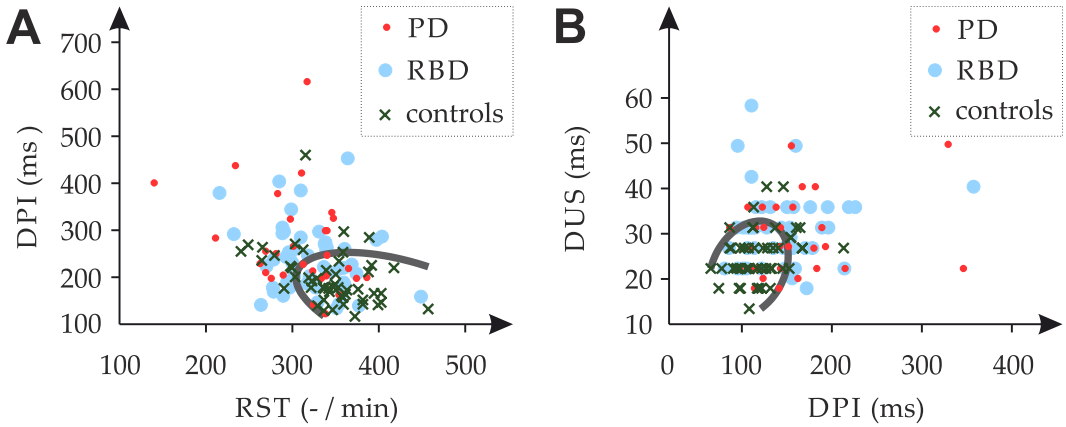


Figure 5. Selected pairs of representative acoustic features depicting the most distinctive speech pattern allowing differentiation between PD and controls. Individual speakers plotted in speech dimensions represented by (**A**) DPI in the monologue and RST in reading task, and (**B**) DPI in the reading task and DUS in monologue. The solid gray line represents the border of discrimination between *speech negatives* (subjects with unchanged speech performance) and *speech positives* (subjects with speech performance closer to PD speakers). The '·' marks represent PD, 'o' for RBD and 'x' for controls. PD = Parkinson's disease, RBD = rapid eye movement sleep behaviour disorder, DPI = duration of pause intervals, RST = rate of speech intervals, DUS = duration of unvoiced stops.

and respiration (Supplementary Fig. S4). In addition, developed algorithm also proved to be suitable for evaluation of both types of common connected speech including reading passage and monologue. While during the reading task the speaker is simply pronouncing ready-made text and thus can provide attention to articulatory planning, spontaneous speech requires more linguistic planning and the ability to formulate thoughts, allowing the speaker to modify the rhythm of his speech to breathe more freely[28]. Indeed, our findings indicate that most speech dimensions were influenced by the type of speaking task and both tasks can provide useful information. Although we did not find any interaction between task and group effects across speech dimensions, we still may assume that the temporal measurements such as speech acceleration and articulation decay are related to the content of speech and thus are better suited for a reading task as was previously shown[29]. It is also prudent to consider the effect of language and gender on speech characteristics. As the natural rhythm of the speech differs across languages[30], the specific speech dimensions designed in the present study cannot be simply transcribed to other languages, but need to be slightly rescaled with respect to common speech performances of healthy speakers of the given language. Interestingly, previous research has also suggested that gender may have a confounding impact on the progression of specific speech impairment in PD due to sexual dimorphism of laryngeal size[31]. In agreement with previous studies reporting a clearly increased incidence of PD in men[32] and a strong male predominance for RBD of up to 90%[33], our PD and RBD groups also consisted mainly of male participants and thus we cannot exclude the possibility that certain speech patterns may be influenced by gender-specific aspects. We should also point out that we did not stratify our PD patients according to the presence of RBD symptoms[23], as we wanted to test our algorithm using a representative sample of PD-related dysarthric speech that could be influenced by various factors such as disease phenotype. Finally, as the primary aim of the current study was to develop an automated monitoring system allowing the assessment of connected speech, we validated our findings based upon the reference to neurological evaluation using UPDRS III and did not evaluate any neuroimaging biomarkers that could support our findings. Recently, it has been shown that methods such as resting state functional magnetic resonance imaging or single-photon emission computed tomography scanning may provide sensitive indicators of early basal ganglia dysfunction[34,35]. Future studies are needed to evaluate our findings in other languages and in relation to other suitable biomarkers sensitive to prodromal neurodegeneration due to PD and other synucleinopathies.

In conclusion, our results indicate that the automated analysis of thoughtfully-selected acoustic features with well-defined pathophysiology from recordings of connected speech can be a reliable tool for monitoring vocalization deficits associated with neurodegeneration based on pathological alpha-synuclein storage, from non-perceptible preclinical to more advanced dysarthria stages. We believe that the procedure can be further elaborated and translated into other languages as well as to the entire spectrum of neurodegenerative disorders manifesting motor speech disorders. The current pilot findings provide novel opportunities for future research on motor speech disorders ranging from traditional laboratory-based analyses, monitoring the effect of therapy and disease progression, to the possibility of high-throughput screening for prodromal neurodegeneration, followed by more detailed analysis if the screen is abnormal.

Methods

Data collection. From 2014 to 2016, a total of 130 Czech native speakers were recruited for the study. For a given large effect size (Cohen's f of 0.4), we determined a minimum sample size of 84 with at least 30 per group by power analysis[36], with the error probability (α) set at 0.05 and a false negative rate (β) set at 0.1 (i.e. power of

0.9), based on a 3-group RM-ANOVA with one covariate (GROUP). Each participant provided written, informed consent and the study was approved by the Ethics Committee of the General University Hospital in Prague, Czech Republic (approval number 67/14 Grant VES AZV 1. LFUK). The study was carried out in accordance with the approved guidelines. Thirty patients (21 men, 9 women) with de novo, untreated PD, mean age 64.9 (standard deviation [SD] 10.9), were diagnosed upon the Parkinson's Disease Society Bank Criteria[37] (Table 1). In addition, 50 subjects (41 men, 9 women), mean age 64.9 (SD 9.1), were diagnosed with idiopathic RBD according to the International Classification of Sleep disorders diagnostic criteria, third edition[38] (Table 1). As a control group, 50 healthy subjects (41 men, 9 women), mean age 63.4 (SD 10.8) years, without a history of neurological or communication disorders, were included in the study.

The study was carried out at a single center. All PD patients were consecutively recruited at their first visit to the clinic and were examined in the drug-naive state, before symptomatic treatment was started. RBD subjects were screened through a web-based online survey[39]. No RBD patient complained of motor or cognitive difficulties or had a history of treatment with antiparkinsonian medication or any other therapy influencing sleep, cognition or motor features. Both diagnosis and evaluation of clinical scales were performed by a well-trained professional neurologist with experience in movement disorders. As the diagnosis of individual PD or RBD subjects was made at evaluation, the specific date recorded for each participant was different, but overall time schedule was the same. Each participant was first scored clinically by the neurologist and subsequently examined during a single session with a speech specialist. Speech data were recorded in a quiet room with a low ambient noise level using a head-mounted condenser microphone (Bayerdynamic Opus 55, Heilbronn, Germany) situated approximately 5 cm from the mouth. Recordings were sampled at 48 kHz with 16-bit resolution. None of the participants underwent speech therapy before the investigation.

Automatic segmentation of connected speech. All samples were preprocessed, and parameterization was established. Subsequently, each speech signal was modelled using a Gaussian mixture model (GMM), representing the most common method used in speech signal processing applications. Estimating all groups at once is not effective because all mixtures are separated imperfectly in single parametric space and false-positive errors as well as false-negative errors might occur. However, precise and robust classification is ensured if individual speech classes are estimated sequentially with respect to the corresponding traits in most differenced parameters. Sequential separation was executed via unique recognition steps where each recognition step separated previous distributions into two fractions (Supplementary Fig. S2).

Preprocessing. The signal was decimated to a sampling rate of 8 kHz, which is sufficient for speech recognition. Signal filtering using a 4th-order high-pass Chebyshev filter was performed to remove frequencies lower than 130 Hz, which include main hum, popping, and other subsonic sounds. Such a high cut-off frequency did not affect recognition but highlights the voiced speech.

Parameterization. The signal was parameterized inside a sliding window of 15 ms, in steps of 5 ms. The PWR, ACR and ZCR were calculated using the following equations:

$$\mathrm{PWR} = \frac{1}{N}\sum_{n=1}^{N} x^2[n] \cdot h[n], \tag{1}$$

$$R_x[k] = \frac{1}{N \cdot \sigma_x^2}\sum_{n=1}^{N}(x[n] - \mu_x) \cdot (x[n+k] - \mu_x), \tag{2}$$

$$\mathrm{ACR} = \frac{1}{N-1}\sum_{k=1}^{N}(R_x[k] - \overline{R}_x), \tag{3}$$

$$\mathrm{ZCR} = \frac{1}{N-1}\sum_{n=1}^{N-1}|\mathrm{sign}(R_x[n+1]) - \mathrm{sign}(R_x[n])|, \tag{4}$$

$$\mathrm{sign}(x[n]) = \begin{cases} 1, & x[n] \geq 0 \\ -1, & x[n] < 0 \end{cases}, \tag{5}$$

where x is a signal in window of length N, h is hamming window, R_x represents the autocorrelation function, σ_x is the standard deviation of the signal, and μ_x is the mean of the signal. Because all parameters can by described by a lognormal distribution, we expressed them on a logarithmical scale. ZCR was computed using the normalized autocorrelation function. This approach emphasizes harmonic frequency and suppresses the noise component of voiced consonants. Thus, all voiced phonemes (vowels and voiced consonants) can be described simply by a unimodal normal distribution of voiced speech, which boosts the sensitivity of detection of voiced speech. ACR was calculated as the variance of the normalized autocorrelation function of the unweighted signal. Signal was also parameterized in the spectrum using the first five of 24 LFCC, representing the low-frequency envelope of the power spectral density.

Sequential separation. The principle of sequential separation consisted in step-by-step recognition of the most differentiated components of speech (Supplementary Fig. S2). Speech was separated in the following order: voiced speech (Supplementary Fig. S2A), unvoiced speech (Supplementary Fig. S2B), and respiration (Supplementary Fig. S2C). The recognition step was executed inside a sliding recognition window. As a result, adaptability to the speech apparatus and environmental noise was achieved over time during speech recording. The GMM of parameters were presupposed in each position of the recognition window. GMM actually involve an unpredictable differentiation of mixtures. Therefore, the number of mixtures was evaluated using the Calinski-Harabasz index over the range $<2; 3>$. The optimal number of mixtures corresponded to a higher index of evaluation. GMM parameters were estimated using the EM-algorithm. Observations were classified via Bayes discriminant rule. Decisions were additionally smoothed by presupposing an Indo-European language family, in which unvoiced speech is accompanied by voiced speech. The values of thresholds of decision smoothing were not settled to exact value but only to approximate the natural timing of the speech apparatus.

Voiced speech. Voiced speech was separated from the whole signal using the parametric space of PWR, ACR, and ZCR within a recognition window of 20 seconds in 6 seconds steps. Voiced speech was identified as the component with the highest mean PWR. Decisions were smoothed using a median filter of the 5^{th} order and the following decision rules: voiced segments shorter than 30 ms were classified as voiceless but voiceless segments shorter than 20 ms were classified back as voiced, as such a short-term control over vocal fold function is hardly possible.

Unvoiced speech. Unvoiced speech represented by unvoiced consonants was separated from unvoiced segments shorter than 300 ms, which included most consonants and excluded most respiration signals. The first five LFCC were used for recognizing unvoiced consonants. The recognition window was 60 seconds long and featured 20 seconds steps. The prolongation of the recognition window compensated for the low occurrence of consonants. Unvoiced speech was identified as the component with the highest mean of the first LFCC, which is related to loudness. Unvoiced speech shorter than 5 ms and unvoiced speech in distance to voiced speech longer than 30 ms were rejected.

Pause. Pauses were defined as unvoiced and non-consonant signals, including the time required for respiration. The minimum duration of pauses was considered to be 30 ms.

Respiration. Respirations were separated from residual segments (excluding voiced speech and consonant segments already classified in the first two steps) longer than 200 ms using the first five LFCC. Respirations were identified as the component with the highest mean of the first LFCC. Respirations shorter than 40 ms were rejected. Inspirations were bounded by silence as the lungs reversed the direction of airflow. Therefore, respirations in distance to voiced speech shorter than 30 ms were classified as unvoiced speech. Gaps between respirations shorter than 400 ms were classified as respirations.

Labels. Each segment was described using labels pertaining to start time, end time, and class. All signal processing and data analysis steps were done in © Matlab (MathWorks).

