Biomedical Signal Processing and Control 66 (2021) 102418

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Advances in Parkinson's Disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects

Laureano Moro-Velazquez [a,*], Jorge A. Gomez-Garcia [b], Julian D. Arias-Londoño [c], Najim Dehak [a], Juan I. Godino-Llorente [b]

[a] Center for Language and Speech Processing, Johns Hopkins University, Baltimore, USA
[b] Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain
[c] Department of Systems Engineering, Universidad de Antioquia, Medellín, Colombia

ARTICLE INFO

Keywords:
Parkinson's disease
Voice and speech
Phonation
Articulation
Automatic detection
Automatic assessment

ABSTRACT

Parkinson's Disease (PD) affects speech in the form of dysphonia and hypokinetic dysarthria. Multiple studies have evaluated PD's influence on different aspects of speech, showing differences between speakers with and without PD. Most recent studies are focused on the proposal of new automatic and objective tools to help in the diagnosis and severity assessment. This comprehensive review identifies the most common features and machine learning techniques employed in automatically detecting and assessing the severity of PD using phonatory and articulatory aspects of speech and voice. We discuss their discriminant properties and literature findings as well as identify common methodological issues that can potentially bias results. The objective is to provide a broad overview of these methods, their advantages and disadvantages, and to identify the most promising methodologies to be explored in future works. We conclude that there is clear evidence that the articulatory and phonatory aspects of speech and voice are relevant for the automatic detection and severity assessment of PD. However, there is no standard methodology sufficiently validated in a clinical trial, and further research is required, especially to develop larger corpora and identify new objective biomarkers.

1. Introduction

Parkinson's Disease (PD) is a chronic and progressive condition caused by the gradual neuronal death in the substantia nigra, implicated in the production of dopamine neurotransmitters, which play a crucial role in motor control. [1]. Despite the fact that PD is the second most common neurodegenerative disease, the average diaganosis time is above two years [2]. Therefore, new precision medicine tools based on patient's signs are needed to assist diagnosis and personalized treatment. To this respect, speech involves complex and precise coordination of the respiratory system, larynx, and supraglottal articulators, being an excellent candidate to provide such diagnostic information [3].

Several studies previous to 1970 described the speech of people with PD [4,5], although it is not until 1969 [6,7] that a group of researchers analyzed more thoroughly the problems of phonation, prosody, and articulation of PD patients. Since then, a plethora of studies have presented evidence that the neurodegenerative processes associated with PD cause dysphonia and dysarthria, particularly *hypokinetic dysarthria* [6,8–10] in different stages of the disease [11]. Dysphonia can be defined as the speaker's incapacity to produce a normal phonation due to the phonatory system's impaired functioning, while dysarthria is more related to problems with articulation when pronouncing words. More specifically, *hypokinetic dysarthria* is characterized by a reduction of loudness and articulation amplitude, slow speech rates combined with rushes of fast speech sometimes, and a decrease of intelligibility mostly. Some specialists suggest that 90% of PD patients suffer from dysarthria [12,3] after a median latency period of 7 years since diagnosis, which is in accordance with the study [13], although other works such as [14,15] point to lower values. The prevalence of dysarthria and dysphonia in PD patients is unclear. Some studies determine a higher prevalence of dysphonia (65.5%) over articulatory impairment (38.5%) when perceptually evaluating 200 patients [15]. These results are consistent with those reported by previous studies [13] employing the same number of patients. However, the mentioned studies employed perceptual methods. More recent studies such as [16] suggest a higher prevalence of articulatory deficits. Whereas this last study uses objective

* Corresponding author.
E-mail address: laureano@jhu.edu (L. Moro-Velazquez).

https://doi.org/10.1016/j.bspc.2021.102418
Received 31 May 2020; Received in revised form 20 November 2020; Accepted 10 January 2021
Available online 1 February 2021

measurements, the employed cohort is much smaller. In this context, literature shows a myriad of studies and approaches investigating the influence of PD in the voice or speech of patients and proposing new biomarkers or automatic detection schemes to support the diagnosis of this condition. The interest in parkinsonian speech has increased with time as shown in Fig. 1, which includes the number of studies per year related to dysarthria in PD and the influence of PD in motor speech aspects between 1956 and May 2020, according to PubMed. In general terms, it is possible to divide these studies into four main groups depending on the analyzed speech aspect, i.e., phonatory, articulatory, prosodic, and cognitive-linguistic:

- Phonatory studies are related to the glottal source and resonant structures of the vocal tract, employing most of the time sustained vowels as acoustic material for the analysis.
- Studies based on articulatory aspects are more diverse as there exist more analysis possibilities: the features or the acoustic measurements analyzed can be extracted from different types of sound segments and can be related to the velocity or acceleration of articulators, type of transitions between segments, or the evolution of formants among others.
- Prosodic studies mainly focus on paralinguistic features such as pitch variation, syllable rate analysis, or the manifestation of emotions in the speech signal.
- Finally, the cognitive-linguistic approaches analyze the deviations in cognitive behavior by examining the vocabulary, sentence complexity, phrase construction, and the existence of word repetitions, among other manifestations.

The speech task or acoustic material that is used in each case is a differentiating key factor of the four aspect groups. In the phonatory analysis, sustained vowels are commonly employed as acoustic material, while in the other three groups, the use of connected speech is necessary. For the latter, monologues, read passages, and Diadochokinetic (DDK) tasks[1] are mostly considered.

Among all the studies evaluating of the patient's speech, it is possible to differentiate between two types of analysis:

- *Perceptual assessment*, where trained evaluators follow a specific protocol such as the Frenchay Dysarthria assessment [17] to rate certain aspects of the speech

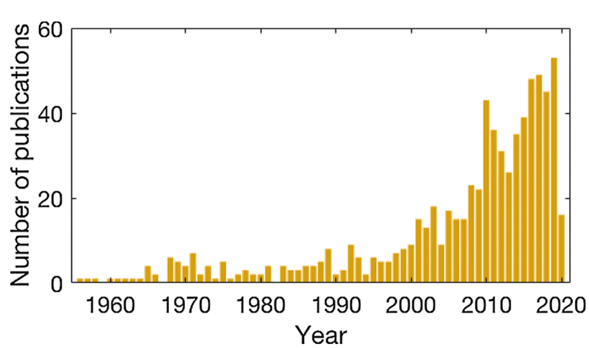

Fig. 1. Number of publications per year from 1956 to May 2020 related to PD, speech, dysarthria, phonatory, articulatory and prosodic aspects found in PubMed.

- *Objective analysis*, which makes use of algorithms to characterize the signal through certain acoustic features.

Some of the works performing objective assessment propose measures or biomarkers to assist PD diagnosis or correlate with its severity. Others propose automatic binary classifiers to distinguish between speakers with and without PD or automatically predict the disease's stage by processing the speech signal.

To this extent, the literature about PD presents reviews concerning the influence of speech treatments [18,19], the effects of the pharmacological and surgical therapy [20–22] on voice and speech, prosody [23], aspects of language [24], speech perception [25], or behavioral treatments for speech [26], among others. Despite that the studies in the literature about parkinsonian speech are abundant, as Fig. 1 shows, no article reviews the findings on the use of the phonatory and articulatory aspects of voice and speech with diagnosis purposes.

Consequently, this article aims to fill this gap, presenting a comprehensive review of the state of the art on PD diagnosis and severity assessment employing signal processing and machine learning techniques focused on phonatory and articulatory aspects of voice and speech. Our objective is to perform a survey of the state of the art to provide a categorization of those studies describing the main speech tasks, features, and signal processing techniques employed in the objective assessment of PD. This review aims to provide a broad overview and discussion of the existing methods and their advantages/disadvantages, and will help to identify the most recommended methodologies to be followed in future works. As secondary objectives, we identify some methodological issues that often arise in the processing of PD speech, and describe the corpora that are publicly available to analyze phonatory and articulatory aspects of parkinsonian speech.

The article is organized as follows: Section 2 introduces the methods to the review, Sections 3 and 4 includes an analysis of phonatory and articulatory aspects, respectively, while Section 5 lists and describes the most common features employed in the state of the art. Lastly, Sections 6 and 7 include the discussion and conclusions of the article.

2. Methods

In this review, we selected articles from PubMed and Google Scholar search after having considered the following selection criteria: documents in English analyzing the effect of PD in voice and speech, proposing new biomarkers or methodologies to detect or assess the disease utilizing speech technologies. The search queries employed to obtain the list of studies can be found in the additional material accompanying this paper. As the review is focused on the articulatory and phonatory aspects of speech from people with PD, we discarded those works analyzing the prosodic or linguistic aspects exclusively, those that we considered that had very limited contributions, and those analyzing the effects of pharmacological therapy or surgery in the patients' speech.

We also excluded studies that we have considered that had methodological issues, such as those using highly unbalanced corpora (average age difference between patients and controls of more than 10 years or unmatched gender between patients and controls in more than 50% of participants) or not using validation procedures to obtain the results (no cross-validation or cross-corpora validation, for instance). When relevant, we included some of these studies in the review, but we indicated the corresponding methodological issue. Likewise, we discarded those originated by conference abstracts that did not have any associated published documents from all records found.

A list of 652 publications were identified in PubMed and 1.190 in Google Scholar until May 2020, from which we selected 192 by carefully reading their title and abstract and according to the mentioned criteria. Then, these studies were analyzed in detail.

Additionally, we included in this review a list of the acoustic features that were identified, along with some representative studies using them and a visualization of their discriminant properties in the Neurovoz

[1] DDK tasks consist of the repetition of words or syllables at a certain syllable rate, which can be steady or increasing until the speakers reach their limit rate. There are two primary manners to perform DDK analysis: by repeating a single syllable, or a sequence of syllables.

corpus [27].

3. Phonatory aspects

3.1. Evidence of the influence of PD in the phonatory system

Studies about the influence of PD in the phonatory system mainly analyze the impairments in phonatory-related structures and muscles like the diaphragm, the muscles connected to the larynx, the vocal folds, or the supra-glottal resonant cavities. The type of acoustic material analyzed in these cases is usually the voice signal from one or several sustained vowels, which can be maintained during a specific time-lapse or modulated in frequency and amplitude. Sustained vowels are expected to generate simple acoustic traces that might lead to consistent and, to some extent, reliable analysis of voice quality. The objective analysis of phonatory aspects is relatively popular since acoustic measures can be directly obtained using available software and libraries such as the Multidimensional Voice Program (MDVP) [28], Praat [29] or OpenSMILE [30].

In one of the earliest articles studying this aspect [13], authors perceptually analyze the dysfunctions of voice and speech on 200 untreated patients (idiopathic and postencephalitic). After assessing their voice quality, the study reports laryngeal and/or vocal tract abnormalities in 89% of the patients, comprising hoarseness, breathiness, tremulousness, and roughness, while only 10% exhibits hypernasality. In a subsequent article [31], authors examine PD patients –32 in this case– through telescopic cinelaringoscopy and report rigidity in the phonatory posture of the larynx and a high correlation with limb rigidity. That means that if the patients' rigidity is higher on the left limbs, the vocal folds exhibit the same effect. This increased muscular activity produces a bowing of the vocal folds and leads to incomplete closure of the glottal gap, phonation weakness, and breathiness [32]. Although this effect cannot always be perceived, it is measurable [33]. In concordance with these observations, other studies suggest that it is possible to detect PD's influence in the vibration of the vocal folds because of an excess of noise and other perturbations in the phonatory signal. This may be caused by abnormal phase closure and phase asymmetry or vocal tremor resulting from abnormal vibrations visualized in the arytenoid cartilages' movements [34]. In a direct relationship with these conclusions, it has been reported that patients present a significantly higher subglottal pressure with respect to control subjects for the same phonation level, indicating that the mean laryngeal resistance of patients is higher too, also correlating with the patient's perception of increased effort to produce phonation [35]. In contrast, [36] reports lower laryngeal resistance and subglottal pressure in PD patients respect to controls. However, the higher average sound pressure level during phonation observed in controls respect to patients in the same study can explain these differences.

