A Diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease

David Montaña[a], Yolanda Campos-Roca[a], Carlos J. Pérez[b,*]

[a]*Department of Computer and Communication Technologies, Universidad de Extremadura, Avda. de la Universidad s/n, 10003 Cáceres, Spain.*
[b]*Department of Mathematics, Universidad de Extremadura, Avda. de Elvas s/n, 06006 Badajoz, Spain.*

Abstract

Background and Objective: A new expert system is proposed to discriminate healthy people from people with Parkinson's Disease (PD) in early stages by using Diadochokinesis tests.
Methods: The system is based on temporal and spectral features extracted from the Voice Onset Time (VOT) segments of /ka/ syllables, whose boundaries are delimited by a novel algorithm. For comparison purposes, the approach is applied also to /pa/ and /ta/ syllables. In order to develop and validate the system, a voice recording database composed of 27 individuals diagnosed with PD and 27 healthy controls has been collected. This database reflects an average disease stage of 1.85 ± 0.55 according to Hoehn and Yahr scale. System design is based on feature extraction, feature selection and Support Vector Machine learning.
Results: The novel VOT algorithm, based on a simple and computationally efficient approach, demonstrates accurate estimation of VOT boundaries on /ka/ syllables for both healthy and PD-affected speakers. The PD detection approach based on /k/ plosive consonant achieves the highest discrimination capability (92.2% using 10-fold cross-validation and 94.4% in the case of leave-one-out method) in comparison to the corresponding versions based on the other two plosives (/p/ and /t/).

[*] Corresponding author
Email addresses: dmongut@unex.es (David Montaña), ycampos@unex.es (Yolanda Campos-Roca), carper@unex.es (Carlos J. Pérez)

Conclusion: A high accuracy has been obtained on a database with a lower average disease stage than previous articulatory databases presented in the literature.

Keywords: Expert system; Acoustic features; Classification; Diadochokinesis (DDK); Parkinson's disease (PD); Speech disorders.

1. Introduction

Expert systems have been used in almost every field of medicine, being diagnosis the dominant decision-making issue. A survey on expert systems for diagnosis support in the field of neurology is presented in [1]. The key element of an expert system is the knowledge base. Complex areas in medicine require extensive knowledge that may be extracted from clinical datasets [2, 3, 4]. The primary aim of this research is to design an expert system for early detection of Parkinson's disease (PD).

PD is the second most common neurodegenerative disorder after Alzheimer's disease. According to the Parkinson's Disease Foundation, an estimated 7 to 10 million people worldwide are living with this medical condition. This disease is a chronic neurodegenerative disorder caused by the progressive degeneration and death of dopaminergic neurons, that play a key role in coordinating the movement at level of muscle tone.

Voice and speech, as dependent on laryngeal, respiratory and articulatory functions, may also be affected in patients with PD [5]. Acoustic analysis on recorded speech signals can help to detect subtle abnormalities in speech that may not be perceptible to listeners [6]. Some authors have considered measures extracted from speech recordings and machine learning techniques to discriminate healthy people from those with PD [7, 8, 9, 10, 11, 12]. These techniques have a great potential to establish efficient biomarkers that may help neurologists in their diagnoses or allow primary care physicians to refer the patient to a neurology unit.

The medical literature describes numerous advantages that may be associated with early intervention in PD [13]. Besides medical treatment, PD patients should have access to other type of services including physiotherapy or speech and language therapy. Therefore, successfully addressing early diagnoses of people with PD is a key issue to improve the patients' quality of life. However, it is estimated that 20% of people with PD remain undiagnosed [14].

An important drawback to test the effectiveness of a detection methodology in an early-stage scenario is the scarcity of data. Therefore, one of the primary goals of this investigation was to build a new voice recording database. In order to check a system for early diagnosis of PD, patients between 1 and maximum 2.5 of the Hoehn and Yahr (H&Y) scale have been considered [15]. This corresponds to cases from very mild to mild stages. To the best of the authors' knowledge, the experimental results shown have been obtained on a database with lower average disease stage (1.85 ± 0.55 according to H&Y scale) in comparison to previous articulatory datasets reported in the literature.

Note that the goal of this work is not to track progression of concrete symptoms or search for a correlation between speech impairment and H&Y stage, but to predict disease presence in an automatic way as early as possible. Stage 1 (H&Y) represents the earliest definable stage of PD with current diagnostic. So if a computer-aided system is able to predict that the disease is present (even in the case of patients at low H&Y stages), this means that this system is able to detect the disease in an automatic way, without the need of a neurologist, with the consequent cost saving. Once the disease is detected by a general practitioner, the patient should be forwarded to the neurology unit for further tracking of the disease.

There are several speaking tasks that could be used to evaluate voice disorders in PD based on the extraction of phonation, articulation or prosody features. Most of the approaches for automatic detection of PD from speech are based on sustained vowel phonations, where the speaker attempts to produce a vowel sound as steady as possible in terms of amplitude and fundamental frequency. This type of vocal task enables the measurement of dysphonic aspects of speech. By using features extracted from sustained vowels, accuracy rates between 73.5% to 100% have been reported depending on the feature selection and classification techniques based on the dataset provided in [7] (see [11]). However, it is important to note that these approaches lead to overoptimistic estimations of the accuracy rates, since they are based on replicated measurements (6 voice recordings per subject) and the dependent nature of the observations has not been taken into account [16, 17]. [12] demonstrate the first classification approach for PD detection that takes into account the underlying within-subject dependence of the replicated recordings by using the dataset provided in [7]. In this case the accuracy rate was 90.4%. In spite of this, it is necessary to highlight the key role that this dataset has played in the development of this research line, allowing the investigation with new linear and nonlinear features.

Articulatory difficulties represent an important manifestation of speech disorders in PD. This fact motivates the search for features extracted from speaking tasks involving quick movements of the articulators. Diadochokinesis (DDK) tests are one of the most common tools to evaluate articulatory impairments in both research and clinical assessment contexts. DDK tasks typically measure the subject's ability to repeat a consonant−vowel (C-V) combination with bilabial, alveolar, and velar places of articulation, quickly and at a rhythmic timing. Subjects are asked to repeat a combination of the three-syllable train, for example, /pa/-/ta/-/ka/, as fast as possible.