Reference hand labels. The following criteria for pause annotations were derived from Fisher and Goberman[40]:

1. A signal interval can be annotated as a pause only if it contains no harmonic spectrum or noise exceeding the noise floor and exhibiting no formant structure similar to that of speech. Pauses can contain respirations isolated from speech or speech artefacts unrelated to the content of speech.
2. Pause preceding speech ends at time of the origin of formant structure accompanied by a harmonic spectrum or noise signal. If speech is initiated by an explosive consonant, then a pause ends at the time of an initial burst of energy.
3. Pause following speech begins at the time of the breaking of the formant structure associated with the previous phoneme.

The following criteria for respiration annotations were established:

1. Respirations were identified as noise substantially exceeding the noise floor with characteristic resonance within the 500–2000 Hz frequency band in pauses longer than 100 ms.
2. Borders of respiration were identified as the time of highest spectral change between pause and respiratory signals.
3. Non-continuous signals of respiration were labelled as homogenous respiration.

All labels were perceptually verified. Labels were described using time of interval start and end, as well as flag of segment origin including pause or respiration.

Algorithm performance evaluation. The gold standard for testing was based on reference hand labels, whereas a tolerance field was assigned to each label. Detected labels lying within the tolerance field were interpreted as true positives (TP). Each hand label was paired with only one detection. Unpaired reference hand labels were interpreted as false negatives (FN). Unpaired detections were interpreted as false positives (FP). The efficiency of detection F was evaluated by the F-score:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{6}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{7}$$

$$\text{F} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{8}$$

The tolerance field of pause was obtained around each label, with bounds corresponding to a quarter of the duration of the corresponding pause. The most interesting information pertaining to the segmentation efficiency was the dependence on pause length. Therefore, efficiency was iteratively computed across pauses longer than the progressive threshold from 50 ms to 300 ms in 50 ms steps. Respirations were evaluated within the tolerance field of duration of the corresponding respiration around each label.

The pause detector for dysarthric speech[17] was set to obtain the best performance result using a 200 bin histogram, a sampling rate of 8 kHz, and same preprocessing procedure used in the currently proposed segmentation method. The voice activity detection[18] was performed using a 8 kHz sampling rate and standard settings.

Acoustic speech features processing. To prevent distortion caused by pauses lasting more than several seconds, all pauses longer than 2 seconds were saturated to a maximal duration of 2 seconds.

RST provides a more robust estimate of speech rate impairment than a simple pause rate measurement as it considers not only pause but both voiced and unvoiced intervals. Voiced intervals provide additional information about impairment of phonatory control, whereas unvoiced intervals about imprecise articulation. The rate, including voiced, unvoiced, and pause intervals, approximates the speech rate complexly, as the speech rate impairment is related to deficits in all dimensions of speech. Each voiced, unvoiced, and pause interval was described by the time of occurrence, determined as the mean time between speech interval start and interval end. The total number of intervals was counted for each moment over the course of measurement. RST was computed as the gradient of the regression line of the time course.

AST determines the extent of timing acceleration. Speech run was split into two halftimes with 25% overlap, which provided a smooth transition between parts. AST was computed as the difference between RSTs of both parts divided by the total duration of the sentence.

DPI evaluates a speaker's ability to initiate speech. Complex speech impairment can cause difficulties in initiating speech, which cause prolongation of pauses. DPI was computed as the median duration of all pause intervals.

EST describes the orderliness or predictability of speech including voiced, unvoiced, pause, and respiratory intervals. Impaired speech associated with hypokinetic dysarthria tends to be more ordered and predictable, as voiced intervals dominate speech at the expense of other types of intervals. Accordingly, decreased entropy is tantamount to impaired speech. Each interval of speech was taken as one observation. Number of all intervals of speech were computed, including number of voiced speech intervals *nv*, number of unvoiced speech intervals *nu*, number of pause intervals *np*, number of respiration intervals *nr*, and total number of intervals *nt*. EST was determined as follows:

$$\text{EST} = -\frac{nv}{nt} \cdot \log_2\left(\frac{nv}{nt}\right) - \frac{nu}{nt} \cdot \log_2\left(\frac{nu}{nt}\right) - \frac{np}{nt} \cdot \log_2\left(\frac{np}{nt}\right) - \frac{nr}{nt} \cdot \log_2\left(\frac{nr}{nt}\right). \tag{9}$$

DUS assesses imprecise articulation manifested by increased noise accompanying stop consonants or even by continuant articulation perceived as fricative. Thus, the duration of detected stop consonants tends to increase. The length of unvoiced consonants forms a bimodal GMM of unvoiced fricative consonants and unvoiced stop consonants. This distinctly varied mixture was recognized using the EM-algorithm. DUS was computed as the median duration of recognized intervals of stop consonants.

DUF measures the temporal quality of articulation. Unvoiced fricatives are characterized by energy concentrated at high frequencies (>2.5 kHz). Information about temporal quality of articulation was measured as temporal damping of the high-frequency bulk. Speech run was split into two halftimes with 25% overlap, which provided a smooth transition between parts. Fricative consonants were recognized in each halftime from bimodal GMM of unvoiced fricative consonants and unvoiced stop consonants using the EM-algorithm. Unvoiced fricatives of each halftime were parameterized using 24 Mel-frequency cepstral coefficients (MFCC). The second MFCC is related to the ratio between energies of the high and low Mel-frequency bands. DUF was computed as the mean of differences between the second MFCC of both halftimes weighted on total duration of speech. DUF was expressed in parts per thousand with the respect to the presumed small range of values.

DVI evaluates the phonatory apparatus with respect to neglecting unvoiced consonants and pauses, voicing during voiceless consonants and simple fusing of multiple roots in word or fusing of independent words. DVI was computed as the mean duration of voiced intervals.

GVI represents the measurement of speaker ability to split voiced segments by pauses. Pauses between voiced intervals refer to the separation of different roots of words, independent words, and sentences in general. Impairment of the phonatory apparatus causes fusing of different roots of words or independent words. Pauses associated with unvoiced speech or respiration were rejected. The length of those clear pauses between voiced segments were described by a bimodal normal distribution. Formal pauses and clear gaps were recognized using the EM-algorithm. GVI was computed as the number of clear gaps per total time of speech.

RSR estimates the breathing rate from detected intervals of respiration. The computation of RSR was designed to be robust against misdetections of respiratory intervals. Each respiration was described by the mean time between respiration start and respiration end. The speed of respiration was determined as the inverted median value of this interval expressed in minutes.

PIR aims to examine respiratory and prosodic function in a context characterized by pause production during a breath group. Commonly, a decrease in breath group together with a decrease in pause rate anticipate decreased PIR due to dysarthria. The number of pauses framed between two respirations was taken into account. PIR was computed as the median number of pauses per respiration.

RLR represents a simple loudness measurement for estimating inspiratory effort. Every obstruction in the respiratory path produces noise. The loudness of inspiratory noise is proportional to inspiratory flow and the resistance of obstruction in the respiratory pathway. Limited chest wall kinematics associated with dysarthria reduce inspiratory flow and noise. The loudness of respiratory noise was referenced to the loudness of speech to make the value of RLR independent of microphone gain. The signal was squared and filtered by a moving average of 120 order and rescaled into loudness using a logarithmical scale. RLR was determined as the difference between the median loudness of respirations and the median loudness of speech.

LRE indicates the ability to convert expiration during speech into inspiration. The impaired ability of respiratory exchange manifests in the prolongation of the interval between expiration and inspiration, which must be minimized in speech breathing. Each respiration was evaluated by the latency time, which was computed as the difference between the time of inspiration and the time of expiration. The time of expiration end was determined from the end of the nearest preceding speech interval. The time of inspiration start was determined from the start of the interval of respective respiration. LRE was computed as the mean of latency times of all detected respirations.

Statistical analysis. For every subject, the final values of 12 acoustic features related to reading passage were computed by averaging the data obtained in two vocal task runs. As all acoustic features were found to be normally distributed by the one-sample Kolmogorov-Smirnov test, statistical analyses were performed using repeated measures analysis of variance with GROUP (PD vs. RBD vs. controls) treated as a between-group factor and TASK (reading passage vs. monologue) treated as a within-group factor. Post-hoc GROUP significance was assessed with the Fisher least-squares difference. Pearson correlations were applied to test for significant relationships. Bonferroni correction for multiple comparisons was applied according to the 12 tests performed with corrected p threshold $= 0.0042$ (i.e., 0.05/12) for $p < 0.05$.

References

1. Kent, R. D., Kent, J. F., Weismer, G. & Duffy, J. R. What dysarthrias can tell us about the neural control of speech. *J. Phonetics* **28**, 273–302 (2000).
2. Graber, S., Hertrich, I., Daum, I., Spieker, S. & Ackermann, H. Speech perception deficits in Parkinson's disease: Underestimation of time intervals compromises identification of durational phonetic contrasts. *Brain Lang.* **82**, 65–74 (2002).
3. Ho, A. K., Bradshaw, J., Iansek, R. & Alfredson, R. Speech volume regulation in Parkinson's disease: Effects of implicit cues and explicit instructions. *Neuropsychologia* **37**, 1453–1460 (1999).
4. Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L. & Gates, S. Speech impairment in a large sample of patients with Parkinson's disease. *Behav. Neurol.* **11**, 131–137 (1998).
5. Duffy, J. R. Motor Speech Disorders: Substrates, Differential Diagnosis and Management, 3rd ed., Mosby, St. Louis (2013).
6. Rusz, J. *et al.* Quantitative assessment of motor speech abnormalities in idiopathic REM sleep behaviour disorder. *Sleep Med.* **19**, 141–147 (2016).
7. Postuma, R. B., Lang, A. E., Gagnon, J. F., Pelletier, A. & Montplaisir, J. Y. How does Parkinsonism start? Prodromal Parkinsonism motor changes in idiopathic REM sleep behaviour disorder. *Brain* **135**, 1860–1870 (2012).
8. Schenck, C. H. *et al.* Rapid eye movement sleep behaviour disorder: devising controlled active treatment studies for symptomatic and neuroprotective therapy-a consensus statement from the International Rapid Eye Movement Sleep Behaviour Disorder Study Group. *Sleep Med.* **14**, 795–806 (2013).
9. Postuma, R. B. *et al.* Identifying prodromal Parkinson's disease: Pre-Motor disorders in Parkinson's disease. *Mov. Disord* **27**, 617–626 (2012).
10. Darley, F. L., Aronson, A. E. & Brown, J. R. Differential diagnostic patterns of dysarthria. *J. Speech Lang. Hear. Res.* **12**, 47–57 (1969).
11. Sussman, J. E. & Tjaden, K. Perceptual measures of speech from individuals with Parkinson's disease and multiple sclerosis: Intelligibility and beyond. *J. Speech Lang. Hear. Res.* **55**, 1208–1219 (2012).
12. Skodda, S., Grönheit, W., Mancinelli, N. & Schlegel, U. Progression of voice and speech impairment in the course of Parkinson's disease: a longitudinal study. *Parkinsons Dis.* **389195** (2013).
13. Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J. & Ramig, L. O. Suitability of dysphonia measurement for telemonitoring of Parkinson's disease. *IEEE T. Biomed. Eng* **56**, 1015–1022 (2009).
14. Novotný, M., Rusz, J., Čmejla, R. & Růžička, E. Automatic evaluation of articulatory disorders in Parkinson's disease. *IEEE/ACM T. Audio Speech Lang. Proces* **22**, 1366–1378 (2014).
15. Stebbins, G. T. & Goetz, C. G. Factor structure of the Unified Parkinson's Disease Rating Scale: motor examination section. *Mov. Disord* **13**, 633–636 (1998).
16. Boersma, P. Praat, a system for doing phonetics by computer. *Glot International* **5**, 341–345 (2001).
17. Rosen, K. *et al.* Automatic method of pause measurement for normal and dysarthric speech. *Clin. Linguist. Phonet.* **24**, 141–154 (2010).
18. International Telecommunication Union. Standardization Sector of ITU. ITU-T G.729, WTSC-96, Geneva, 1–39 (1996).
19. Rusz, J., Čmejla, R., Růžičková, H. & Růžička, E. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *J. Acoust. Soc. Am.* **129**, 350–367 (2011).
20. Postuma, R. B., Gagnon, J. F., Bertrand, J. A., Marchand, D. G. & Montplaisir, J. Y. Parkinson risk in idiopathic REM sleep behavior disorder Preparing for neuroprotective trials. *Neurology* **84**, 1104–1113 (2015).
21. Benecke, J., Rothwell, C., Dick, P. J. R., Day, P. L. & Marsden, C. D. Disturbance of sequential movements in patients with Parkinson's disease. *Brain* **110**, 361–379 (1987).
22. Spencer, K. A. & Rogers, M. A. Speech motor programming in hypokinetic and ataxic dysarthria. *Brain Lang.* **94**, 347–366 (2005).
23. Hammen, V. L. & Yorkston, K. M. Speech and pause characteristics following speech rate reduction in hypokinetic dysarthria. *J. Commun. Disord.* **29**, 429–445 (1996).