3.2. Acoustic measurements and perceptual evaluations

Regarding the range of possible acoustic measurements to characterize PD's influence in patients' phonation, jitter, shimmer, and noise (Harmonic to Noise Ratio (HNR) or Noise to Harmonics Ratio (NHR)) [37] are popular measures. Multiple studies employing a sustained phonation of vowels (the most common is /aː/) from patients and controls, obtain significant differences between groups for the three measurements, being jitter and shimmer higher and HNR lower in patients [38–42]. In contrast, one study [43] finds a more prominent jitter only in patients in the later stages of the disease, while shimmer and NHR yield non-significant statistical differences between patients and controls. Accordingly, the study in [44] reports a significantly higher jitter and relative jitter in PD male patients compared to male controls and no significant differences for shimmer and percent shimmer. Conversely, more recent studies [45] find higher correlation between shimmer and PD than jitter, analyzing the sustained phonation of 40 de-novo PD patients and 40 control subjects. Some authors attribute this behavior to a reduced short-term neuromuscular control of the laryngeal abductory or adductory mechanisms [41]. Besides that, the higher values of NHR in PD patients with respect to controls are partially attributed to the muscular rigidity and tremor present in the larynx. Other works do not find significant differences in jitter and shimmer between groups [32,46,47].

Other acoustic measurements and perceptual ratings have been employed to differentiate between patients and controls, including: maximum phonation time; Grade, Roughness, Breathiness, Asthenia and Strain scale (GRBAS) ratings [32]; nonlinear dynamic analysis (Correlation Dimension D_2) [44]; Glottal to Noise Excitation ratio (GNE) [48]; Soft Phonation Index (SPI) [41] or Fundamental frequency tremor (Fftr) [41], among others, although studies include these features less frequently. Moreover, certain studies analyze formant-related features, such as the Vowel Space Area (VSA) obtained from sustained vowels, to differentiate patients from control speakers. An example of these works is [42], where authors measure triangular Vowel Space Area (tVSA) in addition to jitter, shimmer, and NHR. Results indicate that patients exhibit a decreased tVSA, respect to controls. However, authors only consider 9 control speakers and 9 patients, most of them in a mild to an advanced stage of PD.

Only a few studies perform a longitudinal analysis of the voice of patients. To this respect, authors in [49] conduct a perceptual assessment as well as measures of jitter, shimmer, and NHR over the sustained vowels of 80 patients recorded in two sessions (with a mean time between the examinations of 33 months) and 60 age-matched controls (with a mean time between examinations of 25 months.) The authors find a high correlation between the perceptual assessment of voice and the Hoehn & Yahr (H–Y) rating along time and significantly higher values of shimmer and NHR in male and female patients compared to their respective control groups, while jitter does not provide differentiation between them.

Another example of a longitudinal study can be found in [50], where the authors analyze phonatory, articulatory, and prosodic aspects of 19 Czech patients before and after being treated with symptomatic medication (mostly, levodopa or a dopamine agonist). Jitter, shimmer, HNR, Pitch Period Entropy (PPE), Recurrence Period Density Entropy (RPDE), and Maximum Phonation Time (MPT) are measured, finding statistically significant differences between patients and controls in all of these measures except by MPT and RPDE. These results are found before and after the beginning of the pharmacological therapy, although the voice quality tends to improve after the treatment. Several articles report differences in PD's impact on phonation for males and females in other features besides jitter, shimmer, and noise. For instance, in [51], the authors identify distinct alterations in spectral energy distribution, and abrupt fundamental frequency shifts in female patients, compared to males and controls. In this respect, multiple studies conclude that the differences in the shape and size of men and women's larynx explain the differences in vocal dysfunctions [51,39,44,52]. Likewise, a higher fundamental frequency in PD female patients compared to control speakers has been reported [39].

The fundamental frequency of patients and controls has also been compared in some studies [53,32,41,54,55]. The study [53] reports no significant difference in the average fundamental frequency between 30 patients and 20 controls. However, the patients in this study are 7 years older than controls in average, which could be biasing the results. In another subsequent study employing a corpus with age-matched groups [32], the authors report higher fundamental frequency and fundamental frequency variability in patients (20 participants) than in controls (20 participants) and low correlation between this feature and symptom assessment. Another study employing sustained vowels from 39 patients and 62 controls reports significant differences between the fundamental frequency variability of both groups [41], in concordance with [32]. In [55], the authors analyze the fundamental frequency of 16 patients and

19 controls while receiving perturbed auditory feedback. The study results suggest that the patients tend to have a reduced compensation of their fundamental frequency to the artificial auditory perturbation, in concordance with a subsequent study [56]. Although the fundamental frequency is analyzed in some phonatory-related studies, this feature is more characteristic of works leveraging information from prosody, where the range and variability of this feature tend to provide differentiation between patients and controls [57]. Nevertheless, the fundamental frequency is employed in some phonatory features such as jitter, PPE, Normalized Noise Energy (NNE), or some HNR-related methods, among others. Therefore, a correct estimation of this frequency is crucial since it could undermine the conclusions obtained in some studies. In this respect, many feature extraction toolboxes such as Automatic Voice Condition Analysis (AVCA) [58] or Data Analytics Research and Technology in Healthcare-Voice Analysis Toolbox (DARTH-VAT) [59] offer different techniques to calculate fundamental frequency. In spite of this, many of the analyzed studies do not specify the algorithm employed to determine the fundamental frequency, which complicates the comparison of methodologies.

3.3. Automatic detection and assessment

Within all the works addressing phonatory aspects of people with PD, some of the most recent use certain features as the input of automatic detectors to distinguish between parkinsonian and normophonic voices or predict the severity of the disease in a patient utilizing machine learning technologies.

Many of the analyzed studies perform this detection employing traditional measurements such as jitter, shimmer, or noise, among others, in combination with various classification schemes that usually provide accuracy over 75% [60–63].

In this sense, some works use dysphonia measures to detect PD or assess the severity of PD in telemonitoring scenarios [64–67]. In one of these studies [64], the authors perform a vast number of acoustic measures in 31 speakers (23 with PD). These measurements include, among others, jitter, shimmer, HNR, NHR, obtained using the MDVP, and other complexity measurements such as D_2 and PPE (a new feature proposed by the authors) to discriminate between parkinsonian and normophonic voices. The proposed PPE is added to the study to quantify the abnormal variations of pitch, measuring the entropy of the probability distribution of the pitch variations in the logarithmic scale (semitones), excluding the smooth vibrato and microtremor, commonly present in normophonic speakers. After a previous feature selection to remove redundant information and determine the number of characteristics for optimal classification, these are used to feed an SVM classifier and determine which pairs of features have better discrimination in the SVM high-dimensional space. Results reveal that complexity (RPDE, Detrended Fluctuation Analysis (DFA), and PPE) and noise (HNR) measurements may have better discriminative properties, but unfortunately, the study only provides graphical results.

Although most of the published works have focused on detection, some other studies have addressed the estimation of the patients' degree of affection based on, mainly, the Unified Parkinson's Disease Rating Scale (UPDRS). This is the most popular scale to quantify the severity of PD in the clinical setting. Originally, the first version of the UPDRS was composed of 42 items, each ranging from 0 (normal) to 4 (severe), summing up a total score of 199. The scale is divided into 4 sections according to the characteristics measured [68]: (i) mentation, behaviour, and mood (0–16 points); (ii) activities of daily living (0–52 points); (iii) motor examination (0–108 points); and (iv) complications of therapy (0–23 points). A newer version of the scale was proposed by the Movement Disorder Society [69], conceived to be more complete, homogeneous, and with better clinimetric properties than the UPDRS scale. This new scale is called Movement Disorder Society-UPDRS (MDS-UPDRS), and like the UPDRS scale, is composed of 4 sections that slightly differ from its predecessor. The MDS-UPDRS includes 65 items summing up a maximum of 260. In fact, some of the voice/speech corpora used for PD assessment were labeled using the original UPDRS definition, whilst others use the new one, and most of them are focused on the section III of the scale. The works [65,66,70] use sustained vowels aiming to estimate UPDRS. In [65], the authors employ measurements such as jitter, shimmer, noise, and complexity, among other features, to predict UPDRS through regression trees in a longitudinal study in which the materials consist of almost 6.000 sustained phonations recorded weekly from 52 patients during 6 months as described in [71]. The Mean Absolute Error (MAE) of the estimated values for motor UPDRS (part III of the rating scale) is below 5.8, which is reduced to 2 in the subsequent study [66]. In [70], the authors report a 7.7 Root Mean Square Error (RMSE) in the automatic estimation of UPDRS, using sustained vowels from 168 patients, features measured with openSMILE [30]. However, the differences in results between [70] and [65,66] are difficultly comparable as the studies use different number of patients, metrics (MAE vs. RMSE), and because the authors in [70] employ a cross-validation strategy considering different speakers in training and testing subsets, which is more conservative than the one employed in [65,66], where the authors randomly select the utterances for the training and testing folds, meaning that both subgroups can contain utterances from the same speaker.

An example of a study employing multiple features and several sustained vowels is [33], where authors analyze several techniques to estimate UPDRS and detect PD employing a corpus with 84 patients and 49 controls. Five vowels (/a:/, /e:/, /i:/, /o:/ and /u:/) uttered by all the speakers at minimum, normal and maximum Sound Pressure Level (SPL) are characterized with 350 features, including jitter, shimmer, noise, complexity, and empirical mode decomposition features among others. After using feature selection and random forests for binary detection, a 90% accuracy is obtained, employing all the vowels at minimum SPL (Leave One Out (LOO) cross-validation) and an MAE of 5.70 for the estimation of UPDRS-III.

Other types of features such as Mel-Frequency Cepstral Coefficients (MFCC) have been used in combination with several machine learning schemes as SVM, random forest, or Least Absolute Shrinkage and Selection Operator (LASSO) regression, providing accuracies near 85% in the binary detection of PD with sustained vowels [72,73]. This configuration has been compared with other traditional features in the same studies, and results suggest that it provides better results than jitter, shimmer, or noise, and similar results than complexity measurements [72] or [73] tunable Q-factor wavelet transform.

In a different approach, authors of [74] use Perceptual Linear Prediction (PLP) feature vectors to characterize sustained vowels of 34 subjects (17 patients). These features are the input of an automatic detector system (SVM), which provides a mean accuracy of 76% in discerning patients from controls. In a subsequent work [75], the same authors obtain an accuracy of 91% in a similar approach using MFCC for the same purposes.

In contrast, other related studies use several corpora and a broad set of state-of-the-art acoustic features such as jitter, shimmer, noise, complexity, PLP, Linear Predictive Coding (LPC) or MFCC, [47,76], and reach accuracy below 75%. In the same sense, the study [59] reports a low correlation between most of these traditional features and classes (PD vs. non-PD) employing a large cohort, compared to previous studies using a lower amount of recordings [48]. Similarly, low correlation between jitter, shimmer and noise and non-speech motor manifestations as well as disease duration are reported in a study with 40 de-novo PD patients [45]. The same study, which employs traditional acoustic features, Cepstral Peak Prominence (CPP) and glottal source features such as Harmonic Richness Factor (HRF) in an SVM-based PD detection system, reports Area Under the Curve (AUC) below 0.75 in a LOO cross-validation scheme. Other similar study, employing glottal source features –complexity measurements obtained from the glottal flow signal– and SVM reports a 75% accuracy in the detection of PD [77].

Furthermore, it is worth noting that from the phonatory works employing machine learning approaches and phonatory features providing accuracy over 90% or high correlations between scores or biomarkers and UPDRS, some of them present certain well known methodological issues, such as using the same speakers in testing and training [78], having a corpus with an average age difference between patients and controls of over ten years [79], or not even specifying demographic differences between groups.

Therefore, although there is evidence of the separability properties of some phonatory features concerning the patient and control groups, the analyzed studies usually consider cohorts with less than 200 patients or do not perform cross-corpora validations to evaluate the accuracy of the automatic detection and assessment systems. In this respect, although the literature reports two corpora with larger cohorts; one of them [80] is quite unbalanced in age, and there is no control over the patients (these have not been enrolled in a hospital and it is not possible to know how reliable their diagnosis is); the studies related to the other large cohort [59] report differences with the conclusions obtained in previous studies (which might be influenced by the channel, among other aspects). Likewise, there has not been defined yet, a de-facto recommended configuration to be used in diagnosis tools.