In [18], the authors presented an approach to discriminate PD from healthy controls (HCs) based on features extracted from DDK utterances. This study was based on 24 individuals diagnosed with PD and 22 HCs, all of them Czech native speakers. The task was repeated twice per speaker resulting in the acquisition of 80 utterances in total. The within-subject dependence of the utterances was not taken into account. The approach was based on 13 features representing six different articulatory aspects of speech: vowel quality, coordination of laryngeal and supralaryngeal activity, precision of consonant articulation, tongue movement, occlusion weakening, and speech timing. The authors achieved best success rates of 87.1% (using 10-fold cross-validation) and 88.4% (with the leave-one-out (LOO) method) on a database reflecting disease stages of 2.2 ± 0.5 (H&Y scale).

In [17], the authors performed discrimination of PD from HCs based on phonation, articulation and prosody, by using different speaking tasks (including rapid syllable repetition). Three different languages (Spanish, German and Czech) were considered. The reported accuracy was 99% for Spanish in the case of features extracted from the unvoiced segments of DDK utterances. These segments were modeled by using 12 Mel Frequency Cepstrum Coefficients (MFCCs) and the energy measured over 25 bands based on the Bark scale. However, it is necessary to remark that the patients had a mean H&Y stage of 2.3 ± 0.8, including patients in advanced stages. In the case of Czech language, where the mean stage is slightly lower 2.2 ± 0.5, the reported accuracy when using the unvoiced segments is reduced to 93.1%. In general, the lower the disease stage, the more challenging the diagnosis task is, since the speech impairment is less severe.

DDK utterances cover three types of syllables (/pa/, /ta/ and /ka/), composed of two regions with completely different characteristics (plosive consonant and vowel segments). Some previous studies in clinical assessment contexts have shown irregular articulation of velar stops by speakers with PD. In [19], the authors report imprecise velar contact in PD, after an investigation based on real-time dynamic magnetic resonance imaging. [20] point out that velopharyngeal control may be impaired in PD. In [21], the authors indicate that syllable /ka/ is more impaired than /pa/ and /ta/. This allows to hypothesize that automatic diagnosis based on /k/ segments should provide better performance than in the case of /p/ or /t/ segments. However, to the best of the authors' knowledge, a validation of this hypothesis in an automatic detection scenario has not been reported. In [17] and [18], the authors extract features from different unvoiced segments (/p/, /t/ and /k/), but there is no distinction between the three plosives, that is, it is not considered which one would perform best in the development of a classification approach based on a single plosive. The focus only on one type of syllable saves computational effort since it avoids feature extraction tasks on the other types of syllables. Here three different classification experiments on the three plosive segments have been performed and the results are comparatively analyzed for an early-stage PD database.

Voice Onset Time (VOT) is defined as the duration of the part of the syllable (/pa/, /ta/ or /ka/, in this work) between initial burst and vowel onset. Since all the acoustic features are extracted from VOT segments, an accurate estimation of VOT is necessary. This accuracy must be also guaranteed in the case of dysarthric speech. A simple and intuitive algorithm is proposed, that provides accurate results both for healthy and dysarthric speakers.

Using the proposed segmentation algorithm, several features have been considered that may be sensitive to possible articulatory deficits due to dysarthria. Both temporal and spectral features based on VOT segments have been considered. The former group is based on time durations [6, 18], whereas the latter one includes MFCC-based features [22] and spectral moments [23]. In a different context, but also related to plosive consonants, [24] demonstrate accurate discrimination of articulation place based on a combination of different types of features, including temporal and spectral ones, and in particular MFCC-based features.

Spectral moments measure the shape of the energy distribution in the spectrum. In the context of PD, spectral moments have been applied to the analysis of long-time average spectra from a standard reading sample in [25] and have been used to describe fricatives occurring in the word initial position of a reading passage in [26]. However, to the best of the authors' knowledge, the use of spectral moments extracted from voiceless stop consonants for PD detection has not been reported yet.

The outline of this paper is as follows. Section 2 presents the main information on participants and speech recordings. It also explains the proposed VOT extraction algorithm. Details about the global feature extraction process and the data analysis are also given in Section 2. In Section 3, the experimental results are presented. A discussion is presented in Section 4. Finally, Section 5 shows the conclusion.

2. Methods

2.1. Participants

A total of 54 Spanish native speakers participated in the study, 27 of which (9 women and 18 men) were diagnosed with PD. There are more men than women diagnosed with PD by a ratio 2:1, so the collected database reflects the gender distribution in the global PD population [27]. Their mean age (± standard deviation) was 68.41 ± 10.38 years. All of them were at very mild or mild stages in the H&Y scale (mean disease stage 1.85 ± 0.55). All of them were medicated according to their neurologist's prescription and were members of the *Asociación Regional del Parkinson de Extremadura* (Spain).

Table 1 contains information about the PD-affected patients in this study. This information includes gender, age in years, H&Y stage, treatment and time lapse since last intake in hours.

In addition, 27 HC subjects (9 women and 18 men) with no history of neurological or communication disorders were recruited. Their mean age was 69.22 ± 11.05 years. Age distribution showed no statistically significant differences between both groups (t=0.279, p-value=0.781).

The research protocol was approved by the Bioethical Committee from the University of Extremadura. All subjects signed an informed consent.

2.2. Vocal task and speech recording

The vocal task was to perform steady /pa/-/ta/-/ka/ syllable train repetitions as constantly and as quickly as possible during at least 4 seconds.

The speech recordings were made using a portable computer with an external sound card (TASCAM US322) and a headband microphone (AKG 520) featuring a cardiod pattern and positioned at approximately 4 cm from the

ID	Gender	Age	HY	Treatment	Last dose
PD-01	M	46	1	Azilect, Madopar, Mirapexin	2
PD-02	F	65	2	Azilect, Mirapexin, Stalevo	1.5
PD-03	F	70	2.5	Azilect, Mirapexin, Stalevo	2.5
PD-04	M	70	2	Rolpryna, Simenet Plus	1.5
PD-05	M	75	1	Azilect, Simenet	3
PD-06	M	80	1	Sumial	4.5
PD-07	M	66	2.5	Requip, Stalevo	5
PD-08	F	73	2.5	Azilect, Simenet Plus Retard	5
PD-09	M	53	1	Azilect, Neupro, Stalevo.	1
PD-10	M	77	2	Mirapexin, Simenet	1.5
PD-11	M	60	2	Neupro	2.25
PD-12	M	79	2	Simenet	2.5
PD-13	M	72	1	Azilect	4
PD-14	M	69	2	Azilect, Mirapexin, Simenet Plus	1.25
PD-15	M	74	2	Azilect, Requip, Stalevo	2.5
PD-16	F	69	2	Azilect, Simenet Plus	1.25
PD-17	M	70	2	Azilect, Neupro, Rolpryna	4
PD-18	M	69	1.5	Azilect, Requip, Stalevo	2.75
PD-19	F	56	1.5	Madopar, Mirapexin, Neupro	3
PD-20	M	71	2.5	Azilect, Neupro, Simenet Plus	4.5
PD-21	F	51	2.5	Neupro	0.1
PD-22	F	83	2.5	Azilect, Rivotril, Stalevo	4.5
PD-23	M	60	2.5	Rivotril, Stalevo	0.1
PD-24	M	49	1	Azilect, Mirapexin, Stalevo	1.5
PD-25	M	82	2	Simenet Plus	2
PD-26	F	79	1.5	Madopar	2.75
PD-27	F	79	2	Simenet Plus	4

Table 1: Description of the PD patients.

lips. The digital recording was carried out at a sampling rate of 44.1 KHz and a resolution of 16 bits/sample by using Audacity software[1] (version 2.0.5).

