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.

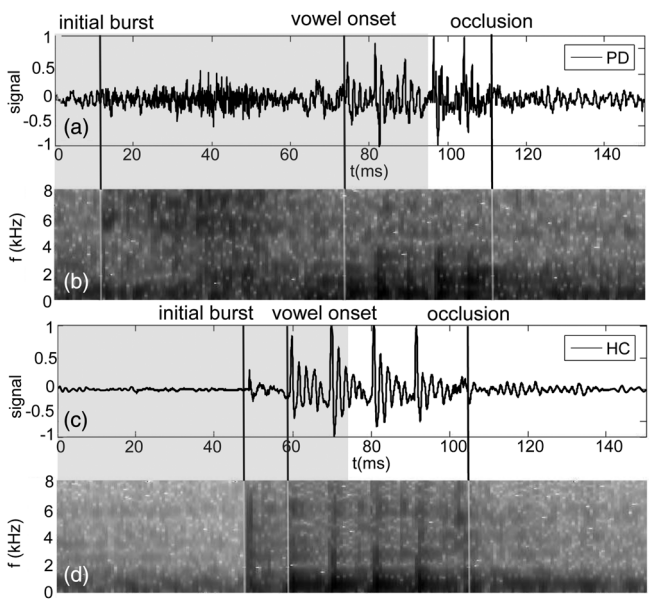

Fig. 1. Examples of syllable and its wideband spectrogram for Parkinsonian (a), (b) and healthy (c), (d) speakers, with marked positions of the initial burst, vowel onset and occlusion. The gray background shows the front part of the syllable and the white background refers to the rear part of the syllable.

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

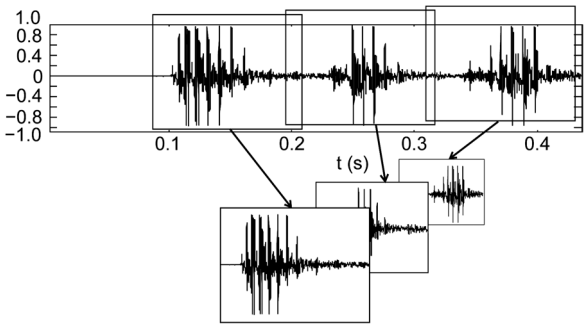

Fig. 2. Detail of an utterance divided by rough segmentation into single syllabic segments.

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.

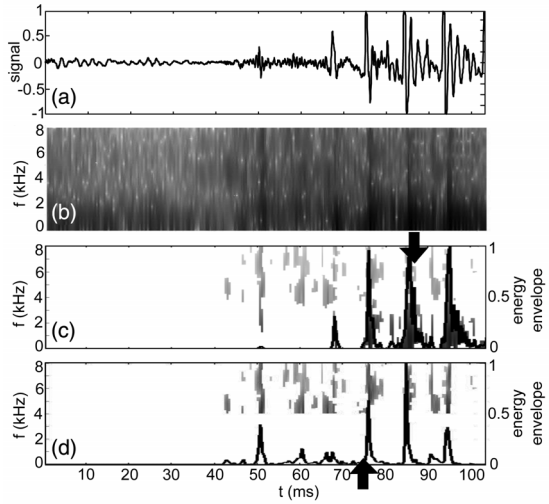

Fig. 3. Signal in the time domain (a), signal spectrogram (b), filtered spectrograms with marked energy envelopes and arrows pointing to spectral centroid positions in the entire frequency range (c), and the upper half of the frequency range (d).

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

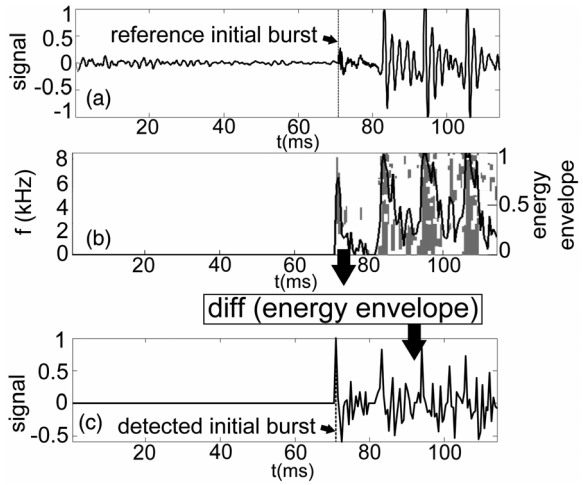

Fig. 4. Fig. 4 Front part of a syllable in the time domain (a), filtered spectrogram with the gray color denoting 1 and the white color denoting 0 and its marked energy envelope (b), and the normalized difference of the energy envelope used for the final initial burst detection (c).