4. Articulatory aspects

4.1. *Evidence of PD's influence in phoneme articulation*

Connected speech contains fluctuations in vocal characteristics such as voice onset, voice termination, and voice breaks, which are considered crucial in quality of voice and speech evaluation. These characteristics are not fully represented with sustained vowels. Several studies in literature point out to an evident influence of PD in phoneme articulation, and study that influence in some specific sounds or articulatory movements. In one of the earliest studies trying to determine parkinsonian patients's articulatory deficits from a phonetic point of view [13], authors analyze the dysfunctions of the larynx, lips, and tongue (back, tip and blades) in 200 idiopathic and postencephalitic untreated PD patients. In that study, two trained listeners evaluate the patients' speech to perceptually assess their voice quality and the misarticulation of sounds. Results show that 45% of the patients exhibit lingual or labial (or both) abnormalities during articulation. Authors report that the errors are mainly concentrated in obstruent consonants, especially stop-plosives, fricatives, and affricates (in this order), that is, the consonants requiring the greatest narrowing or closure during articulation, with a higher proportion of errors found in velar articulations, mostly /k/ and /g/ sounds (this has been confirmed in subsequent works [81–85,63]). Additionally, the study points out that all patients with articulatory impairments have phonatory dysfunctions but not the contrary, (although other recent works suggest that not all patients with articulatory issues present dysphonia [16]). In a subsequent article [81], authors use the same corpus and report incomplete contact of articulators for plosive stops and partial constrictions for fricatives. One of the main conclusions found in that publication is that the intra-speaker consistency of the errors is near 98%, meaning that when a patient misarticulates a phoneme, this error is repeated during the whole session. However, this study does not assess the longitudinal intra-speaker error consistency as it does not include inter-session recordings. Another critical observation found in [81] is that inter-speaker misarticulation consistency reaches 97%, meaning that the vast majority of the patients produce the same error substitution when a phoneme is misarticulated. In most cases, this error consisted of the substitution of stop sounds by fricatives, in a phenomenon commonly known as *spirantization*.

Concerning the influence of PD in specific speech segments, another article [9] analyzes a DDK task (repetition of the syllable sequence /pa-ta-ka/) in 16 patients and 10 controls, finding remarkable differences between groups in the duration of the voicing segments, which tend to be longer in patients and in the silences before bursts (also called stop silences[2]), that tend to disappear. These observations align with the spirantization effects by which the stops are substituted by fricatives, as fricatives are not preceded by stop silences.

Conversely, other studies using features characterizing the articulatory constriction during consonants with spectral moment coefficients [87] only find subtle differences between consonants from parkinsonian and control groups [88]. To this respect, significantly less contrast between voiced and voiceless Voice Onset Time (VOT) has also been found in PD patients in comparison to control speakers [89] as well as longer VOT in patients [90], which aligns with the significant decrease in lung volume initiation and termination during speech production in PD patients respect to control speakers found in other studies [91]. This is contradictory with the findings exposed in [13,81–84] where a strong influence of the disease is found in certain consonants. To this extent, whereas some of the cited studies point towards a higher influence of PD in speech segments requiring a higher narrowing of the vocal tract, it might be difficult to find clearly defined articulatory patterns common to most PD patients. The reason is that hypokinetic dysarthria can be considered a multidimensional impairment [22] and its manifestations may depend on the individual and PD sub-type (postural instability/gait difficulty and tremor-dominant phenotypes, for instance) [92]. It is unclear if the speech tasks' complexity allows a better separability between PD patients and controls when measuring articulatory features. For instance, the study [93] analyzes segment duration and speech error rates in 15 patients with neurological disorders (8 of them with PD) and 15 controls during the production of syllable sequences, ranging from low to high complexity. The most complex speech tasks were those sequences composed by several syllables, including phonemically similar consonants and low phonotactic probabilities (/ja cha va za tha/) in contrast to the low complexity utterances (/ma ka/). Results suggest that patients present significantly longer vowel duration, but the effects caused by the higher complexity of the sequences –as an increase of the speech error rate and a longer duration of speech measures– are similar in both groups. Authors suggest that "*the effects of phonemic similarity and phonotactic probability are primarily attributable to processing by distinct regions of cortex.*" However, only a small cohort of 8 PD patients was considered.

Finally, other authors also consider that the different auditory patterns between patients and controls contribute to both groups' speech differences [94].

4.2. *Kinematic considerations*

Some studies contemplate the articulators' kinematic properties (mainly related to velocity and acceleration of these) as sources of information about the presence or evolution of PD. For instance, the work [95] studies dysarthria in 12 patients, comparing them to 10 controls by performing a labial kinematics analysis at different speech rates. Results show significant differences between patients and controls in the amplitude and velocity of the lips movement at high speech rate (5 to 7 syllables/s). In comparison, no significant differences were found at low speech rates (3 to 5 syllables/s), suggesting that PD-related hypokinetic dysarthria might vary as a function of the speaking rate.

Several studies [96,95,97–99] report lower amplitude and velocity in the jaw and lower lip opening articulatory gestures for patients compared to controls, as well as reduced formant transitions and duration of vocalic segments. In concordance with this, a significantly

[2] The stop silences or silences before bursts, as described in [86] are the pauses or silences before a burst on speech caused by a total constriction or closure of the vocal tract, especially during the articulation of plosives.

reduced second formant (F2) slope for patients (20 subjects) with respect to controls (5 subjects) has been reported [100].

Other articles indirectly evaluate the velocity and acceleration of articulators through the processing of the speech signal. In the study [86], the authors leverage velocity and acceleration of the speech signal's amplitude envelope during a DDK test to obtain several complexity measurements that allow discriminating patients from controls. In the same sense, another study [76] suggests that the velocity and acceleration of MFCC and PLP features obtained from connected speech (commonly known as derivatives) are key in PD detection when using these coefficients to characterize parkinsonian speech.

4.3. Formant measurements and associated features

Several articles in the literature measure the speech formants and their evolution to assess PD. Some of these approaches employ VSA-related features, that are calculated using the formants from vowels extracted from connected speech [101–105] instead of using sustained vowels as a acoustic material. Since formants reflect, to some extent, the position of the tongue, a reduction of the articulation extension of this articulator can subsequently affect the frequency range of the formants. For instance, [104] presents a comparison of PD detection techniques using only vowel segments extracted from sustained vowels, sentence repetitions, reading passages, and monologues of 20 untreated patients and 15 male control speakers. Significant differences between the two groups are obtained for tVSA and Vowel Articulation Index (VAI) features, which are lower in patients than in controls. Compared to other studies that use only sustained vowels, as [42], this study reports a more significant separability between groups when using vowels from monologues. In this sense, the VSA extracted from articulated vowels has been found to provide significant differences between speakers with and without PD [105,104,101,102,49,106,103,107]. In this respect, some studies suggest that the articulation of the vowel /u/ tends to be more affected by PD than /a/ or /i/ [105,104,108]. Moreover, the changes in vowel acoustics caused by dysarthria reflected in measurements such as the VSA can also be perceived by human listeners [109].

In [101], authors record 14 patients before and after starting the Lee Silvermant Voice Treatment (LSVT), 15 patients without this type of treatment, and 14 control speakers. All three groups were recorded twice, with a similar time window between the two sessions. Three read sentences were recorded from each speaker at each session from where three vowels, /i/, /u/, and /a/ were extracted for analysis. Results indicate that the tVSA is smaller in patients, in concordance with [42, 52], and that it increases towards the control group's values after the LSVT. Nevertheless, the same authors propose an alternative type of measurement based on formants, the Formant Centralization Ratio (FCR) [110], a metric not sensitive to gender, and more effective than tVSA in distinguishing PD dysarthric speakers and controls, according to the study's results. In a very similar way, the study [52] (68 patients and 32 controls) reports lower tVSA in patients, especially in males, although results indicate that the use of the VAI [106], using formants of the three vowels, /i/, /u/ and /a/, provides a better differentiation between speakers with and without PD. The same study suggests that tVSA could be insensitive to mild forms of dysarthria. Additionally, a subsequent study [102] (67 patients and 40 healthy speakers) reports that VAI seems to be superior to tVSA in the characterization of vowel articulation decline as PD evolves.

4.4. Automatic detection and assessment

New studies propose automatic systems to detect or assess PD making use of advanced speech processing and machine learning techniques.

For instance, a study [111] includes an analysis of DDK tasks (/pa-ta-ka/) in a corpus of 24 PD patients and 22 controls, reporting up to 88% accuracy in differentiating speakers with and without PD. In that study, all the utterances were subdivided automatically into syllabic segments, and then, the initial burst, onset, and occlusion positions were detected to analyze articulation. Only 13 features were obtained by performing measurements on the distinct syllabic segments, each feature describing a distinct articulatory trait of speech, such as the coordination of laryngeal and tongue activity or the precision of consonant articulation, among others. This study indicates that speech processing can produce powerful indicators of imprecise consonant articulation in PD-related dysarthria. The transitions between phonemes and between voiced and unvoiced segments have also been employed in the automatic detection of PD using spectrograms and Convolutional Neural Networks (CNN) [112], PLP and GaussianMixtureModel-Universal Background Model (GMM-UBM) [76], i-vectors and Probabilistic Linear Discriminant Analysis (PLDA) [76] or MFCC and Band Bark Energies (BBE) with SVM classifiers [113] yielding in most of the cases accuracies over 85%. The analysis of some of these approaches indicates that individual segments associated with specific articulation movements such as those of plosives, provide more relevant information for the detection PD in comparison to detection schemes that do not select or set the attention to any specific segment [113,63]. Similar studies employing voiced and unvoiced segments separately to detect PD obtain better classification results using the unvoiced segments [114]. More recently, new approaches have leveraged the differences between phones [85] or articulation manners [63] of speakers with and without PD, also proposing automatic PD detectors grounded on speaker recognition technologies and using several corpora, providing up to 94% accuracy detecting the disease, including cross-corpora trials.

Other approaches referred previously in this section, as [86], use the velocity and acceleration of the speech signal's amplitude envelope during a DDK test to obtain several complexity measurements to discriminate patients from controls (50 subjects per group). These features were used to feed an SVM classification scheme providing an accuracy of 85% in the detection of speakers with PD.

On the other hand, from all the analyzed articles, a group is focused on the automatic estimation of severity indices, such as UPDRS [115–125]. For instance, in the study [121], MFCC and BBE features combined with speech intelligibility measures –calculated in terms of word accuracy obtained with the Google automatic speech recognizer– have been used to estimate UPDRS-III, obtaining Spearman's correlations of 0.72 with the patients' scores using three corpora: one in Spanish (50 patients), other in German (88 patients) and other in Czech (20 patients).

Moreover, the study [126] combines a set of acoustic features obtained employing OpenSMILE and other features extracted from the analysis of articulators' displacement with several read sentences, words, and DDK tasks from the GITA corpus (50 patients) [127]. This study's novelty is that authors inversely map the articulators' displacement using the speech signal and a Deep Neural Network (DNN) encoder. Pearson's correlation of 0.51 between UPDR UPDRS and estimated scores (using Support Vector Regression (SVR) and DNN) is reported.

Other studies analyze the correlation between UPDRS and articulatory features. For instance, the study [128] uses the Chi-Squared Distance (CSD) coefficient, reporting that this feature can provide a better correlation (0.78) with the UPDRS speech symptom severity rating (UPDRS-S) than other traditional measurements such as HNR or MFCC, although an analysis of the correlation of all the used features with other UPDRS parts such as the motor section rating (UPDRS-III) would have been of interest.

Furthermore, a recent study [129] uses the phonetic posteriors obtained through the use of a DNN scheme to characterize parkinsonian voices as a blend of non-modal voices, suggesting that PD can be represented by a voice quality spectrum composed of 30% breathy voice, 23% creaky voice, 20% tense voice, 15% falsetto voice and 12% harsh voice.

The variety of classification schemes found in the the literature is large, including GMM-UBM [76], forced GMM-UBM [85], multinomial Naive Bayes [62], SVM [130,86], i-vector based schemes [117,76], k-Nearest Neighbor (k-NN), multi-edit-nearest neighbor [130] or decision trees [131]. Some of the analyzed studies employ neural network structures as [132,133], but although these schemes are popular in current machine learning approaches, in general, the small size of most parkinsonian corpora does not allow to exploit the capabilities of these schemes efficiently. In that sense, some other studies propose transfer learning methodologies to use DNN approaches adapted to small training corpora [134,135].

Regarding feature extraction, although MFCC features are commonly used to feed classification systems, several studies compare their performance with PLP coefficients in identical conditions (same classification schemes and corpora) and suggest that PLP coefficients improve the performance of the systems [76,136].