24. Weismer, G. & McNeil, M. Articulatory characteristics of Parkinsonian dysarthria: Segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination. The dysarthrias: Physiology, acoustics, perception, management, College-Hill Press, San Diego 101–130 (1984).
25. Romenets, S. R. *et al.* Rapid eye movement sleep behavior disorder and subtypes of Parkinson's disease. *Mov. Disord* **27**, 996–1003 (2012).
26. Gong, Y. *et al.* Clinical manifestations of Parkinson disease and the onset of rapid eye movement sleep behavior disorder. *Sleep Med.* **15**, 647–653 (2014).
27. 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.* **57**, 1330–1343 (2014).
28. Levelt, W. J. M. *Speaking: From Intention to Articulation. A Bradford Book*. The MIT Press: Cambridge (Massachusetts)–London (England), 1989; 1–545.
29. Skodda, S. & Schlegel, U. Speech rate and rhythm in Parkinson's disease. *Mov. Disord* **23**, 985–992 (2008).
30. Ramus, F., Nespor, M. & Mehler, J. Correlates of linguistic rhythm in the speech signal. *Cognition* **73**, 265–292 (1999).
31. Hertrich, I. & Ackermann, H. Gender-specific vocal dysfunctions in Parkinson's disease: electroglottographic and acoustic analyses. *Ann. Oto. Rhinol. Laryn* **104**, 197–202 (1995).
32. de Lau, L. M. & Breteler, M. M. Epidemiology of Parkinson's disease. *Lancet Neurol.* **5**, 525–535 (2006).
33. Trotti, L. M. REM sleep behaviour disorder in older individuals. *Drugs Aging* **27**, 457–470 (2010).
34. Rolinski, M. *et al.* Basal ganglia dysfunction in idiopathic REM sleep behaviour disorder parallels that in early Parkinson's disease. *Brain* **139**, 2224–2234 (2016).
35. Berg, D. *et al.* MDS research criteria for prodromal Parkinson's disease. *Mov Disord* **30**, 1600–1611 (2015).
36. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
37. Hughes, A. J., Daniel, S. E., Kilford, L. & Lees, A. J. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J. Neurol. Neurosurg. Psychiatry* **55**, 181–184 (1992).
38. American Academy of Sleep Medicine. International Classification of Sleep Disorders, Third Edition: Diagnostic and Coding Manual, Westchester, Illinois: American Academy of Sleep Medicine (2014).
39. Buskova, J., Ibarburu, V., Sonka, K. & Ruzicka, E. Screening for REM sleep behavior disorder in the general population. *Sleep Med.* 24, 147 (2016).
40. Fischer, E. & Goberman, A. M. Voice onset time in Parkinson disease. *J. Commun. Disord.* **43**, 21–34 (2010).

Acknowledgements

This study was supported by the Ministry of Health of the Czech Republic, grant no. 15-28038A and grant no. 16-28914A, and by the Czech Science Foundation, grant no. 16-07879S.

Author Contributions

J.H. and J.R. conceived and designed the experiments. J.H., T.T., K.S., E.R. and J.R. performed the experiments. J.H. and J.R. analysed the data. J.H. developed and tested analysis method used in the project. J.H. wrote the paper. All authors were involved in discussion of results and revision of the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-00047-5

Competing financial interests: The authors declare no competing financial interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix A10

Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease

Rusz J, Hlavnicka J, Tykalova T, Novotny M, Dusek P, Sonka K, Ruzicka E (2018) Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease. *IEEE Trans Neural Syst Rehabil Eng* 26:1495-1507.

Journal name (category): **IEEE Transactions on Neural Systems and Rehabilitation Engineering** (engineering, biomedical/rehabilitation)
Quartile in category: **Q1/Q1**
Impact factor: **3.972**
Number of WoS citations (excluding autocitations): **0**

Smartphone Allows Capture of Speech Abnormalities Associated With High Risk of Developing Parkinson's Disease

Jan Rusz, Jan Hlavnička, Tereza Tykalová, Michal Novotný, Petr Dušek, Karel Šonka, and Evžen Růžička

***Abstract*— Although smartphone technology provides new opportunities for the recording of speech samples in everyday life, its ability to capture prodromal speech impairment in persons with a high risk of developing Parkinson's disease (PD) has never been investigated. Speech data were acquired through a smartphone as well as a professional microphone with a linear frequency response from 50 participants with a rapid eye movement sleep behavior disorder that are at a high risk of developing PD and related neurodegenerative disorders. Additionally, recordings of 30 newly diagnosed, untreated PD patients and 30 healthy participants were evaluated. Acoustic assessment of 11 speech dimensions representing the key aspects of hypokinetic dysarthria in the early stages of PD was performed. Smartphone allowed the detection of speech abnormalities in participants with a high risk of developing PD. Acoustic measurements related to fundamental frequency variability, duration of pause intervals, and rate of speech timing extracted from spontaneous speech were sufficiently sensitive to significantly separate groups (area under curve of 0.85 between PD and controls) and showed very strong correlation and reliability between the professional microphone and the smartphone. Speech-based biomarkers collected through smartphones may have the potential to revolutionize the diagnostic process in neurodegenerative diseases and improve stratification for future neuroprotective therapy in PD.**

Manuscript received March 13, 2018; revised May 23, 2018 and June 21, 2018; accepted June 26, 2018. Date of publication June 29, 2018; date of current version August 7, 2018. This work was supported in part by the Ministry of Health of the Czech Republic under Grant $15-28038A$ and Grant $16-28914A$, in part by the Czech Science Foundation under Grant $16-07879S$, and in part by the Research Center for Informatics under Grant CZ.02.1.01/0.0/0.0/16_019/0000765. *(Jan Rusz and Jan Hlavnička contributed equally to this work.) (Corresponding author: Jan Rusz.)*

J. Rusz is with the Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, 160 00 Prague, Czech Republic, and also with the Centre of Clinical Neuroscience, Department of Neurology, First Faculty of Medicine, Charles University, 120 00 Prague, Czech Republic (e-mail: ruszjan@fel.cvut.cz).

J. Hlavnička, T. Tykalová, and M. Novotný are with the Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, 160 00 Prague, Czech Republic (e-mail: hlavnjan@fel.cvut.cz; tykalter@fel.cvut.cz; novotm26@fel.cvut.cz).

P. Dušek, K. Šonka, and E. Růžička are with the Centre of Clinical Neuroscience, Department of Neurology, First Faculty of Medicine, Charles University, 120 00 Prague, Czech Republic (e-mail: petr.dusek@lf1.cuni.cz; ksonka@lf1.cuni.cz; eruzi@lf1.cuni.cz).

Digital Object Identifier 10.1109/TNSRE.2018.2851787

***Index Terms*— Acoustic analyses, cellular phone, Parkinson's disease, REM sleep behavior disorder, speech disorder.**

I. Introduction

PARKINSON'S disease (PD) is a neurodegenerative disorder characterized by the pathological accumulation of abnormal α-synuclein in the brain leading to the loss of vulnerable neuronal populations including dopaminergic neurons in the substantia nigra. As a consequence, individuals with PD present with bradykinesia, in addition to other principal motor features such as rigidity and/or resting tremor [1]. The incidence of PD in the overall population is estimated to be 1.8 % in persons 65 years of age and older [2]. More effective therapies for neurodegenerative diseases such as PD will become a strategic priority due to the increasing economic burden connected with prolonged life expectancy [3]. However, there currently is no treatment to halt or slow the progression of PD. Available pharmacotherapy and neurosurgical interventions only offer symptomatic alleviation of PD motor symptoms that clinically manifest relatively late in the course of neurodegeneration, at the moment when up to 50 % of the neurons in the substantia nigra is already irrecoverably damaged and up to 80 % of striatal dopamine has been depleted [4], [5]. Therefore, the major reason for the failure to develop neuroprotective therapy may be that the disease progresses for many years before the appearance of clinical signs and then it is simply too late for intervention. The early recognition of PD in prodromal stages has thus crucial implications for the future development of neuroprotective therapy [7], [8].

Idiopathic rapid eye movement sleep (REM) behavior disorder (RBD) is a sleep disorder characterized by the loss of physiologic skeletal muscle atonia during REM sleep, resulting in motor responses related to dream content [8]. During sleep, persons with RBD make rapid movements, kick, scream, talk aloud and commonly cause injury to themselves or their bed partners [9]. Idiopathic RBD is a prodromal marker of synucleinopathies, which are neurodegenerative disorders characterized by pathological α-synuclein deposits in the brain. This group includes PD and two related disorders also manifesting with parkinsonism: dementia with Lewy bodies and multiple

system atrophy. Risk of conversion of idiopathic RBD into PD, dementia with Lewy bodies and less frequently into multiple system atrophy is extremely high (> 80%) [10]–[12]. Conversely, only a small minority of subjects with RBD do not develop a neurodegenerative disorder [6]. The high conversion rate of RBD to neurodegenerative disease provides a unique opportunity to observe the clinical development of PD and related disorders [13], [14]. No other preclinical marker has comparable predictive value to RBD with regard to synucleinopathy development [15].

As the most complex motor skill acquired involving more than 100 muscles, speech is highly susceptible to degeneration of neural structures engaged in motor system control [16]. Indeed, up to 90 % of PD patients develop distinctive speech and voice abnormalities, collectively termed hypokinetic dysarthria, characterized by decreased quality of voice, hypokinetic articulation, hypophonia, monopitch, monoloudness or deficits in timing [17]. In clinical practice, experienced clinicians frequently develop an intuitive sense that PD is present due to characteristic changes in voice. This resonates with the reports of family members, who perceptually note changes in speech long before the occurrence of cardinal PD manifestations and before the diagnosis is established [18]. However, evidence for prodromal motor speech changes in presymptomatic PD is extremely limited [19], [20].

Acoustic analyses allow for the objective quantification of these perceptual impressions of characteristic changes in voice due to PD. That, together with the fact that the recording and processing of human speech is an area with an extensive background of knowledge, makes speech changes an excellent candidate as a diagnostic biomarker of PD. At the time of diagnosis, multidimensional speech impairments represented by abnormal phonation, articulation, and prosody are already detectable in PD patients [21]. The comprehensive screening for these abnormalities typically requires the processing of three basic speaking tasks including prolonged vowel phonation, fast syllables repetition and connected speech. Based upon these paradigms, several representative acoustic measures with well-defined pathophysiological interpretation have been shown to possess an accuracy of more than 80 % in the differentiation between newly diagnosed, untreated PD and healthy controls [21]–[23]. These measures may be considered as speech-related diagnostic biomarkers of PD. In particular, decreased quality of vocal fold function has been evaluated via a paradigm of sustained phonation by measures related to perturbations of acoustic signal [21]. Problems with articulators have been captured through a paradigm of fast syllable repetition and abnormalities related to slower or irregular sequential motion rates (SMR) and imprecise consonant articulation [22]. Finally, several prosodic deviations including monopitch, monoloudness, articulatory decay, inappropriate duration of pauses, and decreased rate of follow-up speech segments have been elicited via natural spontaneous speech [19]–[21].

Unfortunately, all previous findings were based upon speech recordings obtained using a professional condenser microphone, which considerably limits their broader applicability. In this regard, vocal assessment by smartphone provides intriguing potential advances as it is inexpensive, non-invasive, scalable to a large population, simple to administer and can be performed remotely from the patients' home. Furthermore, recordings can be sent to a remote specialized server for processing or even can be processed directly via smartphone by acoustic analyses of speech that can be fully automated [20], [22]. Therefore, acoustic analyses allow the possibility of high-throughput screening, which can be followed by more detailed medical examination if the screen is abnormal. Given the advancement in the development of mobile technologies, collecting data through smartphones continues to be a growing focus, not only in speech-related research. Thanks to their ubiquity, smartphones allow much more frequent collection of data with much lower costs. Indeed, pilot research has demonstrated that embedded accelerometers in smartphones can be successfully used for measuring the extent of PD hand tremor [24]. Very recently, the utility of mobile phones to quantify PD severity and treatment response has been demonstrated by assessing posture, gait, finger tapping, phonation and response time [25]–[27]. However, previous surveys on smartphone measurements did not consider speech assessment in their battery of tests. Moreover, to the best of our knowledge, the sensitivity of any smartphone-based measures for detecting prodromal stages of PD has never been investigated.

Nevertheless, several important issues have to be resolved before considering detailed research in the field of smartphone-based speech biomarkers for prodromal PD. Although several acoustic measures reliable in capturing speech impairment in early, untreated PD are available [19]–[22], their sensitivity for prodromal stages of neurodegeneration has never been thoroughly investigated. Furthermore, the most widely used methods currently available for the evaluation of speech performance are focused on the assessment of dysphonia via paradigm of sustained phonation [28]. It is unclear whether there is a need to use functional PD-related vocal tasks such as sustained phonation or syllable repetition instead of common spontaneous speech, which would provide a more naturalistic setting concerning speech assessment through a smartphone. There is only a single study showing that detection of timing deficits in spontaneous speech is possible in patients at high risk for developing PD [20]. Finally, it is well known that resulting performance of acoustic features may differ across various recording devices and microphones [29]. However, it has not been thoroughly tested whether available acoustic methods for the evaluation of early PD are independent of high-quality microphone and thus whether they can be applied to smartphone recordings acquired in common environments with a low ambient noise level. Specifically, acoustic features related to timing are expected to be preserved even on recordings collected through a low-quality microphone, while amplitude-based measures are expected to be biased by microphone characteristics, hardware filtering, and compression [30].