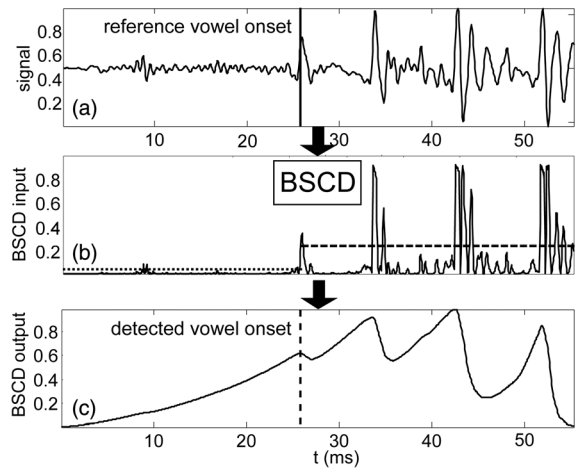

Fig. 5. Original signal with marked reference position (a), input of the BSCD detector represented by the squared original signal and marked BSCD steps (b), and output of the BSCD detector with marked positions of reference and detected V position (c).

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

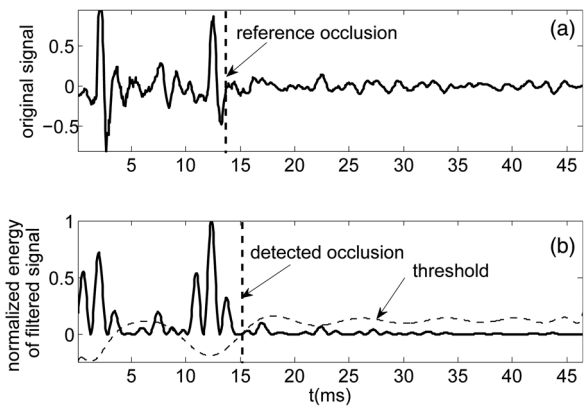

Fig. 6. Rear part of a syllable with marked reference position O (a), and energy of the filtered signal with polynomial threshold and marked position O (b).

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.

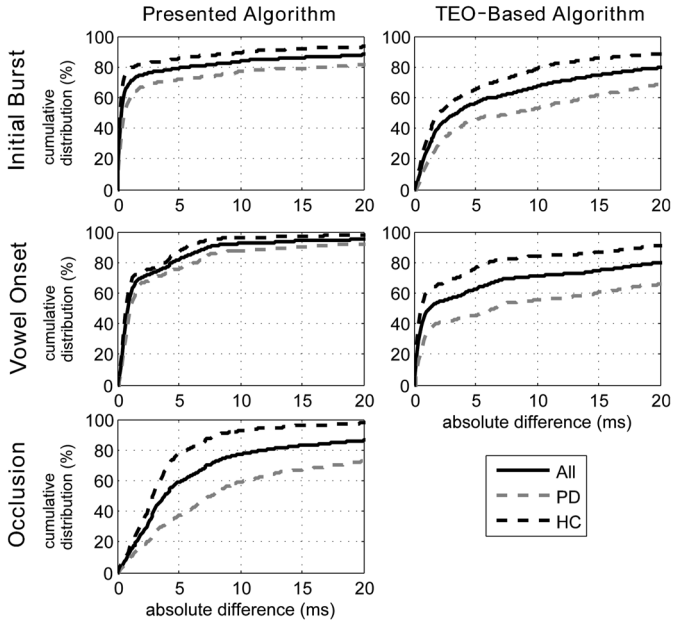

Fig. 7. Cumulative distributions of algorithm performance based on the absolute difference between automatic detection and reference labels. Performances are estimated separately for syllables in Parkinson's disease subjects (PD) and healthy controls (HC), as well as for all syllables together (All). 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

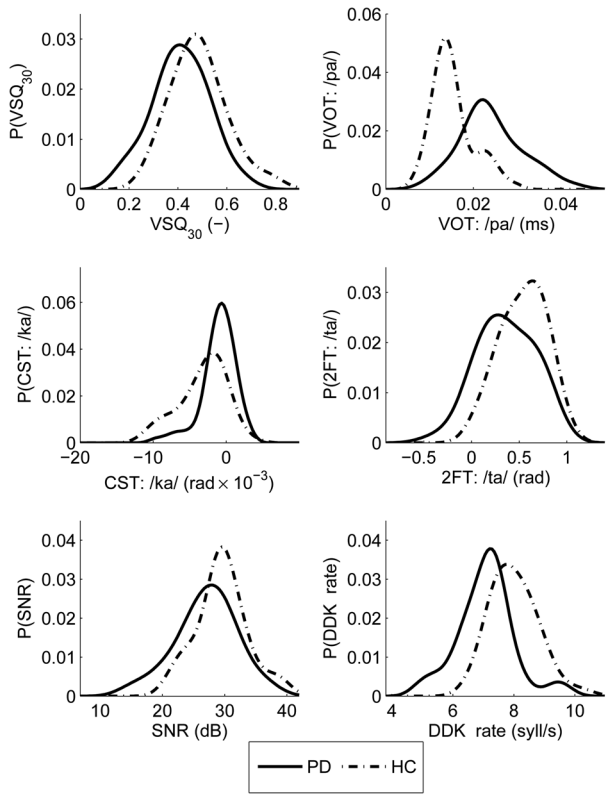

Fig. 8. Probability densities of six representative features with the best SVM classification performance. The vertical axes are the probability densities P(measure) of feature values estimated using the Gaussian kernel density method. The dashdot lines represent the HC group and solid lines the PD group.

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.