5. Common features employed for detection or severity assessment of PD

This section lists some of the most common features employed in the automatic detection and severity assessment of PD. Table 1 includes a list of some of the motor feature families and coefficients analyzed in the literature to detect or assess PD. The table includes comments about the significance of these features' discriminative properties and a reference to the studies using them. All the features in the list have been presented in this article, and the most relevant are discussed in the following section.

Additionally, in order to illustrate the discriminant properties of some of the most used features, we show their probability density functions and t-distributed Stochastic Neighbor Embedding (tSNE) [158], in Figs. 2 and 3, respectively, for 43 speakers with PD and 46 healthy controls (HC) from Neurovoz corpus [85,134]. The objective of Figs. 2 and 2 is not to provide detailed analysis or comparison of features but to illustrate the behavior of the most commonly used families in one corpus containing speech from people with PD. We obtained the phonatory features using two sustained vowels /aː/ of 3 s duration from each speaker, and the articulatory features with a DDK task (/pa-ta-ka/). For the sake of reproducibility, we have selected some of the most used feature families that can be calculated with publicly available software such as jitter, shimmer, noise, MFCC, PLP and BBE. We obtained the phonatory features with the standard configuration of the voice feature extraction toolboxes AVCA [58] and DARTH-VAT [59] and calculated the probability density functions employing a Gaussian kernel. Articulatory features were calculated with the AVCA toolbox considering 10 ms frame length and 5 ms shift. The number of MFCC coefficients employed was 13, while the dimensionality of the Rasta-PLP and BBE vectors was 10 and 20, respectively. These articulatory features were complemented with their respective derivatives ($\Delta + \Delta\Delta$). Although both toolboxes can calculate a large amount of coefficients, only those described in Table 1 are considered.

6. Discussion

In this review, we have analyzed different studies dealing with the influence of PD on the phonatory and articulatory aspects of voice and speech and how such influence can help to design new tools to assist the diagnosis and evaluation of the disease. The literature reports a large amount of features and methods to characterize parkinsonian speech. Some studies employ feature vectors with hundreds of measurements that are used as the input of machine learning classifiers to create automatic detection and assessment tools. In this review we have included those methods that have been widely used in literature or that we considered relevant.

Table 1
Phonatory and articulatory feature families referenced in this review, studies including them and a brief description of their associated findings.

Features	Evidence of PD's influence	Studies including the family of features
	Phonatory Aspect	
Jitter	Several studies find higher values of jitter in PD patients groups than in control groups.	[38,39,43,44,54,46,47]
Shimmer	Several studies find higher values of shimmer in PD patients than in controls.	[38,39,43,44,40,32,41,137,49,50,42,64,65,48,138,70,54,139,46,47]
HNR	Several studies suggest that this value is lower in PD groups respect to control groups.	[38,39,53,40,137,50,64,65,48,128,54]
NHR	Several studies suggest that this value is higher in PD groups respect to control groups.	[43,44,41,137,49,42,64,65,48]
GNE	Several studies suggest that this feature provides good results in automatic detection and assessment of PD using machine learning techniques.	[48,140–144]
Subglottal pressure	Results suggest that when PD and control groups are producing the same speech SPL, the subglottal pressure is higher in patients respect to controls	[35,36]
PPE	Results suggest that this value is higher in PD group respect to control group.	[50,64,65,48]
RPDE	Some studies suggest that this value is slightly higher in PD groups respect to control groups.	[50,64,65,48]
Fundam. frequency	Some studies suggest that this value (and its variability) is higher in PD groups respect to control groups	[53,32,41,54,55]
MPT	Some studies suggest that this value is slightly lower in PD groups respect to control groups.	[53,32,50,64,65,48]
D_2	Results suggest that this value is higher in PD groups respect to control groups.	[44,64,65,48]
SPI	Results suggest that this value is higher in PD patients respect to controls.	[41,33,127,32]
Tremor	The analysis of tremor biomarkers suggests that vocal tremor tends to be higher in PD patients respect to controls.	[13,145,34,39,64,41,146,147]
MFCC	These features provide state-of-the-art results in the automatic detection and assessment of PD.	[75,72,73,47,76]
	Articulatory Aspect	
FCR	This feature tends to be higher in groups of PD patients respect to controls.	[110,148–151]
BBE	These features help providing state-of-the-art results in the automatic detection and assessment of PD.	[113,139,152,132,121]
MFCC	These features provide state-of-the-art results in the automatic detection and assessment of PD.	[48,153,113,128,70]
PLP	These features provide state-of-the-art results in the automatic detection and assessment of PD.	[74,76,47,85,63,136,134]
LPC	These features provide some discriminant properties in the automatic detection and assessment of PD.	[76,154]
Duration of voiced segments	This duration tends to be longer in patients.	[9]
Stop silence length	This duration tends to be shorter in patients.	[9,86]

(continued on next page)

Table 1 (*continued*)

Features	Evidence of PD's influence	Studies including the family of features
VAI	Significant reduced values of VAI are found on patients respect to controls.	[52,102,49,104]
VOT	Result suggest that VOT tends to be longer in patients respect to controls.	[40,155,156]
VSA-related	Reduced values of VSA during articulation are found on patients respect to controls.	[105,104,101–103,42,52,139,157]

6.1. General considerations

One of the first observations after the analysis of the literature is that, although the studies focused on each aspect are abundant, the proportion of approaches to detect or assess the disease objectively and automatically is small. In contrast, many of these studies provide evidence and point out trends about the distinct features that can be used as biomarkers or as inputs of new tools for PD diagnosis but do not use them into validated automatic tools. Moreover, one foremost issue is that some of these studies analyzed a small number of patients. In some cases, this number is less than 20, which leads to high uncertainty in the results procured.

The studies performing automatic detection or assessment of PD can be categorized into two major groups depending on the type of speech task that is employed: those using supervised speech tasks and those using unsupervised speech tasks. The supervised speech task approaches are those with constraints in the speech tasks used, requiring the speaker to pronounce a specific sentence, a sustained vowel, or a DDK task, usually under controlled conditions. In this sense, the repetition of the same sentence or phonetic sequence by all speakers simplifies the pronunciation or phonation comparison between groups (patients and controls) and usually provides higher accuracy in automatic detection schemes. For instance, comparing plosive phonemes between groups to find spirantization manifestations in speakers with PD will be more efficient if all speakers pronounce the same sentence with the same number and type of plosives. However, special attention must be paid to recording transcriptions when collecting data. Speakers with PD could be producing more word errors while reading or reciting than controls, as some studies suggest [159], which would lead to a biased phonemic comparison of utterances (i.e., errors would be produced at a cognitive-word level instead of at phonetic-articulatory level). In other words, when trying to find acoustic-articulatory differences between a recording containing the word *cat* and another containing the word *bat* (produced by misreading the word *cat*), the differences found by an automatic detector might be mistakenly attributable to an articulatory issue caused by dysarthria when the real problem was different. Most studies found in literature use supervised speech tasks.

Additionally, when using read speech for evaluation or automatic detection of PD, the influence of multi-task factors while reading must be considered [160]. Moreover, as PD also influences the patient's vision [161], the impact of PD on speech while reading can be caused not only by articulatory deficits but also by visual impairments. For these reasons, in some studies, speakers are asked to listen and repeat some pre-recorded sentences instead of reading texts [85], in order to discard the influence of eye movement impairments, multi-task factors and cognitive-related issues in read speech.

On the other hand, the unsupervised speech task approaches commonly use spontaneous speech, in which there are neither supervising nor phonetic restrictions on the speech task used as the input of the automatic detector. One advantage of unsupervised speech tasks is that they also allow us evaluating cognitive aspects [162,163]. These systems could be used in both clinical and domestic environments and are suitable to monitor the patient's evolution as they can be part of devices recording large speech sequences in a domestic environment. Studies analyzing large sequences are still emerging, and it is possible to find only limited contributions in the literature to this regard.

The acoustic analysis of elicited speech tasks for automatic detection or assessment of PD is almost unexplored to this extent. This type of task could be considered semi-supervised since although the utterances to be recorded will match a particular pattern (the result of a simple math operation or naming an object in a picture), the cognitive load is different from reading a passage or describing a picture. Therefore, including a new type of cognitive load could improve automatic detection and assessment results (and provide some mechanistic insights), especially when eliciting action verbs, as suggested by several articles [164–168].

In the same manner, the approaches can be language-dependent or independent. Usually, phonatory approaches working with sustained vowels could be considered language-independent, although we cannot always assume that a vowel represented by a grapheme has the same associated sound in different languages. For instance, when asking to speakers with different mother tongues to sustain the vowel represented by the grapheme 'a', it might be produced as /ɐː/, /ɑː/ or /aː/ (according to the international phonetic alphabet) depending on the language. Nevertheless, most of the works found in the literature could be language-independent as these do not rely on the use of a specific sound or phoneme that is unique to certain language. Notwithstanding, applying a particular methodology to different languages could provide different results. Moreover, although many articles analyze the dysarthria of speakers with PD and its associated acoustic features in several languages [169], only a few studies analyze the same schemes in several languages simultaneously or perform cross-language experiments (training with a corpus in one language and testing with another corpus in a different language) [27,85,113,63,157,170,133,171]. In general terms, most of these cross-language experiments report PD detection accuracy between 70 and 80%, which indicates that there are common patterns across languages and that single-language approaches provide higher accuracy.

It is important to remark that there are not too many studies analyzing if PD and other motor-related diseases affect voice and speech differently. In this sense, it is unclear if any of the automatic systems proposed in literature will identify subjects as PD speakers when these have other mimicking conditions, as most of these systems are trained with only two classes: speakers with PD and controls without motor-related diseases. This is important when building diagnosis tools for neurodegenerative diseases since clinicians strive to differentiate between PD and other mimicking conditions in many cases. Consequently, practical tools should be multi-class instead of bi-class. In this sense, some works analyze the differences between idiopathic PD and other neurological conditions such as Friedrich Ataxia (FA), Huntington's Disease (HD), or multiple sclerosis, among others. [6,172–178,93,105,179–181,88,182,156,183]. However, these are scarce and, up to date, insufficient to extract solid conclusions to be applied in the clinical practice. New studies in this direction are crucial because, as remarked in [184], there exist common phenomena and differences in the speech production perturbations appearing for the different motor speech disorders. For instance, the study [100] shows a significantly reduced F2 slope for PD patients with respect to controls. Nevertheless, the analysis of F2 in the speech of stroke patients yields the same results in that study. Without a differentiation between the primary neurological impairments on their influence over speech, it will be more challenging to propose practical tools to detect or assess PD. In this regard, clear differences in prosodic patterns between PD and HD [175] as well as FA [178] patients have been reported. Moreover, [156] reports articulatory differences in progressive supranuclear palsy, multiple system atrophy, and PD, especially when analyzing VOT.