Therefore, the aim of this study was to determine the feasibility of detecting speech impairment in participants with RBD and early untreated PD by smartphone. We endeavored

to provide recommendations on the possible use of smartphone-based vocal diagnostic biomarkers of PD by comparing representative acoustic features obtained collaterally through a smartphone and a professional head-mounted condenser microphone with linear frequency response.

II. Methods

A. Participants

A total of 110 consecutive Czech participants were recruited from 2015 to 2017. The study was approved by the Ethics Committee of the General University Hospital, Prague, Czech Republic and all participants provided written, informed consent. Fifty participants (46 men, 4 women), mean age 66 (SD 8) years, were diagnosed with idiopathic RBD according to the International Classification of Sleep Disorders, third edition diagnostic criteria, including confirmation of REM sleep without atonia by polysomnography [31]. In addition, 30 patients (26 men, 4 women), mean age 62 (SD 11) years, with de-novo PD were diagnosed based on the Parkinson's disease Society Bank Criteria [32]. As a control group, 30 healthy participants (26 men, 4 women), mean age 65 (SD 10) years, without a history of neurological or communication disorders were included in the study. Approximately 50 % of included PD, RBD, and healthy control volunteers also participated in a previous study focused on automated analysis of connected speech in patients with RBD [20], however, characteristics of complex speech impairment obtained through smartphones were not previously investigated.

All PD and RBD patients were scored according to the motor score of the Movement Disorder Society–Unified Parkinson's Disease Rating Scale (MDS-UPDRS III, ranging from 0 to 132, with 0 for no motor manifestation and 132 representing severe motor disturbance) [33]. In particular, the MDS-UPDRS III cut-off score of 32/33 points can be considered for mild/moderate levels and cut-off score of 58/59 points for moderate/severe levels [34]. The MDS-UPDRS III scale includes a speech item for the clinical description of speech severity (ranging from 0 to 4, with 0 representing normal speech and 4 indicating unintelligible speech; Table I). Both diagnosis and evaluation of clinical scales were performed by a well-trained professional neurologist with experience in movement disorders. All PD patients were consecutively recruited at their first visit to the clinic and were examined before symptomatic treatment was started. No PD or RBD patient had a history of therapy with antiparkinsonian medication. None of the RBD participants subjectively complained of motor or cognitive difficulties. Clinical evaluation showed slight motor impairment in 10 RBD participants (20 %) with MDS-UPDRS III > 6 after removal of postural and action tremor. This cut-off score > 6 is considered to be a sensitive indicator of initial parkinsonism [35]. Nine RBD participants (18 %) manifested mild cognitive impairment.

TABLE I
Clinical Characteristics of Newly Diagnosed, Untreated PD and RBD Participants

	PD (n = 30)	RBD (n = 50)
Mean Age (years)	62.3 (SD 11.3)	66.4 (SD 7.8)
Mean symptom duration(years)	1.9 (SD 1.3)	5.5 (SD 4.0)
Mean Hoehn & Yahr score	2.1 (SD 0.4)	n/a
Mean MDS-UPDRS III total	29.4 (SD 12.8)	4.9 (SD 4.3)
Mean MDS-UPDRS III speech item	0.60 (SD 0.56)	0.02 (SD 0.14)
Mild motor symptoms (%)	70	100
Moderate motor symptoms (%)	27	0
Severe motor symptoms (%)	3	0
Men (%)	87	90
Positive history of PD in family (%)	20	4
Antiparkinsonian therapy (%)	0	0
Antidepressant therapy (%)	13	18
Clonazepam therapy (%)	0	0
Other benzodiazepine therapy (%)	3	6
RBD presence (%)	20	100

Captions: PD = Parkinson's disease, RBD = rapid eye movement sleep behavior disorder, SD = standard deviation, MDS-UPDRS = Movement Disorder Society-Unified Parkinson's Disease Rating Scale, n/a = not applicable.

B. Speech Examination

Speech recordings were performed in a quiet room with a low ambient noise level (< 50 dB). Data were recorded using professional head-mounted condenser microphone (Beyerdynamic Opus 55, Heilbronn, Germany) placed approximately 5 cm from participant's mouth and attached to a recorder (Edirol R-09HR, Roland, Shizuoka, Japan; hereafter "microphone"). Simultaneously, data were recorded using a smartphone (Xperia Z1 Compact, Sony, Japan; Figure 1) held by the test participant (hereafter "smartphone"). Participants were instructed to position the smartphone close to their ear in the same way as during a regular telephone call. Speech signals were sampled at 48 kHz with 16-bit resolution using both devices. These settings were chosen as they represent the highest values enabled by the smartphone system. We developed the basic recording application for Android 5.1 system that allowed recordings to be sent to a remote server. No unique settings were adjusted and the application thus worked as a default audio recorder.

Each participant was recorded during single session with a speech specialist who guided the respondent through the standardized recording protocol. All participants were instructed to perform three vocal tasks as follows: (i) sustained phonation of the vowel /a/ per one breath for as long and as steadily as possible; (ii) /pa/-/ta/-/ka/ syllable repetition per one breath performed as fast, steadily and accurately as possible; and (iii) monologue for approximately 90 seconds as narration of a freely-chosen fictional story. The vocal tasks of sustained phonation and syllable repetition were repeated twice for every participant per session.

C. Acoustic Features

Based upon three paradigms of sustained phonation, fast syllable repetition and monologue, we tested 11 representative acoustic measures with a well-defined pathophysiological interpretation that have proven be sufficient in the differentiation between newly diagnosed PD and healthy controls (Table II, Figure 2) [19]–[22].

1) Sustained Phonation: Vowel prolongation has become standard task in PD voice research to elicit the presence

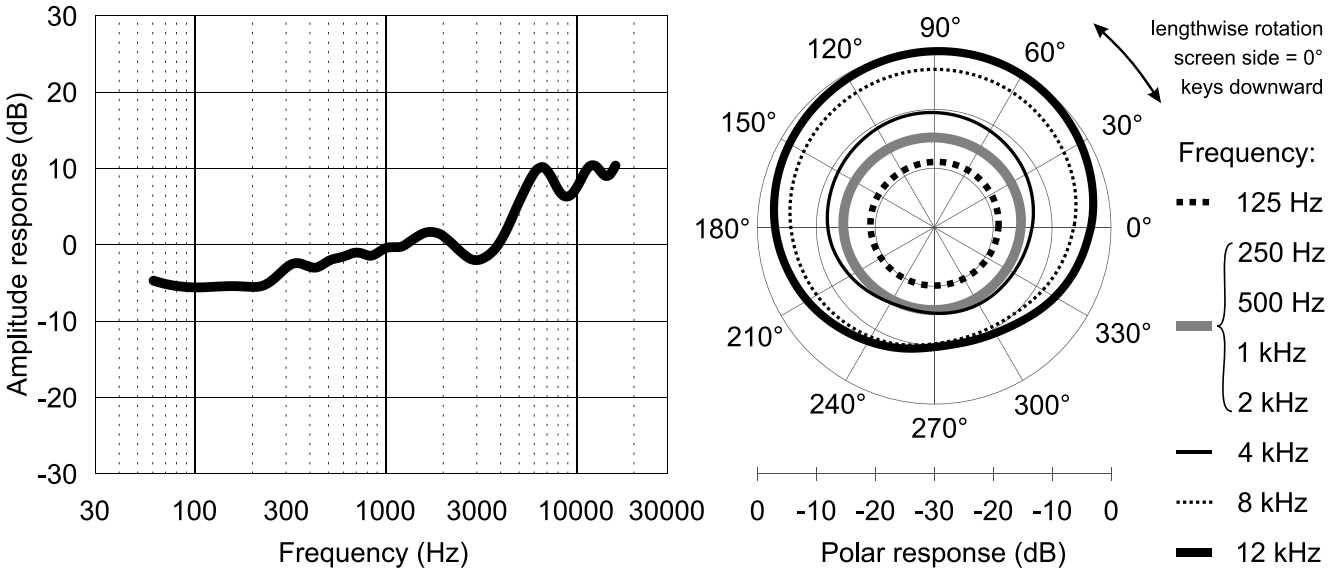


Fig. 1. Amplitude (left) and polar (right) response of Xperia Z1 Compact smartphone measured in anechoic chamber at the distance of 1 m. The recording system showed no proximity effect for the distance of 5 cm. Mean standard error of plotted regression curves is less than 2 dB. Measured self-noise level of the smartphone was 38.8 dB A-weighted.

of dysphonia (i.e. inappropriate vocal fold function), which in hypokinetic dysarthria frequently manifests as a harsh voice [21], [36], [37]. Voice harshness particularly reflects impaired control of stationary pitch during phonation, termed micro-instability of vocal fold vibrations. Moreover, the addition of noise in the speech signal is also typical, indicating incomplete vocal fold closure and turbulent air flow through the vocal folds. The most popular measurements associated with harshness of voice are perturbation measures of jitter, shimmer and harmonics-to-noise ratio (HNR). Jitter represents frequency perturbation, the extent of variation of the voice range. It is defined as the variability of the fundamental frequency from one cycle to the next. Shimmer describes amplitude perturbation, representing roughness of speech. It is defined as the sequence of maximum extent of the signal amplitude within each vocal cycle. HNR is defined as the amount of noise in the speech signal, and represents voice hoarseness. Calculation of these measures was based on an autocorrelation method allowing determination of the frequency and location of each cycle of vibration of the vocal folds, the so-called pitch marks [38]. All measurements of jitter, shimmer, and HNR were calculated using algorithms supplied in the software package Praat [39]. We expected that patients with dysarthria will manifest increased jitter and shimmer values and reduced HNR values.

2) Fast Syllable Repetition: Diadochokinetic (DDK) syllable rates, also known as SMR, are used to evaluate ability to articulate quickly and regularly. While DDK rates capture the inability to perform quick movements, the measurement of DDK regularity allows us to measure the degree of rate variations in a period and assesses the ability of the person to maintain a constant rate of consonant-vowel combinations. Although reduced SMR is more common for spastic dysarthria and irregular SMR for ataxic dysarthria [16], both of these features have been found to be affected in patients with early PD as well as persons at high risk for PD [19]. The DDK rate was measured as the number of syllable vocalizations per second and DDK regularity was defined as the standard deviation of distances between the onsets of following syllables. DDK rate and regularity were detected using a robust algorithm presented by Novotný *et al.* [22]. We expected reduced DDK rates and increased DDK regularity in speakers with PD.

In addition to deficits in SMR, PD patients manifest imprecise coordination between articulatory and phonatory muscles during the DDK task [22]. One widely used method for investigating this imprecision is voice onset time (VOT), which is defined as the interval between the initial articulatory release of a stop consonant and the onset of voicing for the subsequent vowel [22]. Slowing of lip and tongue movements together with difficulties in initiating vocalization in PD leads to prolongation of the VOT interval. VOT was determined as the average length across /p/, /t/, and /k/ consonants extracted from all three /pa/-/ta/-/ka/ syllable repetitions. The individual positions of burst and voicing were detected using a robust algorithm presented by Novotný *et al.* [22].