The recordings were made in a quiet room, which had no acoustic treatment. These recording conditions were consistent during the whole data collection process.

2.3. VOT extraction algorithm

The first step was to develop an algorithm suitable for the extraction of VOT also in cases of dysarthric speech. A novel approach has been proposed. Fig. 1 represents the corresponding block diagram.

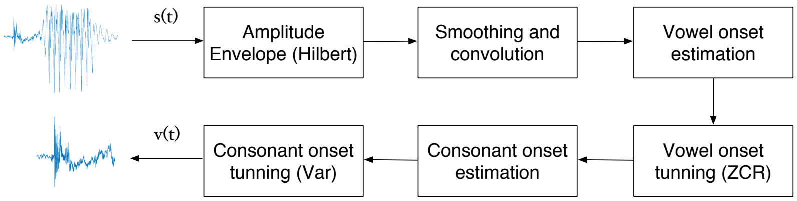
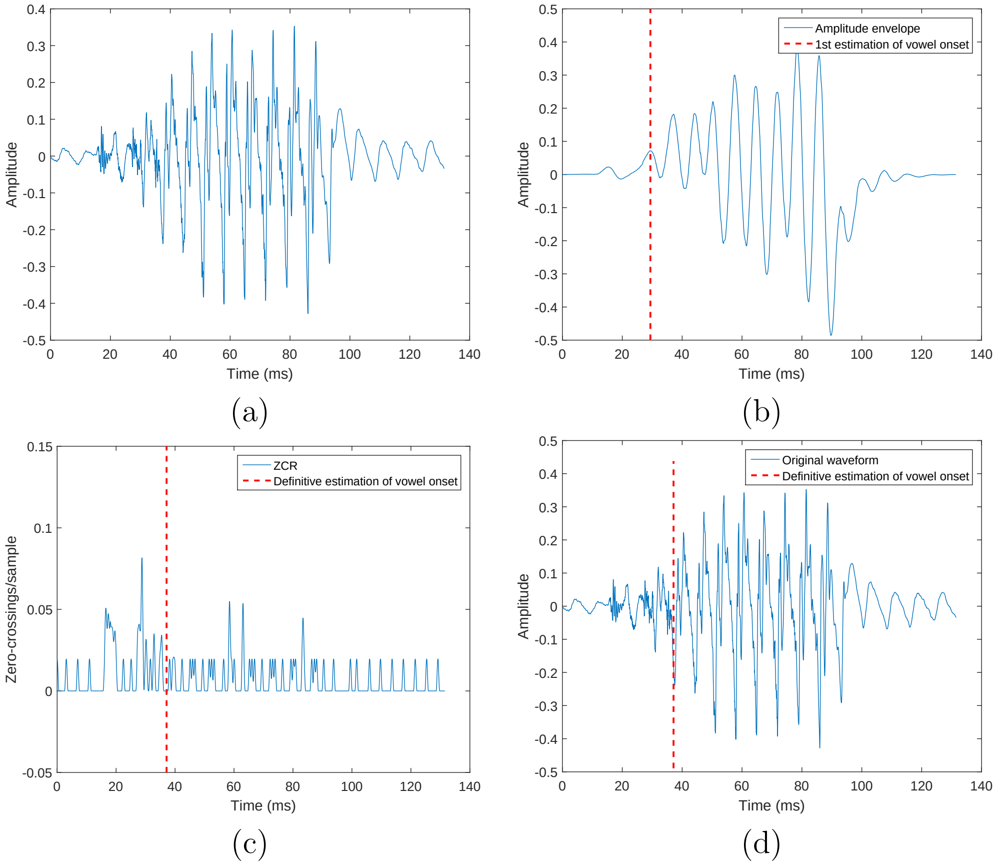

s(t)
Amplitude Envelope (Hilbert)
Smoothing and convolution
Vowel onset estimation
Vowel onset tunning (ZCR)
Consonant onset estimation
Consonant onset tunning (Var)
v(t)

Figure 1: VOT extraction algorithm.

The algorithm for VOT detection is based on the calculation of a smoothed amplitude envelope of the signal by means of Hilbert transform. Then, the envelope is convolved with a differenced Gaussian window of 10 milliseconds. This process allows to simplify the envelope with a second smooth for detection tasks. In general, the energy that represents the vowel is significantly higher than that of the consonant. So, the vowel onset can be estimated as the first prominent peak of the amplitude envelope with respect to the initial burst. The peak amplitude threshold is the mean value of the first half of the convolved amplitude envelope. However, there are subjects that produce very energetic stops. Approaches based exclusively on amplitude envelopes may fail in these cases leading to false vowel onset estimations. In [28], vowel onset estimation is addressed by using a Bayesian Step Changepoint Detector (BSCD). The novel approach presented here is based on an automatic Zero Crossing Rate (ZCR) tuning with the aim of checking that the estimated vowel onset was not set in a consonant region. The first estimation moves to the following peak of the amplitude envelope if the ZCR in the proximity of

[1] https://sourceforge.net/projects/audacity/

the considered peak is higher than the average ZCR. This tuning circumvents false vowel onset estimations in the case of energetic consonants in a more simple and computationally efficient way. Fig. 2 shows these algorithm steps applied on a /ka/ syllable uttered by a PD-affected speaker.



(a)

(b)

(c)

(d)

Figure 2: Vowel onset estimation on /ka/ syllable uttered by a PD patient. a) Original signal, b) amplitude envelope and first vowel onset estimation, c) vowel onset tuning through the ZCR, d) original waveform with final estimated vowel onset.