Moreover, there are other challenges that machine learning approaches have to overcome to derive into more reliable and causally

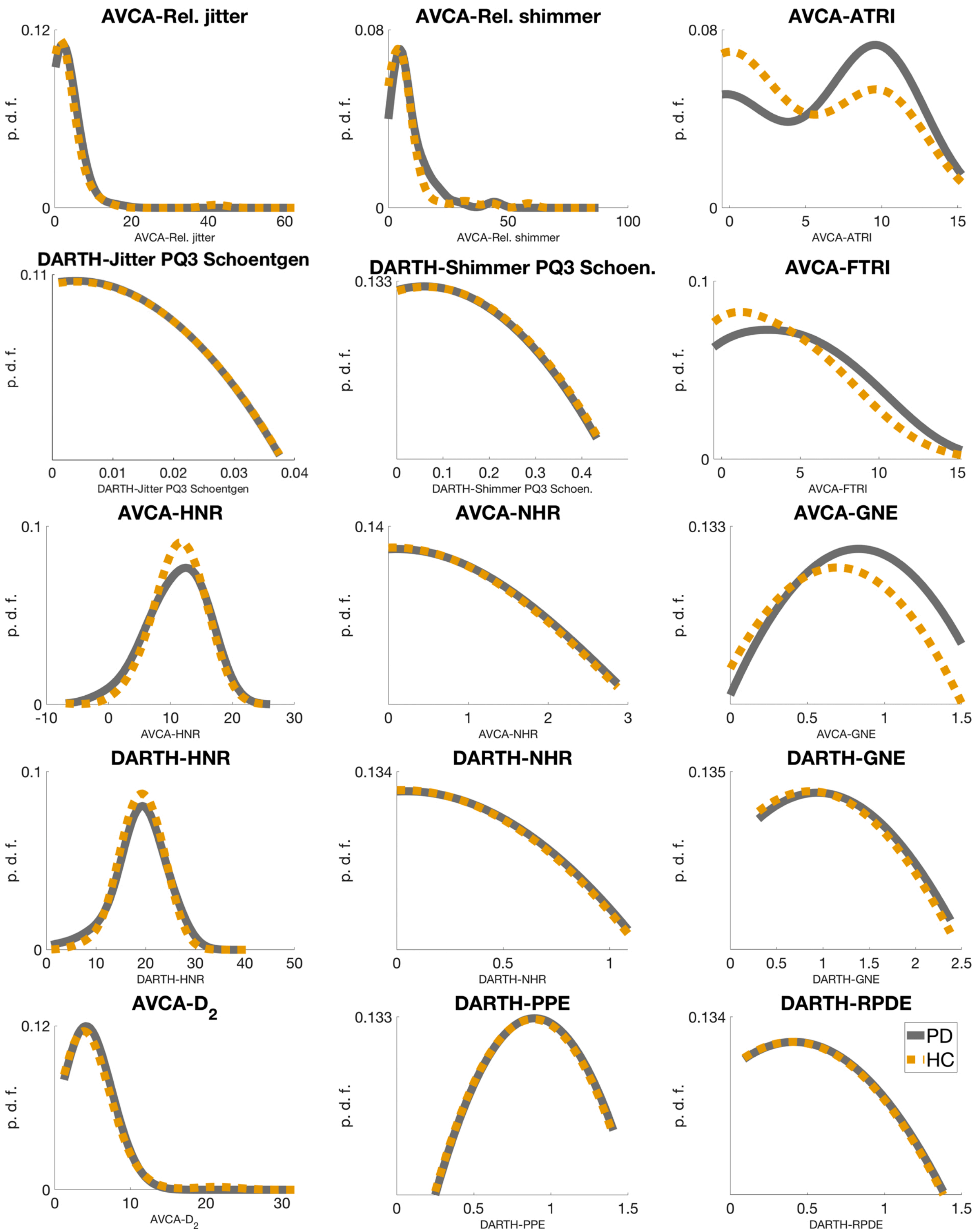

Fig. 2. Probability density functions (p. d. f.) of the average of phonatory features in speakers with PD (solid line) and controls (dashed line). The feature extraction tool referenced in [58] was used to calculate the features labeled as AVCA, while the tool indicated in [59] was employed to calculate the features extracted with the DARTH-VAT. FTRI and ATRI are fundamental frequency and amplitude tremor intensity indices, respectively. Rel. jitter and shimmer are relative jitter and shimmer.

robust diagnosis systems [185]. The vast majority of the analyzed studies proposing new methodologies to detect PD perform associative machine learning and do not consider that the signs found in the analyzed subjects' speech can be caused by factors not related to PD or comorbidities such as depression or dementia. Consequently, discarding causal and counterfactual reasoning in the machine learning process might produce models that lead to poor differential diagnosis performance in real conditions [186].

On the other hand, some studies employ portable devices or mobile phones for speech recording to assess the convenience of the studies for remote patient monitoring [187,114,125,61,188,124,80], although only some compare their results with schemes employing speech

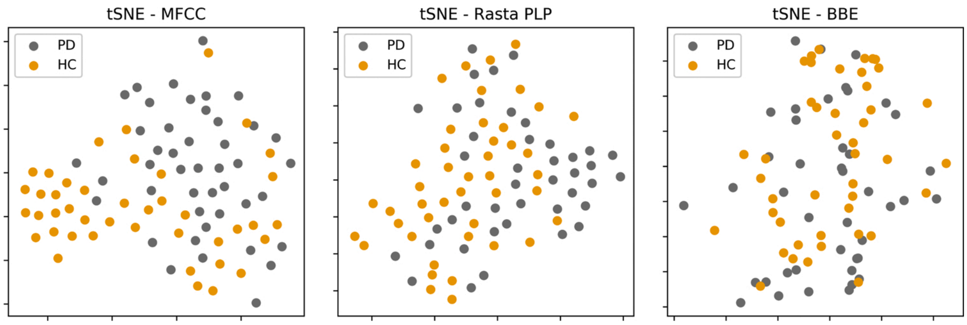

Fig. 3. tSNE plots for the average of articulatory features in speakers with PD (grey) and controls (orange), where each point represents one speaker. The AVCA feature extraction tool referenced in [58] was employed to calculate the articulatory features. The two axes represent the two coefficients from the 2-dimensional embedding calculated with tSNE.

recorded in controlled conditions [61]. The use of smartphones eases performing longitudinal studies, with more patients and recordings, which is crucial since, to date, most of the studies analyze small cohorts. However, this type of study presents certain drawbacks since recordings are noisy, and the acoustic channel conditions change. This variability can be problematic when performing voice quality analysis, as it can lead to substantial differences in the obtained features depending on the smartphone [189,190]. Additionally, if the participants are not enrolled in a clinic program or have a professional diagnosis, the results will have higher uncertainty, as the participants' diagnosis process is not controlled.

6.2. Phonatory approaches

The review of the different phonatory studies allows us to observe that amplitude and frequency perturbations (jitter and shimmer), as well as noise, and their associated features have been widely used along time. Although the literature demonstrates in some cases the influence of PD in jitter and shimmer [38,39,41,40,49,50,42], some works suggest that these features are not good indicators of PD or provide low to moderate differentiation between controls and patients [43,44,32,47,46]. Moreover, some studies suggest that these features are very sensitive to channel variations [190] and require several recordings to minimize intra-speaker variability [191]. In contrast, most of the studies analyzing noise features on parkinsonian voices indicate a more definite relationship between noise and PD's presence or evolution [38,39,64,41,40,48,50,49,42]. The relationship between PD and these features can be partially explained by the rigidity observed in the larynx muscles and its impact on the vocal folds [31,32]. In general terms, several studies point out incomplete closure of the vocal folds in PD patients, producing a breathy voice and, thus, a voice with more noise (lower HNR). In any case, it is difficult to extract precise conclusions regarding some acoustic measurements since different studies include a dissimilar proportion of patients at distinct stages of the disease, various PD sub-types, and treatments. The high variability between the corpora employed in the different studies is one of the main reasons for the diverse conclusions obtained in the different studies, in addition to the use of distinctive methodologies.

In general, caution must be taken when using some acoustic features to detect or assess PD since these are influenced by the disease and other factors, such as age [192,193]. Therefore, these values must be assessed according to the age or the sex of the subject since the normality ranges may vary. Similarly, the existence of any other organic pathology influencing these features must be discarded in advance. Moreover, finding a certain correlation between the variability of a particular acoustic feature and the disease's severity or diagnosis labels might be taken with caution since correlation does not imply causation. A careful analysis of environmental factors or therapy helps to discard correlations that might not be directly caused by the disease.

Complexity measurements or a combination of a large number of features, including cepstral and frequency coefficients (such as MFCC or PLP), have also been employed in automatic detection systems considering phonatory aspects, with unequal results. Therefore, it is difficult to precise which of these features has better discrimination properties. Although there is evidence of the separability properties of some phonatory features, only one study employs a large cohort of patients (1500 patients and 8300 control speakers) [59,194], and the correlation between the analyzed features and the speaker labels provided in that study is weak. This suggests a low relevance of the analyzed features in the differentiation between PD and control speakers, although the large variability of noise conditions and microphones employed in that study, and self-reported labels might be influencing the results –similarly to the studies using the mPower corpus [80]–. Consequently, it is unclear which configuration should be used in automatic detection and assessment systems and its expected accuracy in diagnosis tools employing phonatory aspects. To this respect, new initiatives proposing public challenges to evaluate PD's automatic detection over a corpus, including a large cohort of participants recorded under controlled conditions will be necessary to identify relevant features.

On the other hand, although tremor is a common sign of PD, vocal tremor has not been deeply studied and used in phonatory approaches, and only some works study it with mixed results [13,145,34,39,64,41,146,147,92]. This could be caused by the difficulty of differentiating the parkinsonian tremor from other types of tremor such as microtremor or to the absence of reliable features related to this characteristic. Moreover, the prevalence of vocal tremor in the PD population might be low and not necessarily coincident with the prevalence of limb tremor, according to recent studies [92]. In this sense, tremor features in Fig. 2 (Amplitude Tremor Intensity Index (ATRI) and Fundamental Frequency Tremor Intensity Index (FTRI)) provide better discrimination between speakers with and without PD than the rest of phonatory features when using the sustained vowel $/\mathrm{a}\!:\!/$ in the Neurovoz corpus.

6.3. Articulatory approaches

Some of the analyzed publications related to articulatory movements in speakers with PD indicate that the movement of the structures involved in articulation (lips, tongue, jaw or soft palate) are reduced with respect to standard articulation, fail to contact completely during constriction or in some cases have poor coordination. Therefore, the literature suggests that the combination of these deficiencies can cause irregular or unusual voiced and unvoiced segment boundaries.

Other studies suggest that obstruent consonants, the consonants requiring the greatest narrowing or closure during articulation –plosives and fricatives, especially velars (such as $/\mathrm{g}/$ in "green")–, are the most affected by PD [81–85,63]. In the same sense, several studies find spirantization in patients, being plosives and fricatives the most commonly mistaken. Additionally, a considerable number of articles analyze the

VSA and other features obtained from the vowels in connected speech [101–105], reporting a reduced area for patients, indicating a reduction of the tongue elongation during the articulation of vowels. In this sense, the combination of conclusions of all of these studies suggests that, while obstruent consonants, plosives, fricatives, and vowels tend to provide more information about dysarthria caused by PD, there is not a specific type of phoneme or sound to be used for the development of automatic detectors of PD. Therefore, the speech tasks employed in these systems should be phonetically balanced, to cover as many articulatory movements as possible, but including a certain bias to these types of phonemes.

In the same vein, articles studying the velocity and acceleration of the articulators suggest that some patterns obtained from these kinematic features can reveal PD's presence [95,97–100,86,76]. Provided that one of the most characteristic signs of patients with idiopathic PD is related to impairments in acceleration and velocity of movements (as akinesia, bradykinesia, and hypokinesia), these findings are consistent. However, the velocity and acceleration of articulation based on the speech signal have not been deeply analyzed.

In addition to the influence of the motor impairments associated with PD in speech, one cause of the misarticulation may be related to potential hearing impairment [195] and disturbances in the patient's perception of the duration of occlusion lengths in certain phonemes, as some studies suggest that patients are leaned to perceive occlusion lengths longer than they are, causing errors of identification of phonemes [196]. As an example, in [196], some of the participant patients interpret the German word "boten" –/bo:tn/– (*messengers*) as "boden" — (*floor*). Therefore, one hypothesis is that these perception impairments can contribute to the misarticulation during speech production as patients do not perceive their articulatory errors.

In general, studies implementing automatic detection of PD taking advantage of these observations combining articulatory features and machine learning techniques, typically yield accuracy over 80%. Only a few works perform cross-corpora trials [132,27,63] due to the lack of publicly available corpora in the same language or containing similar speech tasks, which is usually necessary for these types of experiments. However, more cross-corpora analyses are advisable to identify if a method is corpus-dependent and evaluate the influence of different recording conditions/environments.

Regarding the classification techniques used in the existing automatic detection and assessment approaches, although in general, SVM, Gaussian Mixture Model (GMM), decision trees, and i-vector-PLDA are conventional classifiers found in the analyzed articles, it is controversial to judge which technique provides better results since the performance of the classifiers generally depends on the dimensionality and amount of data available to train models. Although it is possible to find some examples successfully employing CNN [133], the use of DNN is usually limited to transfer learning approaches in which the DNN is trained with a sizeable non-parkinsonian corpus and then fine-tuned with a parkinsonian corpus [134,135,197].

6.4. *Phonatory and articulatory aspects comparison*

Most of the analyzed phonatory approaches provide accuracy ranging between 75 and 90%. In the case of articulatory ones, accuracy ranges between 80 and 95%. Additionally, some of the articles including the results of phonatory and articulatory approaches applied to the same corpus [76,47], suggest that the articulatory aspects have better discriminative properties than the phonatory aspects. In this sense, Figs. 2 and 3 indicate that articulatory features might have more discriminative power. Fig. 2 suggests that the probability distribution functions for speakers with and without PD tend to be very similar in many phonatory features. However, Fig. 3 shows clustering of patients and controls, especially for MFCC and Rasta-PLP features.