3) Monologue: Monopitch reflects the reduced ability to perform intonation during speech and can be perceived as reduced melody of speech. Monopitch represents a prominent feature of hypokinetic dysarthria in PD [16], [21]. Monopitch was calculated as the standard deviation of the fundamental frequency (F0 SD), representing the speaker's capability to produce various fundamental frequencies. The pitch sequences were obtained using the software package Praat with the standard autocorrelation-based procedure [39]. The frequency range (minimum and maximum pitch value) was manually adjusted for each recording to avoid pitch doubling and halving; no other intervention was required. To minimize the effect of individual differences between speakers in pitch, the obtained F0 sequences were converted into logarithmic tonal scale (semitones). For example, different pitch ranges such as

TABLE II
OVERVIEW OF APPLIED SPEECH MEASUREMENTS

Deviant speech dimension	Acoustic feature	Definition	Pathophysiological interpretation with respect to hypokinetic dysarthria
Sustained phonation			
Signal perturbations	Jitter	Frequency perturbation, defined as the extent of variation of the voice range, i.e. the variability of the pitch from one cycle to the next.	Deteriorated control of respiratory and laryngeal muscles leads to instable periods of vocal fold opening, causing dysphonic rough speech.
Signal perturbations	Shimmer	Amplitude perturbation, defined as the sequence of maximum extent of the signal amplitude within each vocal cycle.	Deteriorated control of respiratory and laryngeal muscles leads to instable extent of vocal fold opening, causing dysphonic rough speech.
Signal perturbations	HNR	Harmonics-to-noise ratio, defined as the amount of noise in the speech signal.	Reduced rate of airflow and improper control of vocal folds causes increased turbulent noise.
Fast syllable repetition			
Slow sequential motion rates	DDK rate	Diadochokinetic rate, defined as the number of syllable vocalizations per second.	Hypokinesia of speech apparatus makes the movements of articulators slower. Hypokinesia causes deficits in speech timing.
Irregular sequential motion rates	DDK regularity	Diadochokinetic regularity, defined as the standard deviation of distances between following syllable onsets.	Hypokinesia causes deficits in speech timing.
Imprecise consonants	VOT	Voice onset time, defined as the length of the entire consonant from initial burst to vowel onset.	Hypokinesia causes slowing of lip and tongue movements, leading to a longer time required to pronounce individual consonants.
Monologue			
Monopitch	F0 SD	Standard deviation of fundamental frequency contour converted to semitone scale.	Hypokinesia causes reduced amplitude of vocal cord movements, leading to glottal incompetence.
Monoloudness	Int SD	Standard deviation of speech intensity contour extracted from voiced segments.	Hypokinesia leads to decreased amplitude of respiratory and thyroarytenoid muscles.
Articulatory decay	RFA	Resonant frequency attenuation, defined as the differences between the maxima of the second formant region and minima of local valley region called antiformant.	Hypokinesia leads to decrease spectral energy as a result of decayed articulatory movements.
Inappropriate silences	DPI	Duration of pause intervals, defined as the median length of pause intervals.	Hypokinesia of speech apparatus makes initiating of speech difficult, leading to prolonged pause intervals.
Decreased rate of follow-up speech segments	RST	Rate of speech timing, defined as rate of voiced, unvoiced and pause intervals measured as the slope of regression line of total interval count per time. Each interval was described as mean time between onset and offset of interval.	Hypokinetic movements of the speech apparatus lead to reduced stream of voiced,unvoiced and pause intervals.

100-200 Hz or 200-400 Hz were then represented by equal semitone intervals [21]. We expected the values of F0 SD to be reduced as a consequence of hypokinetic dysarthria.

Subsequently, four features were extracted based upon automated robust segmentation, which provides detection of voiced, unvoiced, pause and respiratory intervals [20]. Monoloudness reflects the reduced ability to alternate loudness during connected speech and is a prominent feature of the verbal output of speakers with hypokinetic dysarthria [16], [21]. Monoloudness was computed only on voiced intervals to ensure that the results reflected only respiratory and phonatory control and were not distorted by articulatory deficits. Voiced intervals were extracted from the signal, squared and filtered using a moving average with a widow size of 21.3 ms to obtain a power envelope. The window size of 1024 samples was preferred for convenience in computation; however, any window size from 15 ms to 50 ms can be applied. Monoloudness was computed as the standard deviation of the intensity (Int SD) calculated as power envelope in the logarithmic scale, which provides relative calibration of the signal to a 0 dB intensity level (i.e., the mean value of the power envelope represents 0 dB). We expected reduced Int SD values as a consequence of hypokinetic dysarthria.

Articulatory decay was examined using an acoustic feature of resonant frequency attenuation (RFA) that measures differences between the maxima of the second formant region and minima of the local valley region (so-called antiformant) [19]. Prior to evaluation, only voiced intervals were selected. Subsequently, cepstral liftering was applied on the power spectral density extracted from these voiced speech segments, resulting in the robust detection of local extrema. RFA represents the expressiveness of articulation avoiding dependence on loudness. We expected lower RFA values due to hypokinetic dysarthria.

Duration of pause intervals (DPI) describes the quality of speech timing, as pauses can be heavily influenced by

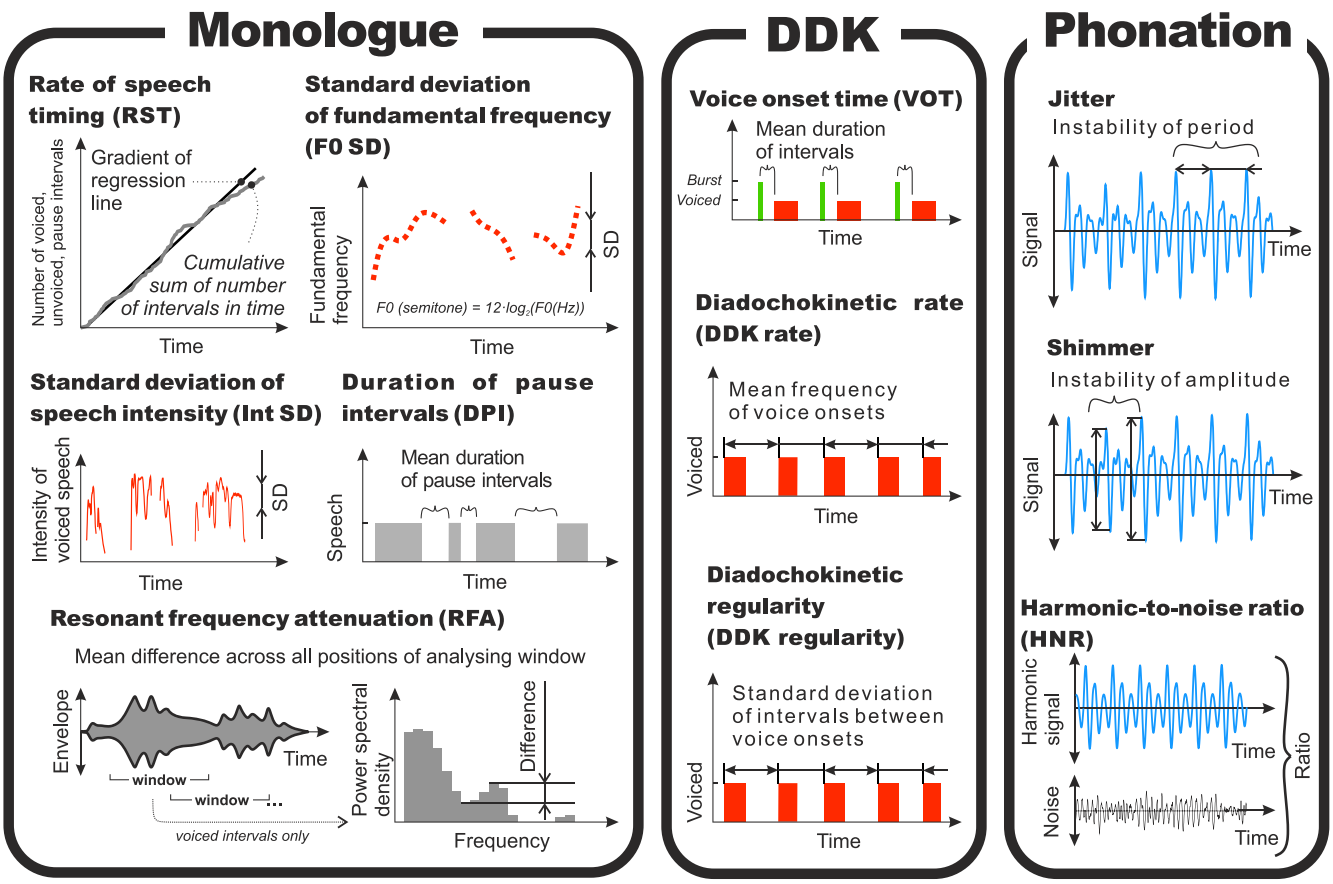


Fig. 2. Graphical illustration of basic principles of individual acoustic speech features. *Captions*: SD = standard deviation, F0 = fundamental frequency.

the ability to properly initiate speech [20]. Complex speech impairment can cause difficulties in initiating speech, which results in prolongation of pauses. DPI was computed as the median duration of all pause intervals.

Rate of speech timing (RST) represents an alternative to pause rate measurement providing more descriptive estimation of speech flow based not on speech/pause segments but voiced, unvoiced, and pause intervals [20]. While voiced intervals reflect the deficits in the phonatory control, unvoiced intervals are influenced by imprecise articulation. Therefore, the rate based upon voiced, unvoiced, and pause intervals reflects overall timing with more complexity, accounting for problems in all dimensions of speech. For further calculation, each voiced, unvoiced, and pause interval was described by the mean time between interval onset and interval end. The total number of intervals was then calculated for every moment over the course of the measurement. Final RST value was obtained as the gradient of the regression line of the time course. We expected reduced RST values as a consequence of hypokinetic dysarthria.

D. Statistical Analyses

In order to provide recommendations on acoustic features that are suitable as smartphone-based vocal diagnostic biomarkers of PD, we performed the following statistical analyses with a pre-selection stage. First, for every participant, the final values of six acoustic features related to sustained phonation and fast syllable repetition were computed by averaging the data obtained in two vocal task runs. As the one-sample Kolmogorov-Smirnov test did not indicate non-normally distributed acoustic features, differences between groups (PD vs. RBD vs. controls) were calculated using one-way analysis of variance with post-hoc Fisher least-squares difference. Second, Pearson correlation was applied to test for significant relationships between pairs of acoustic features applied to recordings obtained from the microphone and smartphone. Third, the magnitude of agreement between recording equipment was measured as the root mean squared error normalized by the mean observed value (NRMSE) [29]. For the final evaluation, we retained only features that showed enough statistical power to separate the investigated groups based on recordings through the microphone (i.e., physiologically sufficient, $p < 0.05$) and in parallel showed very strong correlation and reliability between both recording devices (i.e., technologically sufficient, Pearson's $r > 0.80$ and NRMSE $< 20\%$). The classification performance (sensitivity/specificity) of the retained smartphone-based acoustic features was calculated using binary logistic regression with leave-one-out cross-validation.

TABLE III
OVERVIEW OF RESULTS

Acoustic variable	Microphone			Smartphone			Microphone vs. smartphone		
	PD	RBD	Controls	PD	RBD	Controls	Microphone		
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	ANOVA p	Pearson's r	NRMSE(%)
Sustained phonation									
Jitter (%)	0.69 (0.32)	0.59 (0.24)	0.65 (0.28)	0.73 (0.34)	0.69 (0.29)	0.66 (0.26)	0.28	0.89	21.2
Shimmer (%)	4.71 (1.83)	4.62 (1.81)	4.81 (1.57)	6.55 (2.72)	6.31 (2.40)	6.57 (2.41)	0.89	0.79	32.1
HNR (dB)	19.3 (3.3)	19.8 (3.1)	20.0 (2.9)	15.5 (4.0)	15.9 (3.5)	16.4 (3.5)	0.68	0.92	7.5
Fast syllable repetition									
DDK rate (syll/s)	6.63 (0.84)	6.31 (0.76)	6.64 (0.88)	6.62 (0.83)	6.34 (0.77)	6.61 (0.88)	0.13	0.99	1.6
DDK regularity (ms)	26.0 (11.0)	24.4 (8.1)	24.8 (10.6)	28.4 (10.9)	26.3 (8.3)	28.5 (11.6)	0.77	0.85	21.0
VOT (ms)	24.1 (4.7)	23.3 (4.1)	23.7 (5.5)	24.8 (4.4)	24.1 (4.4)	24.5 (6.6)	0.74	0.82	12.2
Monologue									
F0 SD (st)	2.16 (0.55)	2.55 (0.47)	2.85 (0.73)	2.09 (0.58)	2.47 (0.52)	2.76 (0.75)	< 0.001	0.98	4.7
Int SD (dB)	4.18 (0.97)	4.64 (1.02)	4.45 (0.98)	4.08 (0.87)	4.39 (0.85)	4.37 (0.69)	0.14	0.80	10.9
RFA (dB)	9.0 (1.4)	9.4 (1.8)	10.1 (1.5)	10.6 (1.8)	10.7 (1.7)	10.7 (1.7)	0.04	0.58	15.2
DPI (ms)	284 (113)	240 (83)	208 (48)	275 (115)	236 (66)	202 (48)	0.003	0.90	14.4
RST (-/s)	263 (76)	288 (61)	318 (67)	282 (89)	321 (57)	353 (67)	0.008	0.84	13.7

Captions: HNR = harmonics-to-noise ratio, DDK rate = diadochokinetic rate, DDK regularity = diadochokinetic regularity, VOT = voice onset time, F0 SD = fundamental frequency variability, Int SD = intensity variability, RFA = resonant frequency attenuation, DPI = duration of pause intervals, RST = rate of speech timing, PD = Parkinson's disease, RBD = rapid eye movement sleep behavior disorder, SD = standard deviation, ANOVA = analysis of variance, NRMSE = normalized root mean squared error.

The overall indication of diagnostic accuracy was reported as area under curve (AUC) obtained from operating characteristic curve. Since the RBD group likely represents a mixture of individuals at different stages of prodromal PD, we performed an additional post-hoc analysis in order to examine whether speech abnormalities in the RBD group are associated with an advanced stage of degeneration and high probability of prodromal PD. Therefore, we compared two RBD subgroups defined according to the Movement Disorders Society research criteria for prodromal PD [35]: (i) RBD patients meeting the criteria for probable prodromal PD (hereafter RBD-pPD subgroup) and (ii) asymptomatic RBD patients not meeting the criteria for probable prodromal PD (hereafter RBD-A subgroup). The inclusion into respective RBD subgroups was based upon six factors including age, gender, polysomnography-proven RBD, MDS-UPDRS III score after removal of postural and action tremor, depression requiring antidepressant therapy, and positive family history of PD; please see the study by Berg *et al.* [35] for comprehensive methodic details regarding criteria for the diagnosis of prodromal PD. An independent t-test was applied to compare resulting speech characteristics between RBD subgroups.

III. RESULTS

A. Perceptual Speech Severity

According to the MDS-UPDRS III speech item, 13 PD patients (43 %) and 49 of RBD participants (98 %) perceptually demonstrated normal speech (score of 0), 16 PD patients (53 %) and only 1 RBD participant (2 %) mildly affected speech (score of 1) and 1 PD patient (4 %) moderately affected speech (score of 2).

B. Pre-Selection Stage

Three features related to monopitch, inappropriate silences and decreased rate of follow-up speech segments extracted from monologue were sufficiently sensitive to significantly separate groups based upon data obtained using the professional microphone (F0 SD: $F_{2,107} = 10.9$, $p < 0.001$, $\eta^2 = 0.17$; DPI: $F_{2,107} = 6.1$, $p = 0.003$, $\eta^2 = 0.10$; RST: $F_{2,107} = 5.1$, $p = 0.008$, $\eta^2 = 0.09$) and showed very strong correlation and reliability between the professional microphone and smartphone (F0 SD: $r = 0.98$, $p < 0.001$, NRMSE = 4.7 %; DPI: $r = 0.90$, $p < 0.001$, NRMSE = 14.4 %; RST: $r = 0.84$, $p < 0.001$, NRMSE = 13.7 %) (Table III). In addition, the feature related to articulatory decay was able to statistically differentiate groups (RFA: $F_{2,107} = 3.3$, $p = 0.04$, $\eta^2 = 0.06$) but did not provide enough reliability between both recording devices. Conversely, the feature related to monoloudness demonstrated very strong correlation and reliability between both recording devices (Int SD: $r = 0.80$, $p < 0.001$; NRMSE = 10.9 %) but was not able to statistically differentiate between the investigated groups. Considering the fast syllable repetition task, features representing slow SMR and imprecise consonants showed very strong correlation and reliability between both recording devices (DDK rate: $r = 0.99$, $p < 0.001$; NRMSE = 1.6 %; VOT: $r = 0.82$, $p < 0.001$; NRMSE = 12.2 %) but were not able to statistically separate the investigated groups. In the sustained phonation paradigm, only the feature representing increased noise (HNR) showed very strong correlation and reliability between both recording devices ($r = 0.92$, $p < 0.001$; NRMSE = 7.5 %) but was not able to statistically separate the investigated groups.

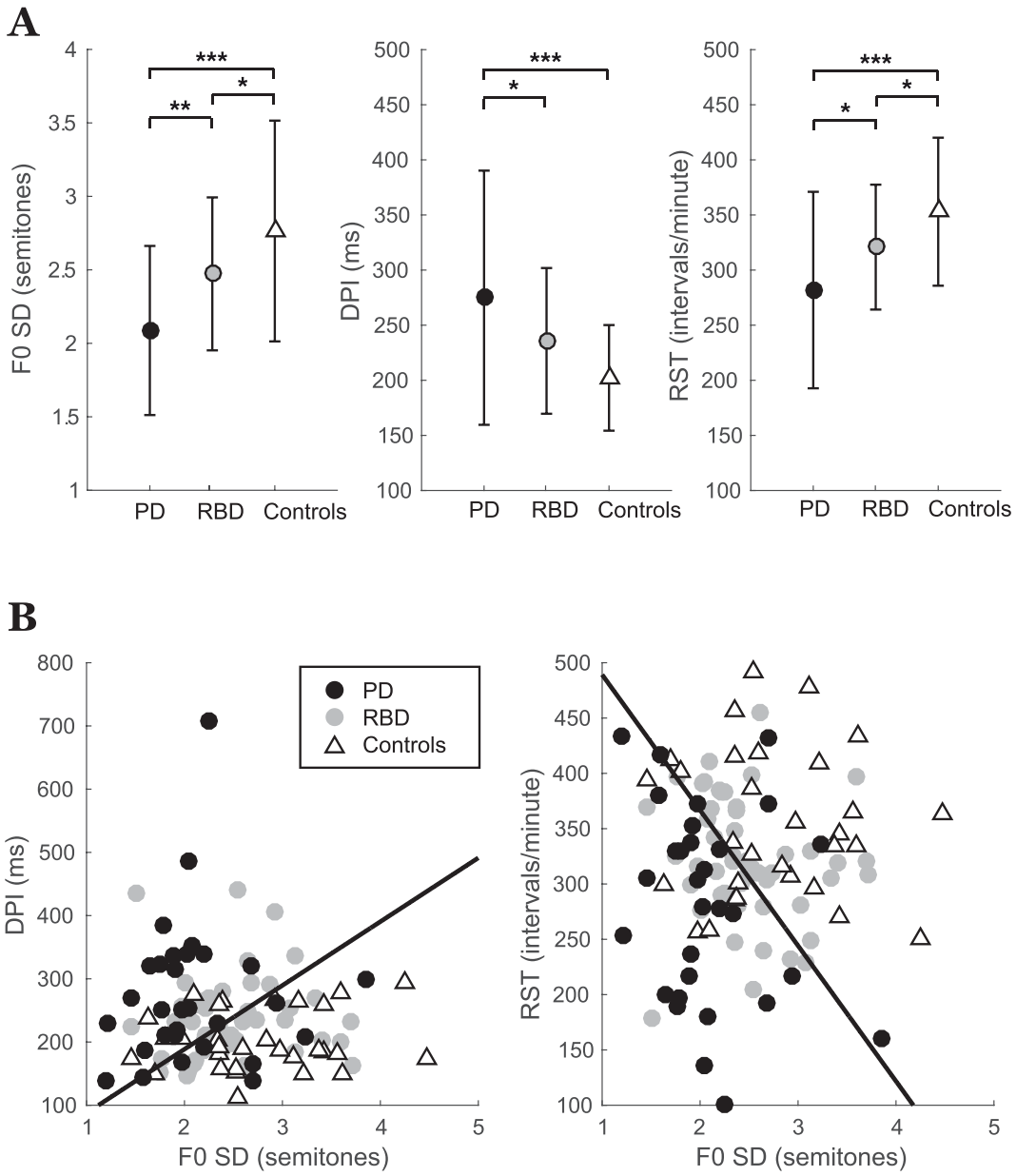


Fig. 3. Results across suitable speech metrics obtained through the smartphone. (A) Group differences between PD, RBD and controls with $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, whereby the symbols represent mean values and error bars represent standard deviation values. (B) Selected pair of smartphone-based speech features plotted in 2D space with optimal decision boundary (black line) between PD and controls. *Captions*: PD = Parkinson's disease, RBD = rapid eye movement sleep behavior disorder, F0 SD = variability of fundamental frequency, DPI = duration of pause intervals, RST = rate of speech timing.

C. Smartphone-Based Evaluation

Features related to monopitch, inappropriate silences and decreased rate of follow-up speech segments recorded through the smartphone significantly separated the investigated groups (F0 SD: $F_{2,107} = 9.5$, $p < 0.001$, $\eta^2 = 0.15$; DPI: $F_{2,107} = 6.4$, $p = 0.002$, $\eta^2 = 0.11$; RST: $F_{2,107} = 7.9$, $p < 0.001$, $\eta^2 = 0.13$) (Figure 3A). Post-hoc analysis detected significant differences in F0 SD between the PD and control groups ($p < 0.001$), between the PD and RBD groups ($p < 0.01$) as well as between the RBD and control groups ($p < 0.05$). Based on DPI, significant post-hoc differences were found between PD and controls ($p < 0.001$) and between PD and RBD ($p < 0.05$). Considering RST, significant post-hoc differences were found between the PD and control groups ($p < 0.001$), between the PD and RBD groups ($p < 0.05$) as well as between the RBD and control groups ($p < 0.05$). A combination of three features (F0 SD, DPI and RST) was able to separate the PD and control groups with AUC 0.85 (sensitivity 75.0%, specificity 78.6%), PD and RBD groups with AUC 0.78 (sensitivity 66.7%, specificity 71.0%) and RBD and control groups with AUC 0.69 (sensitivity 69.8%, specificity 64.7%) (Figure 3B).

D. Comparing RBD Subgroups

As a result of RBD subgroup analysis, 33 RBD patients with the mean age 69 (SD 8) years, mean RBD symptom duration

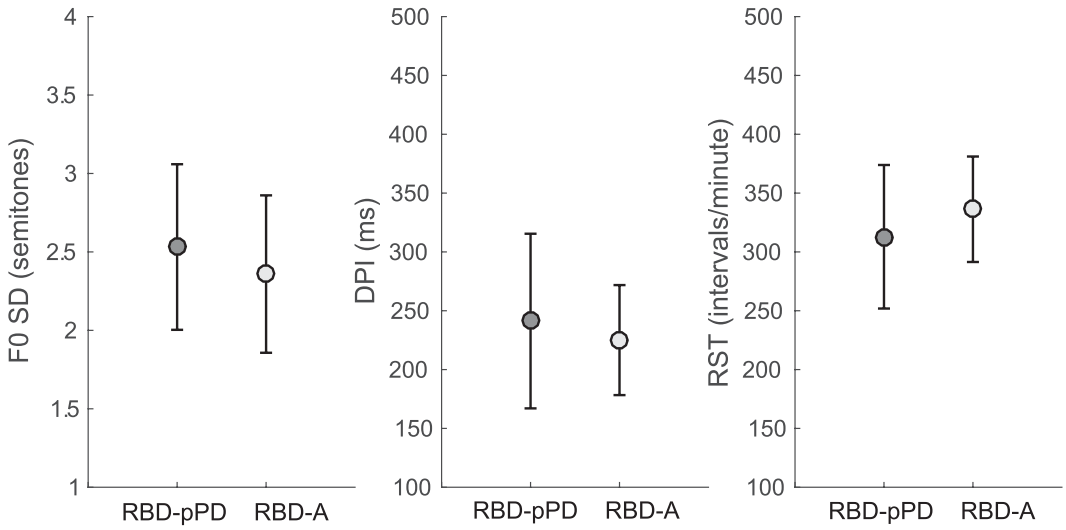


Fig. 4. Results across suitable speech metrics obtained through the smartphone for RBD patients meeting criteria for probable prodromal PD (RBD-pPD) and asymptomatic RBD patients (RBD-A). The symbols represent mean values and error bars represent standard deviation values. No statistically significant differences were found between RBD subgroups.

5.4 (SD 4.5) years and mean MDS-UPDRS III total score 5.8 (SD 4.7) fulfilled the criteria for prodromal PD (RBD-pPD). In addition, 17 RBD patients with the mean age 61 (SD 8) years, mean RBD symptom duration 5.5 (SD 2.9) years, and mean MDS-UPDRS III total score 3.1 (SD 2.4) did not meet the criteria for prodromal PD (RBD-A). None of the features representing monopitch (F0 SD: $t_{48} = -1.1$, $p = 0.27$), inappropriate silences (DPI: $t_{48} = -0.8$, $p = 0.42$) and decreased rate of follow-up speech segments (RST: $t_{48} = 1.4$, $p = 0.17$) recorded through the smartphone significantly separated the RBD-pPD and RBD-A subgroups (Figure 4).

IV. Discussion

Our findings indicate that the detection of speech abnormalities in persons at high risk for PD and related disorders via a smartphone is possible. Smartphones may provide considerable advantages by increasing the number of longitudinal vocal samples from a given individual in an easy and inexpensive way, which is critical not only for researchers performing traditional laboratory-based analyses, but also for pharmaceutical companies developing drugs to treat disorders affecting motor speech performance such as in PD. Based upon a large sample of RBD participants in comparison to early PD and control speakers, we provide recommendation that investigation of monopitch and speech timing is a good starting point to capture preclinical PD via a smartphone. As monopitch, inappropriate silences and decreased rate of follow-up intervals showed sufficient power and reliability to capture prodromal speech abnormalities in this cross-sectional study, they may likely represent biomarkers that appear early enough in the preclinical course of PD. Indeed, we did not find any significant group differences for speech features related to monopitch and speech timing between the asymptomatic RBD subgroup in very early stage of neurodegeneration and the RBD subgroup meeting criteria for prodromal PD. This observation suggests that speech changes may be evaluated to detect approaching conversion to PD within a longer time frame. This assumption appears to be in accordance with previous research showing that vocal and facial akinesia are the first motor signs to develop, with an estimated prodromal interval of 9.8 years before parkinsonism can be clinically diagnosed [6]. However, the predictive value of speech abnormalities needs to be established in further prospective follow-up studies. Observing disease progression over a short time using well-defined and disease-specific biomarkers may significantly help to stratify persons at high risk for developing PD and related disorders for clinical trials, which are crucial for the development of future neuroprotective therapies. The current treatment strategies for PD mainly focus on exercise, various pharmacological approaches, and surgery [40]. Quick, inexpensive, and non-invasive vocal assessment by smartphone may help in the recruitment of appropriate cases into large studies of innovative therapies for prodromal PD and in the future may also bolster early presymptomatic diagnosis of synucleinopathy and enable rapid access to neuroprotective therapy once it will be available.