Once the vowel onset is estimated, the beginning of the initial burst can be detected by analysing the waveform and the spectrum towards the beginning of the signal. [28] used filtered spectrograms and summation along frequency axis to obtain an energy envelope and compute the difference. The approach proposed here includes a finer time-domain tuning step based on the variance. Thus, the procedure to perform initial burst detection is explained next. The spectrogram is calculated from the beginning of the signal to the estimated vowel onset. A first approximation of the beginning of the initial burst is chosen by taking the region of the frame where the energy has a value of at least 1%, with respect to the maximum energy of the interval between the beginning of the syllable fragment and the vowel onset. Then, this estimation is tuned by computing the variance for small fragments of 25 samples (0.6 milliseconds) of this interval. Since the plosive segments (specially /k/) are described by a high variability in time domain, the variance is useful to know when the signal noise ends and the consonant burst starts, because it presents an abrupt transition from one region to the other.

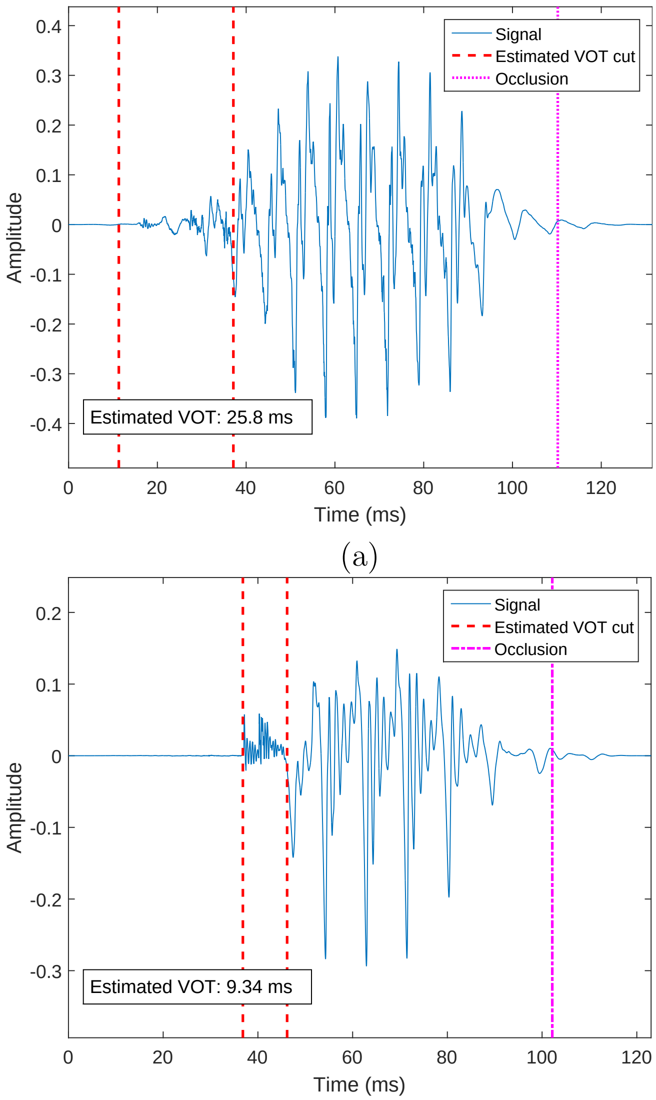
After VOT extraction, the occlusion instant is also estimated by taking the last third of the syllable and checking the point where the energy is at least 1% of the mean frame energy. For illustration, two /ka/ syllable waveforms, corresponding to a parkinsonian and a control voice, are shown in Fig. 3. The positions of initial burst, vowel onset and occlusion are represented by vertical lines on the plots.

The accuracy of the VOT estimation algorithm has been evaluated. For assessment purposes, the VOTs have been manually marked. The performance measure used is the percentage of times the estimated VOT is within certain temporal tolerances of the VOT value calculated from the hand labelled locations of the burst and voice onsets [29]. Thus the performance measure is the percentage of cases in which the absolute temporal deviations fulfill the following equation: $|t_v - \hat{t_v}| < \epsilon$, where $\hat{t_v}$ represents the VOT detected using the proposed algorithm, t_v denotes the manually determined VOT and ϵ denotes the tolerance value. Two tolerance values, also common in the literature, are considered: 10 and 20 ms [30]. Performance results are shown in Section 3.

2.4. Feature extraction

The analysis performed along each syllable of the same type (/pa/, /ta/ or /ka/, depending on the experiment) extracts a set of features with the aim of classifying a specific subject as healthy or PD affected. The set of features are related to measures that explore the voice recordings in time and frequency domains.

Most of the proposed features have VOT as base, a concept that is directly related to the voice articulation process and it can be applied to DDK task in order to obtain suitable descriptors for PD detection. As a result of



(a)

(b)

Figure 3: Estimated VOT and occlusion for /ka/ syllable (a) PD (b) HC.

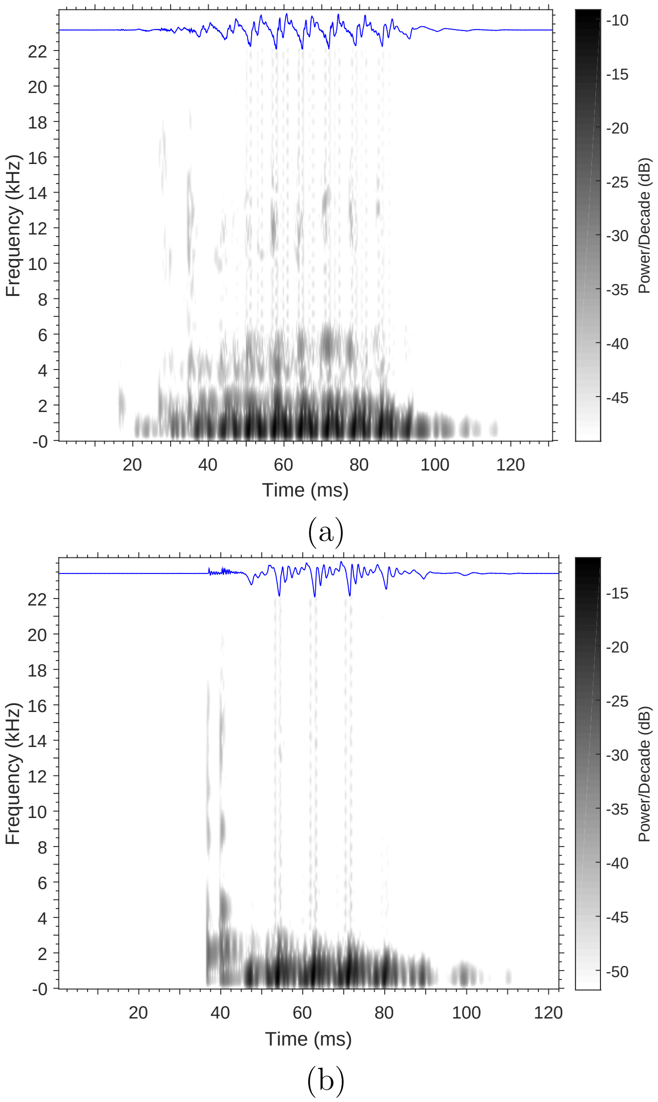
the application of the VOT extraction algorithm described in the previous subsection, a segment that contains the initial burst of the signal, representing the plosive consonant, is delimited. Since the occlusion point has been also estimated, the syllable length is also available.

Values of the different parameters are calculated for each type of syllable (/pa/, /ta/ or /ka/) in the DDK utterance. Although the recording time was initially set to 4 seconds, the waveform length used in the extraction was approximately 3 seconds (after removal of silences, loud breaths and other unwanted elements). This gives an average of 15 syllables (of each type) per utterance.

Six time duration features are considered: VOT, VOT ratio, CVRatio, Vowel Variability Quotient (VVQ), Consonant Variability Quotient (CVQ) and articulation rate. In the case of VOT, VOT ratio and CVRatio, the final features are the mean values of the parameters for the different syllables. VOT ratio is the relationship between VOT and syllable length (in percentage) and CVRatio is the relationship between VOT and vowel region length (in percentage). VVQ is the standard deviation of vowel duration, CVQ is the standard deviation of consonant duration and articulation rate is the number of syllables per second in an utterance.

Concerning frequency domain analysis, two groups of features are considered. The former includes features based on MFCCs. These features are the mean and standard deviation of the frame-based parameters. 13 MFCC parameters ($0-12^{th}$ order) are calculated for each frame. The 0^{th} order one simply represents the average speech energy, and each higher-order MFCC represents increasingly finer spectral detail. Frames have a length of 1024 samples, which corresponds to approximately 23.2 ms, and a 50% overlap. A 12 kHz bandwidth has been considered for the calculation of MFCCs. Energy at frequencies above 4 or 5 kHz has traditionally often been neglected in speech research. The main reason is that the low-frequency portion of the speech spectrum is considered sufficient from the viewpoint of perceptual intelligibility. However, it is assumed that the role of the higher portion of the spectrum is emphasized in this specific application. Fig. 4 shows how a spectrogram from a patient suffering from PD and another one from a HC present relevant differences in energy concentration up to approximately 12 kHz. As it is shown in the literature, velar stops are often spirantized in the case of PD. Spirantization represents the passage of air through an oral constriction when that constriction should be a complete closure that allows no airflow. This phenomenon causes spectrum alteration, showing energy at higher frequency bands in comparison to the case of healthy speakers [31].

The second group of spectral features consists of four spectral moments. In this case, the spectrum is represented by a small number of parameters that encode basic properties of its shape. The idea is borrowed from statistics.



(a)

(b)

Figure 4: Spectrogram for /ka/ syllable (a) PD (b) HC.

The four moments represent the mean, standard deviation, skewness and kurtosis, which describe the central tendency, dispersion, asymmetry, and peakedness of the spectrum. In this work, spectral moments are calculated on smoothed spectra of VOT segments, obtained by Linear Predictive Coding (LPC). The steps involved in this computation are summarized as follows:

1. VOT computation.
2. Windowing and pre-emphasis: The complete VOT segment is windowed (hamming) and pre-emphasized.
3. Spectral envelope computation from LPC coefficients: the order of LPC analysis is equal to the sampling frequency (in KHz) plus 2, rounded to the nearest natural number.
4. Computation of spectral moments from smoothed LPC spectrum, as detailed in the next paragraphs.
5. Averaging: For each utterance and syllable type, each feature represents the average value of each moment calculated across all VOT segments.

The first spectral moment μ_1 is the spectral center of gravity and it is calculated as:

$$\mu_1 = \sum_{k=1}^{K} f_k p_k,$$

where f_k denotes the frequency (in Hz) corresponding to bin k, K represents the number of frequency bins and p_k represents the value of the normalized LPC spectrum at bin k, that is:

$$p_k = \frac{a_k}{\sum_{k=1}^{K} a_k},$$

where a_k represents the LPC spectrum at bin k.

The second spectral moment μ_2 corresponds to the standard deviation, calculated as:

$$\mu_2 = \sqrt{\sum_{k=1}^{K} (f_k - \mu_1)^2 p_k}.$$

Skewness (μ_3) and kurtosis (μ_4) are calculated as:

$$\mu_m = \frac{\sum_{k=1}^{K} (f_k - \mu_1)^m p_k}{\mu_2^m},$$

for $m = 3, 4$.

The final feature set is presented in Table 2. It is an array of 36 features, combining time duration, MFCC-based and moment-based ones.

Time duration features	**Description**
VOT (ms)	Voice Onset Time
VOTratio (%)	VOT to syllable length ratio
CVratio (%)	VOT to vowel length ratio
VVQ (%)	Vowel Variability Quotient
CVQ (%)	Consonant Variability Quotient
Articulation rate (%)	Syllables per second
MFCC-based features	**Description**
$\overline{X}_{MFCC_i}, i = 0 - 12$	Means of $MFCC_i$
$\sigma_{MFCC_i}, i = 0 - 12$	Standard deviations of $MFCC_i$
Spectral moments	**Description**
Mean	Spectral central tendency
Standard deviation	Spectral dispersion
Skewness	Spectral asymmetry
Kurtosis	Spectral peakedness

Table 2: Features extracted from DDK tests.

2.5. Data analysis

Feature selection allows to reduce dimensionality avoiding the risk of overfitting and improving the interpretability of the results. A sequential backward feature selection approach has been used to search for the optimal feature subset in a Support Vector Machine (SVM) classification framework. Sequential backward selection starts from a set with all the features and iteratively deletes the least significant one until a stopping criterion is met.

The classifier engine is based on different kernel functions, namely linear, quadratic and third degree polynomial. Kernel choice has been empirically addressed, by running different experiments. No previous characterization of the dataset has been performed. Grid search has been used to tune the hyper-parameters in the three cases.