These differences in results may be motivated by the fact that the signals employed in phonatory approaches (sustained vowels) are much simpler than those used for articulatory analyses, including less variability and a smaller amount of kinetic information. Moreover, methodologies using connected speech can indirectly characterize certain phonatory aspects, since connected speech also contains vowels and sonorant segments.

In any case, the prevalence of the articulatory and phonatory impairments in populations with PD may be different and not coincidental, as the disease's manifestations in voice and speech vary with patients (and PD sub-types) [92]. Consequently, tools based on only one aspect will favor false-negative rates, with respect to approaches considering several aspects. Therefore, the complementarity between articulatory and phonatory information must be analyzed in detail in future studies.

6.5. *Methodological issues of some automatic detection and assessment approaches*

Some of the examined articles provide high accuracy in the automatic detection of PD or biomarkers with significant separability characteristics between PD patients and controls but, at the same time, contain methodological issues that could be biasing the results. Although some of these issues are known in machine learning applications, their frequency is not negligible.

For instance, the use of recordings from a corpus for PD patients and a different corpus for controls [198,199] can influence or bias classifiers to better differentiate between classes (PD vs. no-PD) due to corpora-based differences such as different background noises, sampling frequency, or microphones (i.e., the channel characteristics.) Other articles reporting high recall in automatic detection of PD train and test their classifiers employing a group of patients that is, on average, at least 10 years older than controls [199,79]. These studies do not consider that age influences speech and that their detectors could be discriminating not only by PD but also by age [200]. To assess these possible biases, providing detailed demographic information of the speakers, including at least age, sex, and stage of the disease, is essential. However, some articles do not include any information about the demographic differences between the analyzed groups, which may compromise their conclusions. In this sense, some articles do not report if the patients' are under any pharmacological therapy –such as levodopa or dopamine agonists– or if these are drug-naive. Whereas it is unclear how these therapies affect the phonatory and articulatory aspects, some studies suggest that these medications may affect prosody [22], which would affect the comparison of results between studies. The same can be said about the moment of recording during the day and its relationship with medication intake, which is not always reported. A recording before medication intake might yield different acoustic measurements than one performed 2 h after medication.

A common methodological issue found in some studies determining UPDRS using machine learning approaches and voice or speech is the lack of information about the time lapse between the UPDRS evaluations and the voice and speech recordings. To this respect, some patients present considerable on-off, motor, and non-motor fluctuations [201–204], that may lead to variable UPDRS values in short periods of time (depending on the patient, these fluctuations can take place in days). Therefore, long periods between UPDRS assessment and voice recordings may derive into inconsistent results in studies automatically determining UPDRS, especially when evaluating patients with known severe fluctuations.

Other studies [62,124] do not employ cross-validation strategies necessary to reduce uncertainty when employing a single corpus with a small number of recordings (usually, no more than three hours in total). Some other authors use cross-validation strategies but include recordings from the same speakers in training and testing subsets. This could be biasing the results as the classifiers could be learning the speaker's identity characteristics and linking them to the speakers' labels, rather than detecting aspects associated with the pathology [65, 79].

On the other hand, some articles [137] indicate that sustained vowels provide better results than other types of materials in distinguishing between voices from speakers with and without PD. In this work, phonatory measures such as jitter, shimmer, HNR, NHR, or pitch statistics are calculated from sustained vowels, words, and short sentences. However, the use of jitter, shimmer, and other phonatory features in running speech can be considered unorthodox since these features aim to characterize the stability of the phonation process principally. For instance, as jitter and shimmer measure frequency and amplitude irregularities of phonation related to phonatory problems, respectively, measuring these features in vowels extracted from the connected speech that include strong modulations caused by prosody and articulation might lead to wrong conclusions. Measuring these features in voiceless consonants has not a precise meaning.

6.6. *Corpora and software availability*

Sharing corpora and code would permit the comparison of methodologies and features for PD detection and assessment, leading to more robust conclusions about the usefulness of some of the cited approaches. It will also permit applying the different methodologies in other domains such as different languages, acoustic environments, or ages. However, not many of the studies share their speech recordings or associated features. From all the analyzed corpora, and to the best of our knowledge, only [64,137,80,205,63] have made the speech recordings or extracted features available to the scientific community. In the same manner, only some studies share their code to reproduce results. In this sense, many of the features are commonly used and can be extracted employing speech feature extraction libraries and software such as the Python-based standalone software OpenSmile[3] [30], Python-based library Surfboard[4] [200], standalone software Praat[5] [29] or Parselmouth-Praat[6] that is a library that allows researchers using the Praat algorithms with a Python interface [206]. Others can be extracted from libraries specialized in voice pathology detection such as AVCA[7] [58] or DARTH-VAT[8] [59,66,207], both in Matlab. Nevertheless, in [139], the authors present an open-source software for speech analysis, namely *Neurospeech*,[9] that permits calculating and analyzing several prosodic, phonatory, and articulatory features. This package is grounded on the contributions of many different authors, although some are not acknowledged.

A thorough comparison of the features provided by these toolboxes and applied to publicly available corpora is recommended for the future work.

7. Conclusions

In this study, we have presented a comprehensive review of state of the art on the use of speech analyzing the phonatory and articulatory aspects to support PD diagnosis and evaluation, with a particular emphasis on those studies proposing approaches to detect PD automatically or to assess its severity. In view of the state of the art, and from an analysis of better and worse practices, we also provide recommendations to be followed in the future.

The aforementioned works are the pillars that have supported the research in the field carried out in the last few years. Thus, the authors acknowledge the value of these works and the seminal contributions of many of them. The criticisms done in this paper are enunciated from a positive point of view, not to hide their scientific or technical value, and are expected to serve other researchers to establish new methodological or validation frameworks, which will make the state of the art move forward more consistently.

Although some early works [13,173,208] point out that PD has a more explicit reflection on patients' phonation than in articulation, more recent studies support the idea that articulatory aspects allow to better differentiate between speakers with and without PD. However, even though the disease's influence on these aspects is noticeable in mid to advanced stages, there is no consensus on which aspect is more appropriate to help with differential diagnosis in the early stages.

Regarding the accuracy of the reviewed systems performing automatic detection of PD, most of the approaches usually yield values ranging between 80% and 95%. The studies performing an automatic assessment of PD severity are more scarce than those performing detection, and their metrics are so diverse that it is not possible to establish an average error rate range to represent state of the art in those cases.

Focusing on the phonatory approaches, although multiple studies have used certain acoustic features such as jitter and shimmer, their usefulness in diagnosis is not apparent yet, and new investigations point out that these features cannot help in distinguishing between PD patients and controls. Other feature families, such as noise or complexity, have shown significant differences between groups of patients and controls but are more suitable for automatic detection or assessment approaches rather than being employed as single biomarkers.

Concerning the articulatory methods, VSA features family has also demonstrated functional differentiation between patients and controls since it characterizes a reduction of the articulatory movements' extension while producing vowels in connected speech. Other features such as PLP or MFCC provide good results in automatic detection and assessment tasks when combined with different classification techniques. The reason is that these features, in combination with machine learning techniques, can indirectly characterize changes in the articulation of vowel formants in addition to the articulation of consonants. Although some studies indicate that vowel articulation, obstruent consonants, fricatives, and affricates are the most affected speech segments in speakers with PD, several studies indicate that all manners of articulation can be affected.

Regarding the speech tasks to be employed in PD detection and assessment, some studies point out that read sentences or monologues can help better to obtain an accurate differential diagnosis as those contain more information about different types of movements and require more coordination than other materials such as DDK tasks or sustained vowels [182]. However, these conclusions must not mislead research on avoiding the use of sustained vowels or DDK tasks under certain circumstances, as the selection of the most appropriate material depends on the analyzed aspect and the features to be measured. These two tasks can be performed similarly by patients with distinct mother tongues and can be more suitable in language-independent approaches than spontaneous or reading speech.

Several methodological issues are identified in automatic detection of PD as the use of acoustic content of the same speakers in training and testing subsets or the evaluation of classifiers with corpora in which the age difference between patients and controls is high.

From all the analyzed studies working with phonatory and articulatory aspects, it is possible to identify some missing analyses which should be addressed. For instance, the use of several corpora in the same study will allow the analysis of the proposed approach's generalization properties and reveal if the different methodologies are corpus-dependent. In this respect, corpora and code's public availability is highly relevant since it will permit methodology comparisons and identify of the most reliable features and biomarkers. Moreover, the use of more elicited speech tasks must be explored since these tasks add more cognitive load while limiting the phonetic content, as in reading tasks. Also, the study of more tremor features in phonatory approaches

3 https://www.audeering.com/opensmile/
4 https://github.com/novoic/surfboard
5 https://github.com/novoic/surfboard/
6 https://parselmouth.readthedocs.io/en/stable/
7 https://github.com/jorgomezga/AVCA-ByO
8 https://www.darth-group.com/software
9 https://github.com/jcvasquezc/NeuroSpeech

and the complementarity between phonatory and articulatory aspects in a single study must be assessed.

Regarding the different languages, only a few studies address multilingual systems' creation or analyze the proposed schemes' behavior in more than one language. In general, different languages share some phones or sounds and contain some unique ones, so the articulatory analysis using certain tasks such as DDK containing phones shared by several languages could be advisable in cross-lingual approaches.

Moreover, the methods employing supervised speech tasks for automatic detection or assessment of PD are dependant on the specific phonemic content contained in the utterances, procuring a more detailed comparison between patients and controls.

Finally, although there are studies analyzing the voice and speech of PD patients since more than 50 years ago, to the best of our knowledge, literature does not report conclusive clinical validations of the automatic detection systems developed using the voice and/or speech as the input source for the diagnosis. This fact may be attributable to the lack of tools providing clear indicators but not because speech cannot be the right candidate as a source of information during diagnosis.

Authors' contributions

Laureano Moro-Velazquez: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Roles/Writing – original draft; Writing – review & editing.

Jorge A. Gomez-Garcia: Data curation; Resources; Software; Validation; Visualization; Roles/Writing – original draft; Writing – review & editing.

Julian D. Arias-Londoño: Formal analysis; Investigation; Validation; Writing – review & editing.

Najim Dehak: Formal analysis; Funding acquisition; Project administration; Resources; Supervision; Writing – review & editing.

Juan I. Godino-Llorente: Conceptualization; Formal analysis; Funding acquisition; Project administration; Resources; Supervision; Writing – review & editing.

Acknowledgements

This work was supported by the Ministry of Economy and Competitiveness of Spain under grant DPI2017-83405-R1.

Declaration of Competing Interest

The authors report no declarations of interest.