In general, our findings related to comparison between the professional microphone and smartphone concur with previous studies demonstrating that recordings obtained through smartphones are useful for detecting vocal changes [29], [41]–[44]. With respect to dysphonia measurements, Praat algorithms for the detection of the fundamental frequency and HNR showed good reliability between the professional microphone and smartphone, which is in perfect agreement with previous surveys [29], [44]. This is likely due to the nature of the fundamental frequency and its harmonic frequencies that represent major events in the frequency domain of speech signal and thus can be detected accurately despite the influence of surrounding factors. The same robustness was not observed for perturbation measures such as jitter and shimmer that rely not only on detection of the fundamental frequency but also on dynamic and spectral characteristics of the recording system.

In concordance with our results, previous data has already shown significant differences between perturbation measures obtained through various recording systems [29], [42], [43]. One possible solution for future studies is to apply cepstral peak prominence as the measure of overall irregularity, which has been shown to be more robust and relevant than jitter and shimmer for dysphonia analysis [45]. Considering speech timing, we observed accurate detection of DPI as well as RST, which is partially in contrast to previous research reporting rather unreliable results related to pause characteristics [29]. These discrepancies may be due to the fact that previous research [29] applied a detector for dysarthric speech based upon a single intensity threshold [46], while we used a robust speech-pause detector [20], which has been shown to outperform this former conventional detector by Rosen et al. [46]. Regarding the fast syllable repetition paradigm, we found reliable detection of the DDK rate as well as VOT between the professional microphone and smartphone. However, to the best of our knowledge, the suitability of measures extracted from SMR has never been tested across different devices. Finally, the features related to signal amplitude measured in the dB scale such as RFA were found to be inaccurate. This result was expected and can be explained by the variable position of smartphone during recording, especially when considering the typical hand tremor of PD patients.

A combination of three features representing monopitch, inappropriate silences and decreased rate of follow-up speech segments was able to distinguish between PD and control groups with an AUC of 0.85 and between RBD and control groups with an AUC of 0.69. The findings related to deficits in speech timing in RBD are not surprising as they have been reported recently [20], although the reliability of their assessment through smartphone has never been investigated. However, the reduced intonation in RBD observed in the present work provides new insight into the production of speech in prodromal parkinsonism. This observation is in concordance with a previous study that reported cases with reduced intonation variability detectable several years before the onset of the first PD motor symptoms [47]. Interestingly, only features extracted from monologue were sufficiently sensitive to statistically separate the investigated groups, suggesting that use of functional PD-related vocal tasks such as sustained phonation and syllable repetition are not essential nor optimal to capture prodromal speech impairment. This finding is in agreement with the general assumption that PD speech performance varies across the specific task performed and that analysis of spontaneous utterances is the best way to assess the impact of PD on speech [48], [49]. Conversely, several previous studies have shown reasonable sensitivity in detecting PD-related speech deficits using mainly the paradigm of sustained phonation (see [28], [36], [50]). One possible explanation is the different complexity of the speaking tasks. While sustained phonation may be useful to elicit some aspects of dysphonia, it is not suitable to detect the wide range of articulatory deficits related to hypokinetic dysarthria. Considering that we investigated very early stages of PD and other synucleinopathies with relatively preserved speech performance, a more articulatory demanding speech task than sustained phonation is presumably required to capture subtle changes of speech due to neurodegeneration. Another explanation could be that previous results were based on a large number of speech features including more advanced linear and non-linear methods, which have been demonstrated to work better in PD than standard perturbation algorithms [51]. The functional vocal tasks may also be insensitive due to the relatively older control group in the present study (mean age of 65 years) compared to previous research (mean age of 58 years) [21], [22]. Higher age causes significant performance decline in perturbation measures [52] as well as in maximum repetition rate [53] during functional vocal paradigms, especially in persons above 60 years of age.

Since analysis and processing of speech disorders in PD has become an attractive scientific discipline in recent years, there is a vast number of vocal characteristics and advanced linear and non-linear methods available with a proven efficiency in separating healthy controls from PD (see [50], [54]–[56]). We carefully stratified and selected only 11 features based mainly upon two criteria. The first criterion was hypothesis-driven and considered that the selected speech metric should have well-defined pathophysiological interpretation with respect to PD and was previously shown to be powerful enough in differentiating between early, drug-naïve PD and healthy controls [19]–[22]. The severity of speech disorder in PD increases with disease progression and certain speech deficits develop later in the course of the disease [37], [54]. Accordingly, previous studies have investigated PD speech characteristics mostly in more advanced stages of disease [28], making it difficult to generalize these observations to prodromal or very early PD. In addition, the vast majority of previous studies have examined speech in PD under dopaminergic medication with different doses and delays from drug initiation [28]. Since dopaminergic treatment can significantly affect speech performance [57], it is difficult to conclude whether the observed changes were due to PD itself or to drug effects. The second criterion was technologically-driven and considered that the selected speech metric was practically useful for smartphone analysis. In other words, the selected speech metric should work properly in common environments with a low ambient noise level and can be extracted and analyzed using automatic process with minimum user control [19]–[22]. Consequently, we did not investigate changes in formant frequencies during vowel production, which is one of the core features of dysarthrias [58]. Indeed, vowel space metrics have been shown to reflect reduced articulatory range of motion present already in the early stages of PD [47], [59], suggesting they may discriminate between controls and patients at high risk for developing PD. However, analysis of vowel space metrics in dysarthric speech is currently based on time-consuming manual identification and analysis of selected corner vowels [48], [58], while reliable automated analysis methods have not yet been introduced. Nevertheless, the 11 acoustic features used in the present study sufficiently cover the multidimensional impairement of hypokinetic dysarthria and, importantly, all these metrics

could be relatively easily fine-tuned to allow fully automated smartphone analysis.

There are some limitations to the present study. We recorded all participants using the same Sony Xperia Z1 Compact. However, signal processing differs from smartphone to smartphone and the resulting changes could be more substantial in more sophisticated and more expensive devices [60]. Therefore, we provided experimentally-measured frequency response of testing smartphone and future studies can compare differences between their smartphones and the smartphone used in the present study. The Sony Xperia Z1 Compact was chosen as it can be considered a mainstream product among available smartphones. It was commercially-available and a relatively inexpensive mid-range smartphone in the year 2015. The motivation for choosing a mid-range product was based on the assumption that in the future, due to continuous advances in technology, even low-cost smartphones will provide at least similar microphone characteristics such as this type, allowing screening of the entire population. Additionally, we decided that the participant should hold the phone during the examination rather than placing the phone on the table or having it held by the investigator, as holding the phone currently represents the most common way of calling through smartphones. As the first step of research in this domain, our recordings were performed in a relatively quiet room with a low ambient noise level. To use the proposed methodology in practice, it might be necessary to collect recordings through smartphones in various realistic scenarios (e.g., with background noise in an urban environment), which would likely introduce additional distortion to the recorded signals. In particular, the most promising biomarker discovered in the present study is monopitch and measures based on the fundamental frequency appear to be robust against the recording system, environmental noise, and their combination [61]. Moreover, the currently proposed approach can already be used to monitor speech performance via smartphone in a common home or clinic environment with a low ambient noise level. The possible applicability of smartphones in detecting prodromal speech abnormalities was demonstrated only in the RBD group, whereas future steps for extension of our approach to other populations with high risk of developing PD such as carriers of genes causing PD are warranted. In agreement with the reported prevalence of RBD [62], our dataset is composed particularly of male participants and thus we cannot exclude the possibility that the observed speech patterns may be influenced by gender-specific aspects. The majority of our PD patients did not exhibit RBD symptoms while PD with and without RBD may represent distinct disease phenotypes [63]. In addition, the majority of RBD patients convert to either PD or dementia with Lewy bodies [10]. Nevertheless, the perceptual investigation revealed that dementia with Lewy bodies manifests very similar hypokinetic and monotonic patterns of dysarthria compared to PD [64]. Indeed, values of measures related to monopitch and inappropriate silences in our RBD participants clearly intermediated between those of PD patients and healthy controls, indicating a certain independence of monopitch and inappropriate silences on parkinsonian phenotype.

V. Conclusion

Our work represents the first step toward the use of smartphone technology for the evaluation of prodromal vocal impairment due to neurodegenerative synucleinopathy. Biomarkers collected through a smartphone including speech performance have the potential to revolutionize the diagnostic process in these neurodegenerative diseases and improve stratification for future neuroprotective therapy against PD and related neurodegenerative disorders. Future studies should consider complementing the present methods with other prospective and well-defined PD-specific features and test their reliability in common, realistic environmental scenarios. In addition, longitudinal studies are needed to confirm and further elaborate our findings and to show the sensitivity of speech parameters as potential diagnostic biomarkers. We envisage that our findings will support a continuum of technology solutions in this area and increase the accessibility of voice monitoring for a wider PD and prodromal PD populations.

References

[1] W. Poewe *et al.*, "Parkinson disease," *Nature Rev. Disease Primers*, vol. 23, p. 17013, Mar. 2017, doi: 10.1038/nrdp.2017.13.

[2] M. C. de Rijk *et al.*, "Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic diseases in the elderly research group," *Neurology*, vol. 54, no. 11, pp. S21–S23, 2000.

[3] L. J. Findley, "The economic impact of Parkinson's disease," *Parkinsonism Related Disorders*, vol. 13, pp. S8–S12, Sep. 2007, doi: 10.1016/j.parkreldis.2007.06.003.

[4] M. C. Rodriguez-Oroz *et al.*, "Initial clinical manifestations of Parkinson's disease: Features and pathophysiological mechanisms," *Lancet Neurol.*, vol. 8, no. 12, pp. 1128–1139, 2009, doi: 10.1016/S1474-4422(09)70293-5.

[5] H. Bernheimer, W. Birkmayer, O. Hornykiewicz, K. Jellinger, and F. Seitelberger, "Brain dopamine and the syndromes of Parkinson and Huntington Clinical, morphological and neurochemical correlations," *J. Neurol. Sci.*, vol. 20, no. 4, pp. 415–455, 1973, doi: 10.1016/0022-510X(73)90175-5.

[6] R. B. Postuma, J.-F. Gagnon, J.-A. Bertrand, D. G. Marchand, and J. Y. Montplaisir, "Parkinson risk in idiopathic REM sleep behavior disorder preparing for neuroprotective trials," *Neurology*, vol. 84, no. 11, pp. 1104–1113, 2015, doi: 10.1212/WNL.0000000000001364.

[7] C. H. Schenck *et al.*, "Rapid eye movement sleep behavior disorder: Devising controlled active treatment studies for symptomatic and neuroprotective therapy—A consensus statement from the international rapid eye movement sleep behavior disorder study group," *Sleep Med.*, vol. 14, no. 8, pp. 795–806, 2013, doi: 10.1016/j.sleep.2013.02.016.

[8] B. Högl, A. Stefani, and A. Videnovic, "Idiopathic REM sleep behaviour disorder and neurodegeneration—An update," *Nature Rev. Neurol.*, vol. 14, pp. 40–55, Nov. 2018, doi: 10.1038/nrneurol.2017.157.

[9] B. Boeve, "REM sleep behavior disorder: Updated review of the core features, the REM sleep behavior disorder-neurodegenerative disease association, evolving concepts, controversies, and future directions," *Ann. NY Acad. Sci.*, vol. 1184, no. 2, pp. 15–54, 2010, doi: 10.1111/j.1749-6632.2009.05115.x.

[10] A. Iranzo *et al.*, "Neurodegenerative disorder risk in idiopathic REM sleep behavior disorder: Study in 174 patients," *PLoS ONE*, vol. 9, no. 2, p. e89741, 2014, doi: 10.1371/journal.pone.0089741.

[11] A. Iranzo *et al.*, "Neurodegenerative disease status and post-mortem pathology in idiopathic rapid-eye-movement sleep behaviour disorder: An observational cohort study," *Lancet Neurol.*, vol. 12, no. 5, pp. 443–453, 2013, doi: 10.1371/journal.pone.0089741.

[12] C. H. Schenck, B. F. Boeve, and M. W. Mahowald, "Delayed emergence of a parkinsonian disorder or dementia in 81% of older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder: A 16-year update on a previously reported series," *Sleep Med.*, vol. 14, no. 8, pp. 748–774, 2013, doi: 10.1016/j.sleep.2012.10.009.

[13] R. B. Postuma, A. E. Lang, J. F. Gagnon, A. Pelletier, and J. Y. Montplaisir, "How does parkinsonism start? Prodromal Parkinsonism motor changes in idiopathic REM sleep behaviour disorder," *Brain*, vol. 135, no. 6, pp. 1860–1870, 2012, doi: 10.1016/S1474-4422(13)70056-5.

[14] R. B. Postuma, J. F. Gagnon, S. Rompré, and J. Y. Montplaisir, "Severity of REM atonia loss in idiopathic REM sleep behavior disorder predicts Parkinson disease," *Neurology*, vol. 74, no. 5, pp. 239–244, 2010, doi: 10.1212/WNL.0b013e3181ca0166.

[15] R. B. Postuma *et al.*, "Identifying prodromal Parkinson's disease: Pre-motor disorders in Parkinson's disease," *Movement Disorders*, vol. 27, no. 5, pp. 617–626, 2012, doi: 10.1002/mds.24996.

[16] J. R. Duffy, *Motor Speech Disorders. Substrates, Differential Diagnosis and Management*, 3rd ed. St. Louis, MO, USA: Elsevier Health Sciences, 2013.

[17] A. K. Ho, R. Iansek, C. Marigliani, J. L. Bradshaw, and S. Gates, "Speech impairment in a large sample of patients with Parkinson's disease," *Behav. Neurol.*, vol. 11, no. 3, pp. 131–137, 1999.

[18] J. W. Tetrud, "Preclinical Parkinson's disease: Detection of motor and nonmotor manifestations," *Neurology*, vol. 41, no. 2, pp. 69–71, 1991.

[19] J. Rusz *et al.*, "Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behaviour disorder," *Sleep Med.*, vol. 19, pp. 141–147, Mar. 2016, doi: 10.1016/j.sleep.2015.07.030.

[20] J. Hlavnička, R. Čmejla, T. Tykalová, K. Šonka, E. Růžička, and J. Rusz, "Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder," *Sci. Rep.*, vol. 7, no. 1, p. 12, 2017, doi: 10.1038/s41598-017-00047-5.

[21] J. Rusz, R. Čmejla, and H. Růžičková, and E. Růžička, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease," *J. Acoust. Soc. Amer.*, vol. 129, no. 1, pp. 350–367, 2011, doi: 10.1121/1.3514381.

[22] M. Novotný, J. Rusz, R. Čmejla, and E. Růžička, "Automatic evaluation of articulatory disorders in Parkinson's disease," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 9, pp. 1366–1378, Sep. 2014, doi: 10.1109/TASLP.2014.2329734.

[23] J. R. Orozco-Arroyave *et al.*, "Automatic detection of Parkinson's disease in running speech spoken in three different languages," *J. Acoust. Soc. Amer.*, vol. 139, no. 1, pp. 481–500, 2016, doi: 10.1121/1.4939739.

[24] R. LeMoyne, T. Mastroianni, M. Cozza, C. Coroian, and W. Grundfers, "Implementation of an iPhone for characterizing Parkinson's disease tremor through a wireless accelerometer application," in *Proc. Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2010, pp. 4954–4958, doi: 10.1109/IEMBS.2010.5627240.

[25] S. Arora *et al.*, "Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study," *Parkinsonism Related Disorders*, vol. 21, no. 6, pp. 650–653, 2015, doi: 10.1016/j.parkreldis.2015.02.026.

[26] A. Zhan *et al.*, "Using smartphones and machine learning to quantify Parkinson disease severity: The mobile Parkinson disease score," *JAMA Neurol.*, vol. 75, no. 7, pp. 876–880, 2018, doi: 10.1001/jamaneurol.2018.0809.

[27] F. Lipsmeier *et al.*, "Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial," *Movement Disorders*, to be published, doi: 10.1002/mds.27376.

[28] L. Brabenec, J. Mekyska, Z. Galaz, and I. Rektorová, "Speech disorders in Parkinson's disease: Early diagnostics and effects of medication and brain stimulation," *J. Neural Trans.*, vol. 124, no. 3, pp. 303–334, 2017, doi: 10.1007/s00702-017-1676-0.

[29] A. P. Vogel, K. M. Rosen, A. T. Morgan, and S. Reilly, "Comparability of modern recording devices for speech analysis: Smartphone, landline, laptop, and hard disc recorder," *Folia Phoniatrica Logopaedica*, vol. 66, no. 6, pp. 244–250, 2014, doi: 10.1159/000368227.

[30] C. A. Kardous and P. B. Shaw, "Evaluation of smartphone sound measurement applications," *J. Acoust. Soc. Am.*, vol. 135, no. 4, pp. EL186–EL192, 2014, doi: 10.1121/1.4964639.

[31] American Academy of Sleep Medicine, "International classification of sleep disorders," in *Diagnostic and Coding Manual*, 3rd ed. Westchester, IL, USA: American Academy of Sleep Medicine, 2014.

[32] R. B. Postuma, M. Jin, L. Wang, B. Qin, and K. Wang, "MDS clinical diagnostic criteria for Parkinson's disease," *Movement Disorders*, vol. 30, no. 12, pp. 1591–1601, 2015, doi: 10.1002/mds.26424.

[33] C. G. Goetz *et al.*, "Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, 2008, doi: 10.1002/mds.22340.

[34] P. Martínez-Martin *et al.*, "Parkinson's disease severity levels and MDS-unified Parkinson's disease rating scale," *Parkinsonism Related Disorders*, vol. 21, no. 21, pp. 50–54, 2015, doi: 10.1016/j.parkreldis.2014.10.026.

[35] D. Berg *et al.*, "MDS research criteria for prodromal Parkinson's disease," *Movement Disorders*, vol. 30, no. 12, pp. 1600–1611, 2015, doi: 10.1002/mds.26431.

[36] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed Eng.*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009, doi: 10.1109/TBME.2008.2005954.

[37] S. Skodda, W. Grönheit, N. Mancinelli, and U. Schlegel, "Progression of voice and speech impairment in the course of Parkinson's disease: A longitudinal study," *Parkinsons Disorders*, vol. 2013, no. 3, p. 389195, Dec. 2013, doi: 10.1155/2013/389195.

[38] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proc. Inst. Phonetics Sci.*, vol. 17. Amsterdam, The Netherlands, 1993, pp. 9–110.

[39] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot Int.*, vol. 5, pp. 341–345, Jan. 2002.

[40] S. Sarkar, J. Raymick, and S. Imam, "Neuroprotective and therapeutic strategies against Parkinson's disease: Recent perspectives," *Int. J. Mol. Sci.*, vol. 17, no. 6, p. 904, 2016, doi: 10.3390/ijms17060904.

[41] A. Guidi *et al.*, "Smartphone application for the analysis of prosodic features in running speech with a focus on bipolar disorders: System performance evaluation and case study," *Sensors (Basel)*, vol. 15, pp. 28070–28087, Nov. 2015, doi: 10.3390/s151128070.

[42] E. Lin, J. Hornibrook, and T. Ormond, "Evaluating iPhone recordings for acoustic voice assessment," *Folia Phoniatrica Logopaedica*, vol. 64, no. 3, pp. 122–130, 2012, doi: 10.1159/000335874.

[43] V. Uloza *et al.*, "Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameters and voice pathology screening," *Eur. Arch. Otrohinolaryngol.*, vol. 272, no. 11, pp. 3391–3399, 2015, doi: 10.1007/s00405-015-3708-4.

[44] C. Manfredi *et al.*, "Smartphones offer new opportunities in clinical voice research," *J. Voice*, vol. 31, no. 1, pp. 111.e1–111.e7, 2017, doi: 10.1016/j.jvoice.2015.12.020.

[45] S. N. Awan, N. Roy, D. Zhang, and S. M. Cohen, "Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: Development of clinical cutoff scores," *J. Voice*, vol. 30, no. 2, pp. 130–144, 2016, doi: 10.1016/j.jvoice.2015.04.009.

[46] K. Rosen *et al.*, "Automatic method of pause measurement for normal and dysarthric speech," *Clin. Linguistics Phonetics*, vol. 24, no. 2, pp. 141–154, 2010, doi: 10.3109/02699200903440983.

[47] B. T. Harel, M. S. Cannizzaro, H. Cohen, N. Reilly, and P. J. Snyder, "Acoustic characteristics of Parkinsonian speech: A potential biomarker of early disease progression and treatment," *J. Neurolinguistics*, vol. 17, pp. 439–453, 2004, doi: 10.1016/j.jneuroling.2004.06.001.

[48] J. Rusz, R. Cmejla, and T. Tykalova, "Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task," *J. Acoust. Soc. Amer.*, vol. 134, no. 4, pp. 2171–2181, 2013, doi: 10.1121/1.4816541.

[49] D. Kempler and D. Van Lancker, "Effect of speech task on intelligibility in dysarthria: A case study of Parkinson's disease," *Brain Lang.*, vol. 80, no. 3, pp. 449–464, 2002, doi: 10.1006/brln.2001.2602.

[50] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, May 2012, doi: 10.1109/TBME.2012.2183367.

[51] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," *J. Roy. Soc. Interface*, vol. 8, pp. 842–855, Nov. 2011, doi: 10.1098/rsif.2010.0456.

[52] C. L. Lortie, M. Thibeault, M. J. Guitton, and P. Tremblay, "Effects of age on the amplitude, frequency and perceived quality of voice," *Age*, vol. 37, no. 6, p. 117, 2015, doi: 10.1007/s11357-015-9854-1.

[53] S. Knuijt, J. Kalf, B. Van Engelen, A. Geurts, and B. de Swart, "Reference values of maximum performance tests of speech production," *Int. J. Speech Lang. Pathol.*, vol. 26, pp. 1–9, Sep. 2017, doi: 10.1080/17549507.2017.1380227.

[54] C. E. Stepp, "Relative fundamental frequency during vocal onset and offset in older speakers with and without Parkinson's disease," *J. Acoust. Soc. Amer.*, vol. 133, no. 3, pp. 1637–1643, 2013, doi: 10.1121/1.4776207.

[55] J. R. Williamson, T. F. Quatieri, B. S. Helfer, G. Ciccarelli, and D. D. Mehta, "Segment-dependent dynamics in predicting Parkinson's disease," in *Proc. 16th Annu. Conf. Int.-Speech-Commun.-Assoc. (INTERSPEECH)*, vols. 1–5, 2015, pp. 518–522.

[56] K. De Keyser *et al.*, "The relationship between speech production and speech perception deficits in Parkinson's disease," *J. Speech Lang. Hearing Res.*, vol. 59, pp. 915–931, Jan. 2016, doi: 10.1044/2016_JSLHR-S-15-0197.

[57] S. Pinto, C. Ozsancak, E. Tripoliti, S. Thobois, P. Limousin-Dowsey, and P. Auzou, "Treatments for dysarthria in Parkinson's disease," *Lancet Neurol.*, vol. 3, no. 9, pp. 547–556, 2004, doi: 10.1016/S1474-4422(04)00854-3.

[58] H. Kim, M. Hasegawa-Johnson, and A. Perlman, "Vowel contrast and speech intelligibility in dysarthria," *Folia Phoniatrica Logopaedica*, vol. 63, no. 4, pp. 187–194, 2011, doi: 10.1159/000318881.

[59] S. Skodda, W. Visser, and U. Schlegel, "Vowel articulation in Parkinson's disease," *J. Voice*, vol. 25, no. 4, pp. 467–472, 2011, doi: 10.1016/j.jvoice.2010.01.009.

[60] J. Lebacq, J. Schoentgen, G. Cantarella, F. T. Bruss, C. Manfredi, and P. DeJonckere, "Maximal ambient noise levels and type of voice material required for valid use of smartphones in clinical voice research," *J. Voice*, vol. 31, no. 5, pp. 550–556, 2017, doi: 10.1016/j.jvoice.2017.02.017.

[61] Y. Maryn, F. Ysenbaert, A. Zarowski, and R. Vanspauwen, "Mobile communication devices, ambient noise, and acoustic voice measures," *J. Voice*, vol. 31, no. 2, pp. 48.e11–248.e23, 2017, doi: 10.1016/j.jvoice.2016.07.023.

[62] L. M. Trotti, "REM sleep behaviour disorder in older individuals: Epidemiology, pathophysiology, and management," *Drugs Aging*, vol. 27, no. 6, pp. 457–470, 2010, doi: 10.2165/11536260-000000000-00000.

[63] S. R. Romenets *et al.*, "Rapid eye movement sleep behavior disorder and subtypes of Parkinson's disease," *Movement Disorders*, vol. 27, no. 8, pp. 996–1003, 2012, doi: 10.1002/mds.25086.

[64] J. Muller *et al.*, "Progression of dysarthria and dysphagia in postmortem-confirmed parkinsonian disorders," *Arch. Neurol.*, vol. 58, no. 2, pp. 259–264, 2001.