Classifier performance was measured using accuracy rate ($(TN+TP)/n$), sensitivity ($TP/(TP+FN)$) and specificity ($TN/(TN+FP)$), where TP is the true positive, TN is the true negative, FP is the false positive, FN is the false negative and n is the total number of subjects.

Cross-validation is used in order to assess the model generalization performance [32]. First, a stratified 10-fold cross-validation scheme is considered. The stratification is used to allow a balanced distribution in the folds between healthy subjects and people with PD. This process is performed 100 times and mean and standard deviation (mean ± sd) of the measures obtained in each iteration are computed. In addition, a LOO cross-validation is performed, which is k-fold cross-validation taken to its extreme, with k equal to the number of samples.

Independent t-tests have also been applied and the differences have been considered statistically significant when p-values are lower than 0.05.

Data analysis as well as feature extraction has been implemented in MATLAB environment.

3. Results

The performance of the VOT estimation algorithm is presented in Table 3. The values are percentages referring to the proportion of syllables whose automatically and manually detected VOT measures fall below a given tolerance value (10 or 20 ms).

Syllable	/ka/		/pa/		/ta/	
Tolerance	10 ms	20 ms	10 ms	20 ms	10 ms	20 ms
All	77.8	94.4	81.5	92.6	77.8	87.0
HC	74.1	88.9	85.2	92.6	77.8	88.9
PD	81.5	100.0	77.8	92.6	77.8	85.0

Table 3: Accuracy of VOT extraction algorithm (%), given tolerance values, for /ka/, /pa/ and /ta/ syllables. The tolerance values (10 and 20 ms) represent absolute deviations.

The results show high accuracies. Besides, there are no statistically significant differences between PD and HC groups for each one of the three syllables (p-values ≥ 0.2). Thus, the algorithm is robust to be applied both to HC and PD-affected speakers. In the case of PD patients, the algorithm is especially accurate when applied to /ka/ syllables, on which the proposed detection system is based. All the estimations deviate less than 20 ms from their manually labelled values and 81.5% of them deviate by less than 10 ms.

Classifier performance was estimated by using the previously mentioned specifications. The results are summarized in Tables 4 and 5. The best performance has been obtained by using the approach based on /ka/ syllables. With this approach, the best accuracy rates are 92.2%, by using 10-fold cross-validations, and 94.4%, with LOO validations. The approach based on /pa/ syllable demonstrates accuracy rates of 82.4% and 85.2%, whereas the third approach, based on /ta/ syllable, shows 79.2% and 87%, respectively. All of these results have been obtained by using a linear kernel function. The low standard deviation values achieved in the 10-fold cross-validation for the three plosives and with all the kernel functions allow to consider the accuracy rates reliable.

Approach based on /k/			
	Accuracy(%)	Specificity (%)	Sensitivity (%)
Linear	**92.2±2.7**	97.3±3.6	87.3±3.6
Quadratic	84.7±3.9	87.4±5.4	82.1±5.3
Poly3	82.7±3.4	91.9±4.8	73.7±4.4
Approach based on /p/			
	Accuracy(%)	Specificity (%)	Sensitivity (%)
Linear	82.4±2.7	82.6±3.5	82.4±3.5
Quadratic	77.4±4.1	76.8±5.3	77.9±5.5
Poly3	77.9±3.9	74.6±5.8	81.6±5.2
Approach based on /t/			
	Accuracy(%)	Specificity (%)	Sensitivity (%)
Linear	79.2±3.3	80.6±4.6	77.6±5.0
Quadratic	77.4±4.1	71.2±5.6	83.9±5.1
Poly3	74.5±3.7	68.6±5.8	81.0±4.8

Table 4: Classification performance based on 10-fold cross-validation.

Concerning the best approach, based on /k/ plosive, the subset of features that are fed to the linear SVM classifier after the automatic feature selection process is composed of:

- Two temporal features: VVQ and articulation rate. As a standard deviation of a temporal measure across the different syllables in the utterance, VVQ feature provides information about stability of timing. Patients with PD show less stable vowel duration due to impaired muscle control. Articulation rate is affected in dysarthric speakers due to reduced muscle speed. [6] found statistical significances between PD and HC groups in the articulation rate.

- One spectral moment: the skewness. This feature has been selected with the three considered kernel functions and thus plays an important role in the classification performance. Considering the four spectral

Approach based on /k/			
	Accuracy(%)	Specificity (%)	Sensitivity (%)
Linear	**94.4**	100	88.9
Quadratic	85.2	85.2	85.2
Poly3	85.2	96.3	74.1
Approach based on /p/			
	Accuracy(%)	Specificity (%)	Sensitivity (%)
Linear	85.2	85.2	85.2
Quadratic	77.8	77.8	77.8
Poly3	79.6	74.1	85.2
Approach based on /t/			
	Accuracy(%)	Specificity (%)	Sensitivity (%)
Linear	87.0	85.2	88.9
Quadratic	75.9	70.4	81.5
Poly3	75.9	70.4	81.5

Table 5: Classification performance based on LOO cross-validation.

moments, the asymmetry of spectral shape demonstrates to have the highest discrimination capability.

- Eleven features based on MFCCs: 4 of them are mean values whereas the other 7 features are standard deviations. Therefore the feature selection process emphasizes variability of MFCCs among frames over mean values, keeping only some middle-order MFCC mean values (5^{th} and $7-9^{th}$ orders).

The combination of temporal and spectral features has been shown to be successful in discriminating PD and HC. Besides the good accuracy rate obtained, the sensitivity and specificity provide important information. The specificity is estimated as 97.3% (10-fold) or 100% (LOO). The 95% confidence interval is $(96.6\%, 98.0\%)$ with 10-fold cross-validation. This means there is a very high proportion of healthy people that are correctly identified as not having PD. In the considered application scenario, false positive cases detected in primary health care would be forwarded to the neurology unit. A high specificity in this context implies that the additional cost due to false alerts is very small. The sensitivity is also high: 87.3% (10-fold) and 88.9% (LOO). The 95% confidence interval is $(86.6\%, 88.0\%)$ with 10-fold cross-validation. This means that the system is being accurate to detect PD.

The results support the argument that important information for PD detection in early stages can be extracted from articulatory tasks such as the DDK test, specially with /k/ plosive, following the proposed approach.

4. Discussion

Novel biomarkers are proposed for PD detection in early stages based on articulatory tasks. The proposed method concentrates specifically on the velar stop. Several medical papers, cited before, support the choice of /k/ segments, since they report impaired velar closure in PD. However, to the best of the authors' knowledge, a comparison of the three unvoiced plosives in the context of automatic classification experiments has not been reported yet.