References

[1] R.F. Pfeiffer, Z.K. Wszolek, M. Ebadi, Parkinson's Disease, CRC Press, 2013.
[2] A.J. Hughes, S.E. Daniel, Y. Ben-Shlomo, A.J. Lees, The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service, Brain 125 (2002) 861–870.
[3] J.R. Duffy, Motor Speech Disorders: Substrates, Differential Diagnosis, and Management, Elsevier Health Sciences, 2013.
[4] H.A. Kaplan, S. Machover, A. Rabiner, A study of the effectiveness of drug therapy in parkinsonism, J. Nerv. Ment. Dis 119 (1954) 398–411.
[5] M. Sarno, Speech impairment in parkinson's disease, Arch. Phys. Med. Rehabil. 49 (1968) 269–275.
[6] F.L. Darley, A.E. Aronson, J.R. Brown, Differential diagnostic patterns of dysarthria, J. Speech Lang. Hear. Res. 12 (1969) 246.
[7] F.L. Darley, A.E. Aronson, J.R. Brown, Clusters of deviant speech dimensions in the dysarthrias, J. Speech Lang. Hear. Res. 12 (1969) 462–496.
[8] H. Ackermann, W. Ziegler, Articulatory deficits in parkinsonian dysarthria: an acoustic analysis, J. Neurol. Neurosurg. Psychiatry 54 (1991) 1093–1098.
[9] J. Kegl, H. Cohen, H. Poizner, Articulatory consequences of parkinson's disease: perspectives from two modalities, Brain Cogn. 40 (1999) 355–386.
[10] P. Blanchet, G. Snyder, Speech rate deficits in individuals with parkinson's disease: a review of the literature, J. Med. Speech Lang. Pathol. 17 (2009) 1–7.
[11] S. Polychronis, F. Niccolini, G. Pagano, T. Yousaf, M. Politis, Speech difficulties in early de novo patients with parkinson's disease, Parkinsonism Relat. Disord. (2019).
[12] J. Müller, G.K. Wenning, M. Verny, A. McKee, K.R. Chaudhuri, E.A. Jellinger, Progression of dysarthria and dysphagia in postmortem-confirmed parkinsonian disorders, Arch. Neurol. 58 (2001) 259.
[13] J.A. Logemann, H.B. Fisher, B. Boshes, E.R. Blonsky, Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients, J. Speech Hear. Disord. 43 (1978) 47–57.
[14] L. Hartelius, P. Svensson, Speech and swallowing symptoms associated with parkinson's disease and multiple sclerosis: a survey, Folia Phoniatr. Logop. 46 (1994) 9–17.
[15] A.K. Ho, R. Iansek, C. Marigliani, J.L. Bradshaw, S. Gates, Speech impairment in a large sample of patients with parkinson's disease, Behav. Neurol. 11 (1999) 131–137.
[16] J. Rusz, C. Bonnet, J. Klempíř, T. Tykalová, E. Baborová, M. Novotný, A. Rulseh, E. Růžička, Speech disorders reflect differing pathophysiology in parkinson's disease, progressive supranuclear palsy and multiple system atrophy, J. Neurol. 262 (2015) 992–1001.
[17] P. Enderby, Frenchay dysarthria assessment, Brit. J. Disord. Commun. 15 (1980) 165–173.
[18] L.O. Ramig, C. Fox, S. Sapir, Speech treatment for parkinson's disease, Expert Rev. Neurother. 8 (2008) 297–309.
[19] M. Trail, C. Fox, L.O. Ramig, S. Sapir, J. Howard, E.C. Lai, Speech treatment for parkinson's disease, NeuroRehabilitation 20 (2005) 205–221.
[20] G.M. Schulz, M.K. Grant, Effects of speech therapy and pharmacologic and surgical treatments on voice and speech in parkinson's disease: a review of the literature, J. Commun. Disord. 33 (2000) 59–88.
[21] S. Skodda, Effect of deep brain stimulation on speech performance in parkinson's disease, Parkinson's Dis. 2012 (2012).
[22] L. Brabenec, J. Mekyska, Z. Galaz, I. Rektorova, Speech disorders in parkinson's disease: early diagnostics and effects of medication and brain stimulation, J. Neural Transm. 124 (2017) 303–334.
[23] S. Skodda, U. Schlegel, Speech rate and rhythm in parkinson's disease, Mov. Disord. 23 (2008) 985–992.
[24] D. Ingvar, Serial aspects of language and speech related to prefrontal cortical activity. a selective review, Hum. Neurobiol. 2 (1983) 177–189.
[25] L.C. Kwan, T.L. Whitehill, Perception of speech by individuals with parkinson's disease: a review, Parkinson's Dis. 2011 (2011).
[26] C. Atkinson-Clement, J. Sadat, S. Pinto, Behavioral treatments for speech in parkinson's disease: meta-analyses and review of the literature, Neurodegener. Dis. Manag. 5 (2015) 233–248.
[27] L. Moro-Velazquez, J. Gomez-Garcia, J.I. Godino-Llorente, J. Rusz, S. Skodda, F. Grandas, J.-M. Velazquez, J.R. Orozco-Arroyave, E. Noth, N. Dehak, Study of the automatic detection of parkison's disease based on speaker recognition technologies and allophonic distillation, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, pp. 1404–1407.
[28] K. Elemetrics, Multi-Dimensional Voice Program (MDVP).[Computer program.], Pine Brook, NJ: Author, 1993.
[29] P. Boersma, Praat: Doing Phonetics By Computer, 2006. http://www.praat.org/.
[30] F. Eyben, M. Wollmer, B. Schuller, Opensmile: the munich versatile and fast open-source audio feature extractor, in: Proceedings of the 18th ACM international conference on Multimedia, ACM, 2010, pp. 1459–1462.
[31] D.G. Hanson, B.R. Gerratt, P.H. Ward, Cinegraphic observations of laryngeal function in parkinson's disease, Laryngoscope 94 (1984) 348–353.
[32] I. Midi, M. Dogan, M. Koseoglu, G. Can, M. Sehitoglu, D. Gunal, Voice abnormalities and their relation with motor dysfunction in parkinson's disease, Acta Neurol. Scand. 117 (2008) 26–34.
[33] J. Mekyska, Z. Galaz, Z. Mzourek, Z. Smekal, I. Rektorova, I. Eliasova, M. Kostalova, M. Mrackova, D. Berankova, M. Faundez-Zanuy, et al., Assessing progress of parkinson's disease using acoustic analysis of phonation, in: Bioinspired Intelligence (IWOBI), 2015 4th International Work Conference on, IEEE, 2015, pp. 111–118.
[34] K.S. Perez, L.O. Ramig, M.E. Smith, C. Dromey, The parkinson larynx: tremor and videostroboscopic findings, J. Voice 10 (1996) 354–361.
[35] J. Jiang, T. O'Mara, H.-J. Chen, J.I. Stern, D. Vlagos, D. Hanson, Aerodynamic measurements of patients with parkinson's disease, J. Voice 13 (1999) 583–591.
[36] M.J. Hammer, S.M. Barlow, Laryngeal somatosensory deficits in parkinsonś disease: implications for speech respiratory and phonatory control, Exp. Brain Res. 201 (2010) 401–409.
[37] P.H. Dejonckere, P. Bradley, P. Clemente, G. Cornut, L. Crevier-Buchman, G. Friedrich, P. Van De Heyning, M. Remacle, V. Woisard, A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques, Eur. Arch. Oto-Rhino-L 258 (2001) 77–82.
[38] L.A. Ramig, I.R. Titze, R.C. Scherer, S.P. Ringel, Acoustic analysis of voices of patients with neurologic disease: rationale and preliminary data, Ann. Otol. Rhinol. Laryngol. 97 (1988) 164–172.
[39] F.J. Jiménez-Jiménez, J. Gamboa, A. Nieto, J. Guerrero, M. Orti-Pareja, J. A. Molina, E. García-Albea, I. Cobeta, Acoustic voice analysis in untreated patients with parkinson's disease, Parkinsonism Relat. Disord. 3 (1997) 111–116.
[40] J. Rusz, R. Cmejla, H. Ruzickova, E. Ruzicka, Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease, J. Acoust. Soc. Am. 129 (2011) 350–367.
[41] Y. Tanaka, M. Nishio, S. Niimi, Vocal acoustic characteristics of patients with parkinson's disease, Folia Phoniatr. Logop. 63 (2011) 223–230.
[42] Y. Bang, K. Min, Y. Sohn, S. Cho, Acoustic characteristics of vowel sounds in patients with parkinson's disease, NeuroRehabilitation (2013).

[43] J.H. Rhonda, J.M. Oates, J.P. Debbie, J.H. Andrew, Voice characteristics in the progression of parkinson's disease, Int. J. Lang. Commun. Disord. 35 (2000) 407–418.
[44] D.A. Rahn, M. Chou, J.J. Jiang, Y. Zhang, Phonatory impairment in parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis, J. Voice 21 (2007) 64–71.
[45] M. Novotný, P. Dušek, I. Daly, E. Růžička, J. Rusz, Glottal source analysis of voice deficits in newly diagnosed drug-naïve patients with parkinson's disease: Correlation between acoustic speech characteristics and non-speech motor performance, Biomed. Signal Process. Control 57 (2020) 101818.
[46] J. Schoentgen, A. Kacha, F. Grenez, Joint analysis of vocal jitter, flutter and tremor in vowels sustained by normophonic and parkinson speakers. Models and Analysis of Vocal Emissions for Biomedical Applications, 2019, 37.
[47] L. Moro-Velázquez, J.A. Gómez-García, N. Dehak, J.I. Godino-Llorente, Analysis of phonatory features for the automatic detection of parkinson's disease in two different corpora. Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), 2019, 33.
[48] 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 Trans. Biomed. Eng. 59 (2012) 1264–1271.
[49] S. Skodda, W. Grönheit, N. Mancinelli, U. Schlegel, Progression of voice and speech impairment in the course of parkinson's disease: a longitudinal study, Parkinson's Dis. 2013 (2013).
[50] J. Rusz, R. Cmejla, H. Ruzivckova, J. Klempivr, V. Majerova, J. Picmausova, J. Roth, E. Ruvzivcka, Evaluation of speech impairment in early stages of parkinson's disease: a prospective study with the role of pharmacotherapy, J. Neural Transm. 120 (2013) 319–329.
[51] I. Hertrich, H. Ackermann, Gender-specific vocal dysfunctions in parkinson's disease: electroglottographic and acoustic analyses, Ann. Otol. Rhinol. Laryngol. 104 (1995) 197–202.
[52] S. Skodda, W. Visser, U. Schlegel, Vowel articulation in parkinson's disease, J. Voice 25 (2011) 467–472.
[53] A. Yücetürk, H. Yılmaz, M. Eğrilmez, S. Karaca, Voice analysis and videolaryngostroboscopy in patients with parkinson's disease, Eur. Arch. Otorhinolaryngol. 259 (2002) 290–293.
[54] F. Majdinasab, S. Karkheiran, M. Soltani, N. Moradi, G. Shahidi, Relationship between voice and motor disabilities of parkinson's disease, J. Voice 30 (2016) 768–e17.
[55] D. Abur, R.A. Lester-Smith, A. Daliri, A.A. Lupiani, F.H. Guenther, C.E. Stepp, Sensorimotor adaptation of voice fundamental frequency in parkinson's disease, PloS one 13 (2018) e0191839.
[56] F. Mollaei, D.M. Shiller, S.R. Baum, V.L. Gracco, Sensorimotor control of vocal pitch and formant frequencies in parkinson's disease, Brain Res. 1646 (2016) 269–277.
[57] L. Moro-Velazquez, N. Dehak, A review of the use of prosodic aspects of speech for the automatic detection and assessment of parkinson's disease, Automatic Assessment of Parkinosnian Speech. Communications in Computer and Information Science (CCIS) Series, 1295 (2020) 53–73.
[58] J. Gómez-García, L. Moro-Velázquez, J.D. Arias-Londono, J.I. Godino-Llorente, On the design of automatic voice condition analysis systems. part iii: review of acoustic modelling strategies, Biomed. Signal Process. Control Rev. (2019).
[59] A. Tsanas, S. Arora, Biomedical speech signal insights from a large scale cohort across seven countries: the parkinson's voice initiative study. Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), 2019, pp. 45–48.
[60] P. Khojasteh, R. Viswanathan, B. Aliahmad, S. Ragnav, P. Zham, D. Kumar, Parkinson's disease diagnosis based on multivariate deep features of speech signal, in: 2018 IEEE Life Sciences Conference (LSC), IEEE, 2018, pp. 187–190.
[61] J. Rusz, J. Hlavnička, T. Tykalová, M. Novotný, P. Dušek, K. Šonka, E. Růžička, Smartphone allows capture of speech abnormalities associated with high risk of developing parkinson's disease, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (2018) 1495–1507.
[62] D. Meghraoui, B. Boudraa, T. Merazi-Meksen, M. Boudraa, Parkinson's disease recognition by speech acoustic parameters classification. Modelling and Implementation of Complex Systems, Springer, 2016, pp. 165–173.
[63] L. Moro-Velazquez, J.A. Gomez-Garcia, J.I. Godino-Llorente, F. Grandas-Perez, S. Shattuck-Hufnagel, V. Yagüe-Jimenez, N. Dehak, Phonetic relevance and phonemic grouping of speech in the automatic detection of parkinson's disease, Sci. Rep. 9 (2019) 1–16.
[64] M.A. Little, P.E. McSharry, E.J. Hunter, J. Spielman, L.O. Ramig, et al., Suitability of dysphonia measurements for telemonitoring of parkinson's disease, IEEE Trans. Biomed. Eng. 56 (2009) 1015–1022.
[65] A. Tsanas, M. Little, Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests, IEEE Trans, Biomed. Eng. 57 (2010) 884–893.
[66] A. Tsanas, M.A. Little, P.E. McSharry, L.O. Ramig, Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average parkinson's disease symptom severity, J. R. Soc. Interface 8 (2011) 842–855.
[67] S. Arora, V. Venkataraman, A. Zhan, S. Donohue, K. Biglan, E. Dorsey, M. Little, Detecting and monitoring the symptoms of parkinson's disease using smartphones: a pilot study, Parkinsonism Relat. Disord. 21 (2015) 650–653.
[68] R. Bhidayasiri, P. Martinez-Martin, Clinical Assessments in Parkinson's Disease: Scales and Monitoring, 1 ed., vol. 132, Elsevier Inc., 2017 https://doi.org/10.1016/bs.irn.2017.01.001.
[69] C. Goetz, B. Tilley, S. Shaftman, G. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. Stern, R. Dodel, B. Dubois, R. Holloway, J. Jankovic, J. Kulisevsky, A. Lang, A. Lees, S. Leurgans, P. LeWitt, D. Nyenhuis, C. Olanow, O. Rascol, A. Schrag, J. Teresi, J. van Hilten, N. LaPelle, Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results, Movement Disorders 23 (2008) 2129–2170, https://doi.org/10.1002/mds.22340.
[70] A. Bayestehtashk, M. Asgari, I. Shafran, J. McNames, Fully automated assessment of the severity of parkinson's disease from speech, Comput. Speech. Lang. 29 (2015) 172–185.
[71] C.G. Goetz, G.T. Stebbins, D. Wolff, W. DeLeeuw, H. Bronte-Stewart, R. Elble, M. Hallett, J. Nutt, L. Ramig, T. Sanger, et al., Testing objective measures of motor impairment in early parkinson's disease: feasibility study of an at-home testing device, Mov. Disord. 24 (2009) 551–556.
[72] Y. Camnos-Roca, F. Calle-Alonso, C.J. Perez, L. Naranjo, Computational diagnosis of parkinson's disease from speech based on regularization methods, in: 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, 2018, pp. 1127–1131.
[73] C.O. Sakar, G. Serbes, A. Gunduz, H.C. Tunc, H. Nizam, B.E. Sakar, M. Tutuncu, T. Aydin, M.E. Isenkul, H. Apaydin, A comparative analysis of speech signal processing algorithms for parkinson's disease classification and the use of the tunable q-factor wavelet transform, Appl. Soft Comput. 74 (2019) 255–263.
[74] A. Benba, A. Jilbab, A. Hammouch, Voice analysis for detecting persons with parkinson's disease using PLP and VQ, J. Theor. Appl. Inf. Technol. 70 (2014).
[75] A. Benba, A. Jilbab, A. Hammouch, Detecting patients with parkinson's disease using mel frequency cepstral coefficients and support vector machines, Int. J. Electr. Eng. Inform. 7 (2015) 297.
[76] L. Moro-Velazquez, J.A. Gomez-Garcia, J.I. Godino-Llorente, J. Villalba, J. R. Orozco-Arroyave, N. Dehak, Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect parkinson's disease, Appl. Soft Comput. 62 (2018) 649–666.
[77] E. Belalcazar-Bolanos, J. Arias-Londono, J. Vargas-Bonilla, J. Orozco-Arroyave, Nonlinear glottal flow features in parkinson's disease detection, in: 2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA), IEEE, 2015, pp. 1–6.
[78] T.J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D.C. Atkins, R.H. Ghomi, Parkinson's disease diagnosis using machine learning and voice, in: 2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), IEEE, 2018, pp. 1–7.
[79] S. Singh, W. Xu, Robust detection of parkinson's disease using harvested smartphone voice data: a telemedicine approach, Telemed. J. E Health (2019).
[80] B.M. Bot, C. Suver, E.C. Neto, M. Kellen, A. Klein, C. Bare, M. Doerr, A. Pratap, J. Wilbanks, E.R. Dorsey, et al., The mpower study, parkinson disease mobile data collected using researchkit, Sci. Data 3 (2016) 160011.
[81] J.A. Logemann, H.B. Fisher, Vocal tract control in parkinson's disease, J. Speech Hear. Disord. 46 (1981) 348.
[82] G. Weismer, M. McNeil, Articulatory characteristics of parkinsonian dysarthria: segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination. The Dysarthrias: Physiology, Acoustics, Perception, Management, 1984, pp. 101–130.
[83] J.A. Robbins, J.A. Logemann, H.S. Kirshner, Swallowing and speech production in parkinson's disease, Ann. Neurol. 19 (1986) 283–287.
[84] E.Q. Wang, L.V. Metman, R.A. 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, J. Med. Speech Lang. Pathol. 14 (2006) 323.
[85] L. Moro-Velazquez, J.A. Gomez-Garcia, J.I. Godino-Llorente, J. Villalba, J. Rusz, S. Shattuck-Hufnagel, N. Dehak, A forced gaussians based methodology for the differential evaluation of parkinson's disease by means of speech processing, Biomed. Signal Process. Control. 48 (2019) 205–220.
[86] J. Godino-Llorente, S. Shattuck-Hufnagel, J. Choi, L. Moro-Velázquez, J. Gómez-García, Towards the identification of idiopathic parkinson's disease from the speech. new articulatory kinetic biomarkers, PloS one 12 (2017) e0189583.
[87] K. Forrest, G. Weismer, P. Milenkovic, R.N. Dougall, Statistical analysis of word-initial voiceless obstruents: preliminary data, J. Acoust. Soc. Am. 84 (1988) 115–123.
[88] K. Tjaden, V. Martel-Sauvageau, D.J. R, B.D. R, G.L. A, M.R. M, S. M, Consonant acoustics in parkinson's disease and multiple sclerosis: comparison of clear and loud speaking conditions, Am. J. Speech Lang. Pathol. 26 (2017) 569.
[89] J.A. Whitfield, A. Reif, A.M. Goberman, Voicing contrast of stop consonant production in the speech of individuals with parkinson disease on and off dopaminergic medication, Clin. Linguist. Phonet. 32 (2018) 587–594.
[90] M. Pettorino, W. Gu, P. Pólrola, P. Fan, Rhythmic characteristics of parkinsonian speech: a study on mandarin and polish, INTERSPEECH (2017) 3172–3176.
[91] J.E. Huber, M. Darling-White, Longitudinal changes in speech breathing in older adults with and without parkinson's disease, in: Seminars in Speech and Language, vol. 38, Thieme Medical Publishers, 2017, pp. 200–209.
[92] T. Tykalová, J. Rusz, J. Švihlík, S. Bancone, A. Spezia, M.T. Pellecchia, Speech disorder and vocal tremor in postural instability/gait difficulty and tremor dominant subtypes of parkinson's disease, J. Neural Transm. 127 (2020) 1295–1304.
[93] K.J. Reilly, K.A. Spencer, Sequence complexity effects on speech production in healthy speakers and speakers with hypokinetic or ataxic dysarthria, PLoS One 8 (2013).
[94] K. De Keyser, P. Santens, A. Bockstael, D. Botteldooren, D. Talsma, S. De Vos, M. Van Cauwenberghe, F. Verheugen, P. Corthals, M. De Letter, The relationship between speech production and speech perception deficits in parkinson's disease, J. Speech Lang. Hear. Res. 59 (2016) 915–931.

[95] M.P. Caligiuri, The influence of speaking rate on articulatory hypokinesia in parkinsonian dysarthria, Brain Lang. 36 (1989) 493–502.
[96] K. Forrest, G. Weismer, G.S. Turner, Kinematic, acoustic, and perceptual analyses of connected speech produced by parkinsonian and normal geriatric adults, J. Acoust. Soc. Am. 85 (1989) 2608–2622.
[97] P. Svensson, C. Henningson, S. Karlsson, Speech motor control in parkinson's disease: a comparison between a clinical assessment protocol and a quantitative analysis of mandibular movements, Folia Phoniatr. Logop. 45 (1993) 157–164.
[98] K. Forrest, G. Weismer, Dynamic aspects of lower lip movement in parkinsonian and neurologically normal geriatric speakers' production of stress, J. Speech Lang. Hear. Res. 38 (1995) 260–272.
[99] A.B. Walsh, Basic parameters of articulatory movements and acoustics in individuals with parkinson's disease, Mov. Disord. 27 (2012) 843–850.
[100] Y. Kim, G. Weismer, R.D. Kent, J.R. Duffy, Statistical models of F2 slope in relation to severity of dysarthria, Folia Phoniatr. Logop. 61 (2009) 329–335.
[101] S. Sapir, J.L. Spielman, L.O. Ramig, B.H. Story, C. Fox, Effects of intensive voice treatment (the lee silverman voice treatment [lsvt]) on vowel articulation in dysarthric individuals with idiopathic parkinson disease: acoustic and perceptual findings, J. Speech Lang. Hear. Res. 50 (2007) 899–912.
[102] S. Skodda, W. Grönheit, U. Schlegel, Impairment of vowel articulation as a possible marker of disease progression in parkinson's disease, PLoS One 7 (2012) e32132.
[103] J. Whitfield, A. Goberman, Articulatory acoustic vowel space: application to clear speech in individuals with parkinson's disease, J. Commun. Disord. (2014).
[104] J. Rusz, R. Cmejla, T. Tykalova, H. Ruzickova, J. Klempir, V. Majerova, J. Picmausova, J. Roth, E. Ruzicka, Imprecise vowel articulation as a potential early marker of parkinson's disease: effect of speaking task, J. Acoust. Soc. Am. 134 (2013) 2171–2181.
[105] K. Tjaden, J. Lam, G. Wilding, Vowel acoustics in parkinson's disease and multiple sclerosis: comparison of clear, loud, and slow speaking conditions, J. Speech Lang. Hear. Res. 56 (2013) 1485–1502.
[106] N. Roy, S.L. Nissen, C. Dromey, S. Sapir, Articulatory changes in muscle tension dysphonia: evidence of vowel space expansion following manual circumlaryngeal therapy, J. Commun. Disord. 42 (2009) 124–135.
[107] S.-C. Hsu, Y. Jiao, M.J. McAuliffe, V. Berisha, R.-M. Wu, E.S. Levy, Acoustic and perceptual speech characteristics of native mandarin speakers with parkinson's disease, J. Acoust. Soc. Am. 141 (2017) EL293–EL299.
[108] L. Moro-Velazquez, J.J. Cho, S. Watanabe, M.A. Hasegawa-Johnson, O. Scharenborg, H. Kim, N. Dehak, Study of the performance of automatic speech recognition systems in speakers with parkinson's disease, Interspeech 2019 (2019).
[109] K.L. Lansford, J.M. Liss, Vowel acoustics in dysarthria: mapping to perception, J. Speech Lang. Hear. Res. (2014).
[110] S. Sapir, L. Ramig, J. Spielman, C. Fox, Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech, J. Speech Lang. (2010).
[111] M. Novotny, J. Rusz, R. mejla, E. Rka, Automatic evaluation of articulatory disorders in parkinson s disease, IEEE Trans. Audio Speech Lang. Process. 22 (2014) 1366–1378.
[112] J. Vásquez-Correa, T. Arias-Vergara, J. Orozco-Arroyave, E. Nöth, A multitask learning approach to assess the dysarthria severity in patients with parkinson's disease, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2018, pp. 456–460.
[113] 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, J. Acoust. Soc. Am. 139 (2016) 481.
[114] J.C. Vásquez-Correa, T. Arias-Vergara, J.R. Orozco-Arroyave, J.F. Vargas-Bonilla, J.D. Arias-Londoño, E. Nöth, Automatic detection of parkinson's disease from continuous speech recorded in non-controlled noise conditions, Sixteenth Annual Conference of the International Speech Communication Association (2015).
[115] A. Frid, E.J. Safra, H. Hazan, L.L. Lokey, D. Hilu, L. Manevitz, L.O. Ramig, S. Sapir, Computational diagnosis of parkinson's disease directly from natural speech using machine learning techniques, in: 2014 IEEE International Conference on Software Science, Technology and Engineering, IEEE, 2014, pp. 50–53.
[116] T. Grosz, R. Busa-Fekete, G. Gosztolya, Assessing the degree of nativeness and parkinson's condition using gaussian processes and deep rectifier neural networks, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2015, pp. 919–923.
[117] G. An, D.G. Brizan, M. Ma, M. Morales, A.R. Syed, A. Rosenberg, Automatic recognition of unified parkinson's disease rating from speech with acoustic, i-vector and phonotactic features, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2015, pp. 508–512.
[118] J. Kim, M. Nasir, R. Gupta, M. Segbroeck, D. Bone, M. Black, Z.I. Skordilis, Z. Yang, P. Georgiou, S. Narayanan, Automatic estimation of parkinson's disease severity from diverse speech tasks, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2015, pp. 914–918.
[119] S. David, G. Kiss, K. