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

 0001-4966/2011/129(1)/350/18/$30.00

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.

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

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 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.

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.

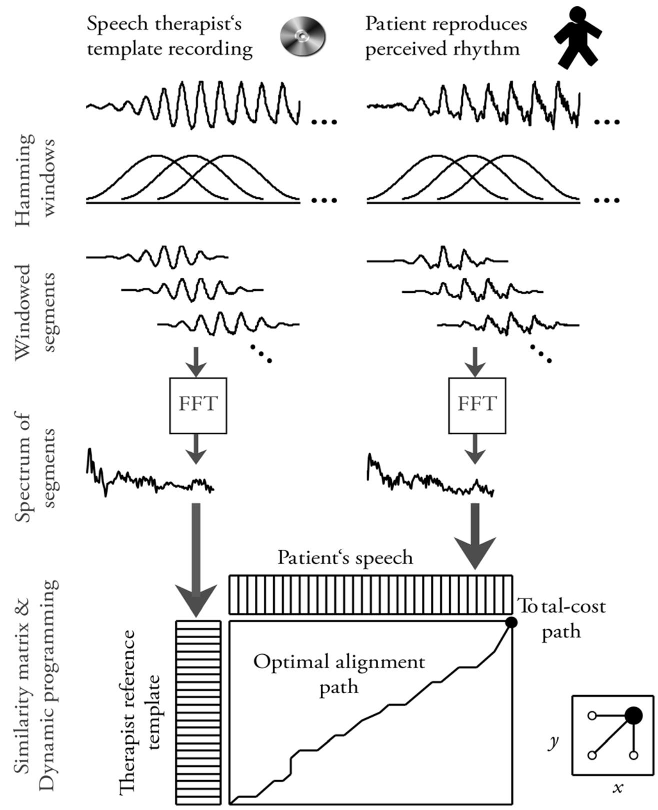

FIG. 1. The alignment process using DTW for measurement of ability to reproduce perceived rhythm. The resulting alignment path may be visualized as a low valley of Euclidean distance scores between speech spectrum segments of patients and speech therapist template recording, meandering through the hilly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

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

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

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]

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

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

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]

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

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 $\omega(1) \le \varepsilon$ and $\omega(2) \le \varepsilon$ are not contradictory for an arbitrary non-negative

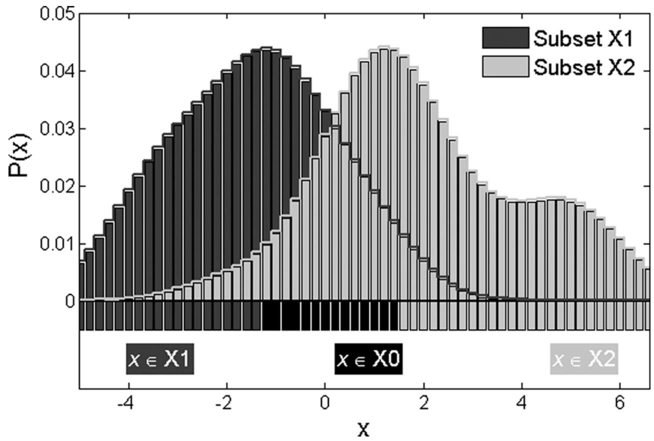

FIG. 2. The top part of the figure shows a selected example of probability densities. The bottom of the figure shows the result of Wald task classification. As a result, the dark-gray shaded bars are the regions in which the feature x assumes the value from the subset $X1$ and the light shaded bars are 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 PD Mean	PD SD	HC Mean	HC SD	Redundant to measurement?	Difference between groups	Retained for Wald task classification?
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)	$\boldsymbol{p < 0.01}$	No
03. Jitter:RAP (%)	0.88	0.81	0.38	0.52	Jitter(local,PPQ5,DDP)	$\boldsymbol{p < 0.01}$	No
04. Jitter:PPQ5 (%)	0.83	0.75	0.32	0.32	Jitter(local,RAP,DDP)	$\boldsymbol{p < 0.01}$	**Yes**
05. Jitter:DDP (%)	2.65	2.42	1.14	1.56	Jitter(local,RAP,PPQ5)	$\boldsymbol{p < 0.01}$	No
06. Shimmer:local (%)	7.51	4.97	2.72	2.27	Shimmer(APQ[3,5,11],DDA)	$\boldsymbol{p < 0.001}$	**Yes**
07. Shimmer:APQ3 (%)	3.69	2.57	1.39	1.36	Shimmer(local,APQ5,DDA)	$\boldsymbol{p < 0.001}$	No
08. Shimmer:APQ5 (%)	4.37	3.07	1.45	1.13	Shimmer(local,APQ[3,11],DDA)	$\boldsymbol{p < 0.001}$	No
09. Shimmer:APQ11 (%)	6.32	3.85	2.20	1.64	Shimmer(local,APQ5)	$\boldsymbol{p < 0.001}$	No
10. Shimmer:DDA (%)	11.07	7.71	4.17	4.07	Shimmer(local,APQ[3,5])	$\boldsymbol{p < 0.001}$	No
11. NHR (-)	0.16	0.27	0.02	0.04	No	$\boldsymbol{p < 0.01}$	**Yes**
12. HNR (dB)	16.01	7.36	24.02	5.61	No	$\boldsymbol{p < 0.001}$	**Yes**
Articulation							
[TASK 2] DDK task							
13. DDK rate (syll/s)	6.22	0.63	7.16	0.73	No	$\boldsymbol{p < 0.001}$	**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	$\boldsymbol{p < 0.001}$	**Yes**
16. RRIS (dB/s)	2.75	1.51	1.16	1.12	No	$\boldsymbol{p < 0.001}$	**Yes**
17. RFPC (-)	0.46	0.17	0.60	0.09	No	$\boldsymbol{p < 0.01}$	**Yes**
18. SDCV (-)	0.14	0.03	0.18	0.03	No	$\boldsymbol{p < 0.001}$	**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	$\boldsymbol{p < 0.001}$	**Yes**
21. Intensity SD (dB)	5.93	1.05	7.55	1.62	No	$\boldsymbol{p < 0.001}$	**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	$\boldsymbol{p < 0.01}$	**Yes**
[TASK 5] Monolog							
25. F0 SD (semitones)	1.53	0.32	2.44	0.65	No	$\boldsymbol{p < 0.001}$	**Yes**
26. Intensity SD (dB)	7.05	1.41	8.75	1.51	No	$\boldsymbol{p < 0.001}$	**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	$\boldsymbol{p < 0.01}$	**Yes**
[TASK 6] Stress patterns							
29. F0 SD (semitones)	2.06	0.81	2.78	0.62	No	$\boldsymbol{p < 0.01}$	**Yes**
30. Intensity SD (dB)	6.40	1.07	7.84	1.97	No	$\boldsymbol{p < 0.01}$	**Yes**
[TASK 7] Emotional sentences							
31. F0 SD (semitones)	2.59	0.74	3.82	0.56	No	$\boldsymbol{p < 0.001}$	**Yes**
[TASK 8] Rhythmic text							
32. Rhythm (-)	2.65	0.55	2.27	0.28	No	$\boldsymbol{p < 0.01}$	**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

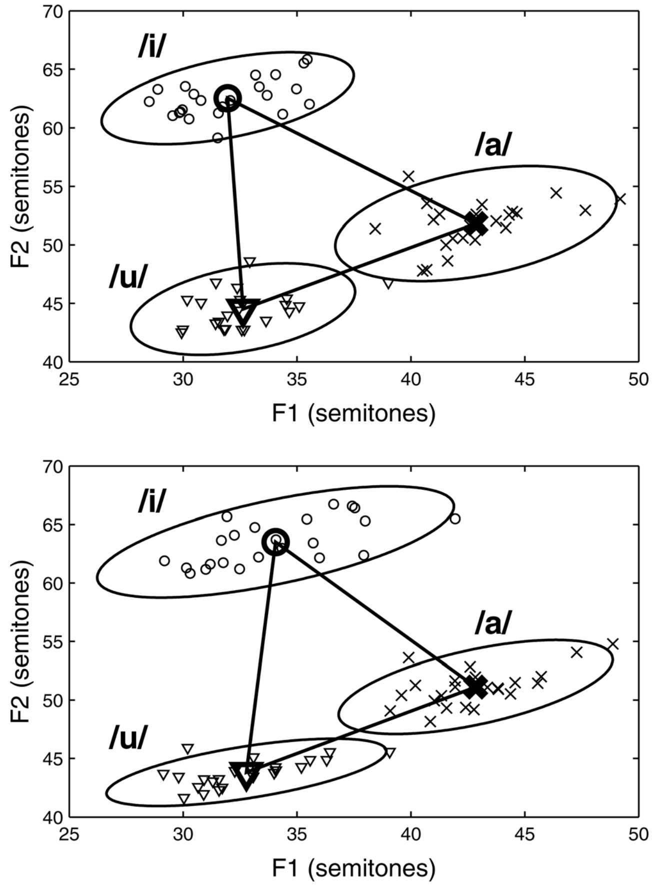

FIG. 3. The vowel triangle space areas for subjects with PD (up) and HC subjects (down).

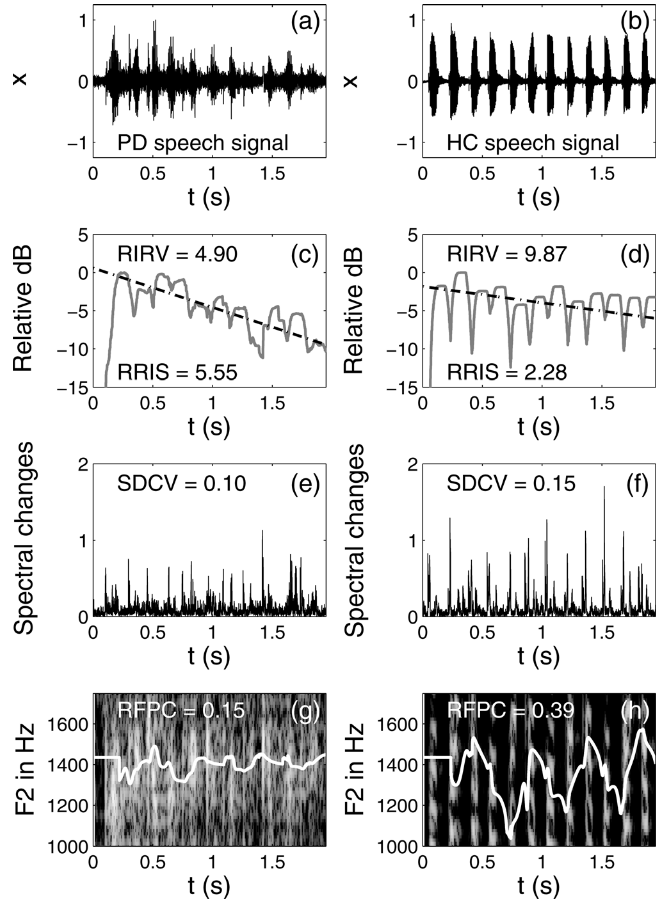

FIG. 4. Details of new articulation measures performed on rapid steady syllable repetition. (a and b) Speech signals of rapid steady /pa/-/ta/-/ka/ syllables repetition; (c and d) light gray lines represent obtained intensity sequences, dashdot lines represent the RRIS. The RIRV is computed as SD from 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

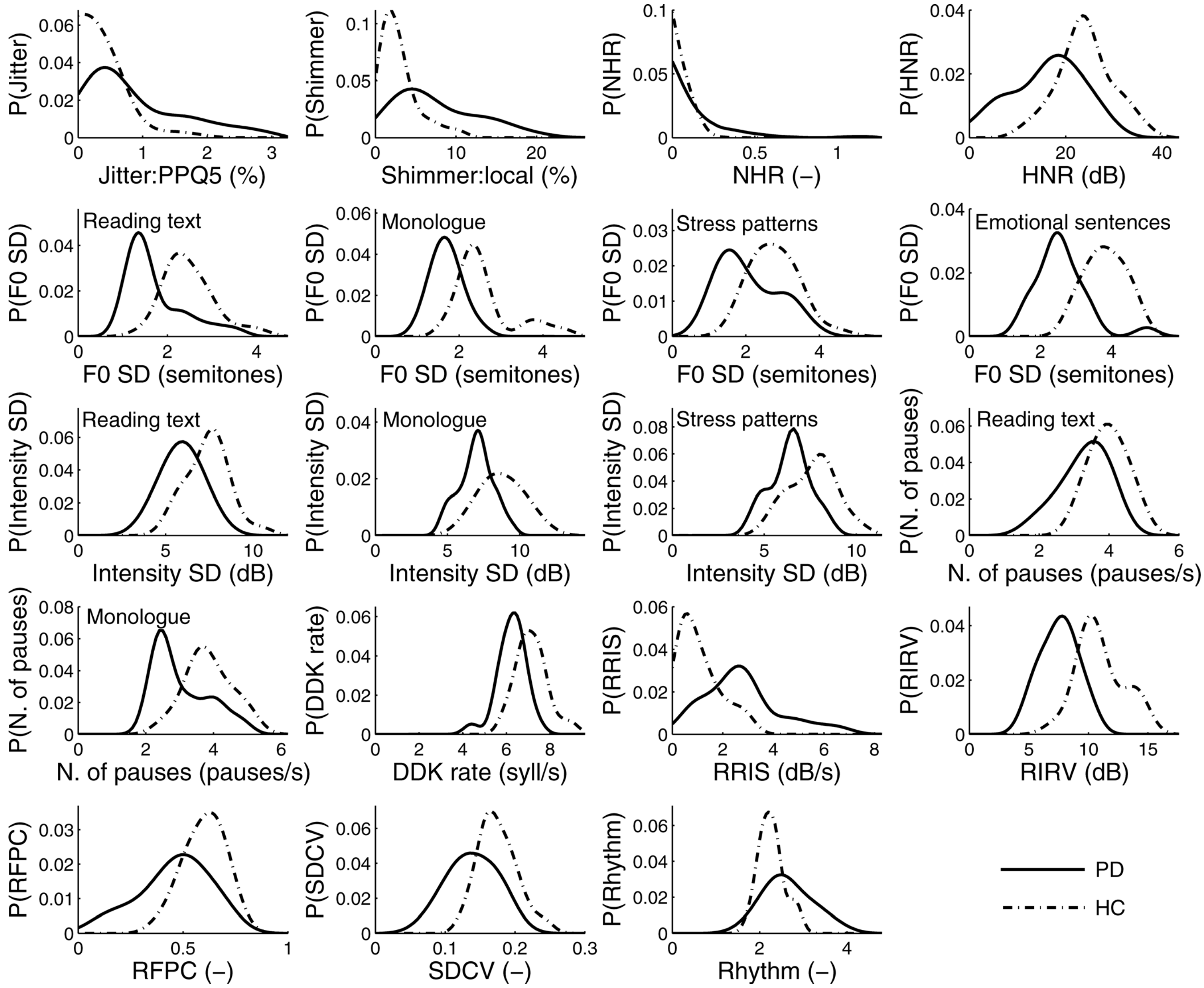

FIG. 5. Selected probability densities of all representative measures (features) in preparation for Wald task classification. The vertical axes are the probability densities P("measure") of the normalized features values of each measure. The dashdot lines are for HC speakers, the solid lines for PD subjects.

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	$\sum$ 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**
$\sum$ |Ratings|	**5**	**14**	**14**	**11**	**12**	**23**	**7**	**9**	**17**	**11**	**28**	**5**	**28**	**13**	**10**	**7**	**12**	**6**	**13**			

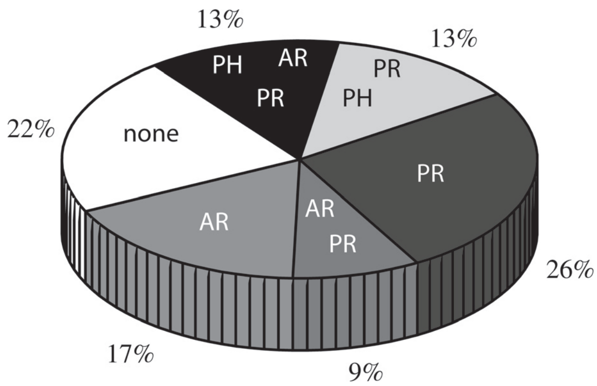

FIG. 6. Details of affected speech dimensions in PD: PH, phonation; AR, articulation; PR, prosody.

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.