There has been little effort specifically targeting early-stage PD. The relevance of this work is increased by the fact that the results are obtained on a new database, composed of recordings from subjects at an earlier stage of the disease (H&Y scale) in comparison to previous investigations based on articulatory features. The mean stage is 1.85 ± 0.55, in comparison to the databases in [17] and [18], with mean stages of 2.2 ± 0.5 and 2.3 ± 0.8, respectively.

In the context of PD detection, due to the difficulty of recruiting patients, it has become usual to conduct experiments with replicated recordings and assess the performance of a certain approach by using independence-based classification methods. The misuse of replicated recordings artificially increases the sample size, leads to a diffuse criterion to decide when a subject should be classified as suffering from PD, and can produce over-optimistic results [12, 16]. In [18] the authors used two replications per subject, so independence was not satisfied. In this work, as well as in [17], only one measure per subject is used. Nevertheless, [18] provided a relevant methodological advance in PD detection based on articulatory features.

Concerning recording conditions, the effects of hostile noisy environments on the performance of many speech-based applications (such as automatic speech recognition) have been deeply explored, but these effects are much less explored for automatic diagnosis of PD. This research line is at an earlier stage and most of the contributions use data collected in laboratory setups with little or no background noise [7]. [33] analyzes the effect of acoustic conditions on different algorithms to detect PD from speech. The results show that background noise considerably impacts the performance of the different algorithms. Therefore a certain control of the acoustic environment, in particular of noise, must be assumed. This does not mean that an acoustically isolated room is required, but a quiet room with a low ambient noise level.

Feature calculation relies on the use of an accurate and computationally efficient VOT extraction algorithm. This algorithm provides an accurate estimation of VOT boundaries from /pa/, /ta/ and /ka/ syllables also in the case of dysarthric voices.

The introduction of the spectral moments extracted from unvoiced stops has allowed to identify a promising new feature (skewness extracted from the velar stop). The weakening of the velar stop /k/ due to spirantization changes the ratio between low and high-frequency energy, having a direct impact on the spectral skewness.

Articulatory features are promising to discriminate PD from speech disorders coming from other types of pathologies resulting in dysphonia, for example, laryngeal diseases. Databases including different types of disorders affecting speech are not often found. An interesting recent investigation on different types of pathologies affecting speech is presented in [34]. This paper evaluates the accuracy of different methods, but only considers recordings of sustained phonations. The incorporation of articulatory features may represent a definitive step forward for the implementation of this technology in the clinical praxis, in a scenario with multiple possible pathologies.

The current classification approach is not cost-sensitive. Medical diagnosis can be addressed as a cost-sensitive classification problem, with different penalties assigned to different misclassification errors. In this concrete application this would mean to trade-off between the financial and emotional cost of a false positive (unnecessary visit to neurologist) versus the clinical cost of a false negative (no access to early intervention). A deeper investigation taking these different aspects into account could lead to develop a cost-sensitive approach. This definitively requires medical and health financial management viewpoints.

5. Conclusion

A novel system (based on DDK voice recordings) is proposed for automatic detection of early-stage PD. After comparison with the other voiceless stops (/p/ and /t/), the proposed approach is based specifically on the velar one (/k/), which demonstrates the highest discrimination capability.

The method involves temporal and spectral features. High accuracy rates of 92.2% (10-fold cross-validation) or 94.4% (LOO cross-validation) have been obtained by using the proposed approach. These results have been obtained on a database with lower disease stages (H&Y scale) in comparison to previous investigations based on DDK recordings.

A goal for the immediate future is to increase the database by recording new subjects. The obtained results are achieved on a database, which is composed of 54 subjects (27 suffering from PD and 27 HC). Acoustic data collection is a troublesome task in the case of PD due to the difficulty in recruiting patients. DDK test constitutes a harder speaking task than sustained vowel tasks, so the collection of large databases is, in this case, even more challenging. A larger database would also allow to take gender issues into account. Being voice and speech highly conditioned by gender, separation by gender is expected to further enhance the system performance.

Future direction of research may include also tele-monitoring applications. These applications require considering robustness under adverse environmental conditions, thus searching for robust solutions should be addressed before these systems are implemented on mobile platforms. Besides the benefit for the patient, tele-monitoring also opens up the possibility of a massive collection of data. In the last decades, physicists have demonstrated the power of quantum systems for information processing [35]. The advantages of quantum computing can be exploited in order to develop more powerful algorithms in the fields of signal processing and machine learning. Although still in the very first stages of development, quantum computing and the massive collection of data may transform the field of medical diagnosis.

In its current state, the proposed approach can be used as a noninvasive, low-cost tool to help family physicians screen for and identify PD in early stages and refer the person to the neurology unit. Early detection of PD allows to improve the patient's quality of life through effective medical treatment as well as other non-medical therapies, e.g., physiotherapy or speech and language therapy.

Conflict of interest

The authors declare that there is no conflict of interest related to this paper.

Acknowledgements

Thanks to the anonymous participants and to Rosa María Muñoz for providing support on the voice recordings and information from the patients. We are grateful to the *Asociación Regional de Parkinson de Extremadura* and *Confederación Española de Personas con Discapacidad Física y Orgánica* for providing support in the experiment development. We also thank two referees for comments and suggestions which have highly improved both the readability and the content of this paper.

This research has been supported by project MTM2014-56949-C3-3-R (MINECO) and projects IB16054, GR15052, and GR15106 (Junta de Extremadura/European Regional Development Funds, EU).

References

[1] M. Josefiok, T. Krahn, J. Sauer, A survey on expert systems for diagnosis support in the field of neurology, in: Intelligent Decision Technologies, Springer, 2015, pp. 291–300.

[2] A. Bhardwaj, A. Tiwari, Breast cancer diagnosis using genetically optimized neural network model, Expert Systems with Applications 42 (10) (2015) 4611–4620.

[3] Q. A. Rahman, L. G. Tereshchenko, M. Kongkatong, T. Abraham, M. R. Abraham, H. Shatkay, Utilizing ECG-based heartbeat classification for hypertrophic cardiomyopathy identification, IEEE Transactions on Nanobioscience 14 (5) (2015) 505–512.

[4] A. Paul, D. P. Mukherjee, Mitosis detection for invasive breast cancer grading in histopathological images, IEEE Transactions on Image Processing 24 (11) (2015) 4041–4054.

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

[6] J. Rusz, R. Cmejla, H. Ruzickova, E. Ruzicka, Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease, Journal of Acoustical Society of America 129 (1) (2011) 350–367.

[7] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, L. O. Ramig, Suitability of dysphonia measurements for telemonitoring of Parkinson's disease, IEEE Transactions on Biomedical Engineering 56 (4) (2009) 1015–1022.

[8] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, L. O. Ramig, Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease, IEEE Transactions on Biomedical Engineering 59 (5) (2012) 1264–1271.

[9] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, O. Kursun, Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings, IEEE Journal of Biomedical and Health Informatics 17 (4) (2013) 828–834.

[10] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, E. Nöth, Analysis of speech from people with Parkinson's disease through nonlinear dynamics, in: T. Drugman, T. Dutoit (Eds.), Advances in Nonlinear Speech Processing, Vol. LNAI 7911 of Lecture Notes in Artificial Intelligence, Springer-Verlag, 2013, pp. 112–119.

[11] M. Hariharan, K. Polat, R. Sindhu, A new hybrid intelligent system for accurate detection of Parkinson's disease, Computer Methods and Programs in Biomedicine 113 (3) (2014) 904–913.

[12] C. J. Pérez, L. Naranjo, J. Martín, Y. Campos-Roca, A latent variable-based Bayesian regression to address recording replication in Parkinson's disease, in: EURASIP (Ed.), Proceedings of the 22nd European Signal Processing Conference (EUSIPCO-2014), IEEE, Lisbon, Portugal, 2014, pp. 1447–1451.

[13] D. L. Murman, Early treatment of Parkinson's disease: opportunities for managed care, The American Journal of Managed Care 18 (7) (2012) S183–8.

[14] A. Schrag, Y. Ben-Shlomo, N. Quinn, How valid is the clinical diagnosis of Parkinson's disease in the community?, Journal of Neurology, Neurosurgery & Psychiatry 73 (5) (2002) 529–534.

[15] M. M. Hoehn, M. D. Yahr, Parkinsonism: onset, progression and mortality, Neurology 17 (5) (1967) 427–442.

[16] L. Naranjo, C. J. Pérez, Y. Campos-Roca, J. Martín, Addressing voice recording replications for Parkinson's disease detection, Expert Systems With Applications 46 (2016) 286–292.

[17] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, K. Daqrouq, S. Skodda, J. Rusz, E. Nöth, Automatic detection of Parkinson's disease in running speech spoken in three different languages, Journal of the Acoustic Society of America 139 (1) (2016) 481–500.

[18] M. Novotny, J. Rusz, R. Cmejla, E. Ruzicka, Automatic evaluation of articulatory disorders in Parkinson's disease, IEEE/ACM Transactions on Audio, Speech, and Language Processing 22 (9) (2014) 1366–1378.

[19] S. S. Kumaran, S. Gudwani, M. Saxena, M. Behari, Study of articulatory movement from the single slice dynamic imaging of the vocal tract in Parkinsonism, Vol. 21, 2013, Ch. Proceedings of the International Society for Magnetic Resonance in Medicine, pp. 2843–2843.

[20] M. J. Hammer, S. M. Barlow, K. E. Lyons, R. Pahwa, Subthalamic nucleus deep brain stimulation changes velopharyngeal control in Parkinson's disease, Journal of Communication Disorders 44 (1) (2011) 37–48.

[21] E. Q. Wang, L. V. Metman, R. A. E. Bakay, J. Arzbaecher, B. Bernard, D. M. Corcos, Hemisphere-specific effects of subthalamic nucleus deep brain stimulation on speaking rate and articulatory accuracy of syllable repetitions in Parkinson's disease, Journal of Medical Speech-Language Pathology 14 (4) (2006) 323–334.

[22] S. Davis, P. Mermelstein, Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, IEEE Transactions on Acoustics, Speech and Signal Processing 28 (4) (1980) 357–366.

[23] K. Forrest, G. Weismer, P. Milenkovic, R. N. Dougall, Statistical analysis of word-initial voiceless obstruents: preliminary data, Journal of the Acoustical Society of America 84 (1988) 115–123.

[24] J. W. Lee, J. Y. Choi, H. G. Kang, Classification of stop place in consonant-vowel contexts using feature extrapolation of acoustic-

phonetic features in telephone speech, The Journal of the Acoustic Society of America 131 (2) (2012) 1536–1546.

[25] L. K. Smith, A. M. Goberman, Long-time average spectrum in individuals with Parkinson's disease, NeuroRehabilitation 35 (2014) 77–88.

[26] P. McRael, K. Tjaden, Spectral properties of fricatives in Parkinson's disease, Journal of the Acoustic Society of America 104 (1998) 1854–1854.

[27] I. N. Miller, A. Cronin-Golomb, Gender differences in Parkinson's disease: clinical characteristics and cognition, Movement Disorders 25 (16) (2010) 2695–2703.

[28] M. Novotny, J. Pospisil, R. Cmejla, J. Rusz, Automatic detection of voice onset time in dysarthric speech, in: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing, 2015, pp. 4340–4344.

[29] E. Fischer, A. M. Goberman, Voice onset time in Parkinson disease, Journal of Communication Disorders 43 (1) (2010) 21–34.

[30] C. Y. Lin, H. C. Wang, Automatic estimation of voice onset time for word-initial stops by applying random forest to onset detection, The Journal of the Acoustical Society of America 130 (1) (2011) 514–525.

[31] K. Chenausky, J. MacAuslan, R. Goldhor, Acoustic analysis of PD speech, Parkinson's Disease 2011 (ID435232) (2011) 1–13.

[32] A. Webb, Statistical Pattern Recognition, John Wiley and Sons, Chichester, 2002.

[33] J. C. Vásquez-Correa, J. Serrá, J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, E. Nöth, Effect of acoustic conditions on algorithms to detect Parkinson's disease from speech, in: IEEE International Conference on Acoustics, Speech and Signal Processing 2017, IEEE, 2017.

[34] J. R. Orozco-Arroyave, E. A. Belalcazar-Bolaños, J. D. Arias-Londoño, J. F. Vargas-Bonilla, S. Skodda, J. Rusz, K. Daqrouq, F. Hönig, E. Nöth, Characterization methods for the detection of multiple voice disorders: Neurological, functional, and laryngeal diseases, IEEE Journal of Biomedical and Health Informatics 19 (6) (2015) 1820–1828.

[35] S. Imre, L. Gyongyosi, Advanced Quantum Communications - An Engineering Approach, Wiley-IEEE Press, 2012.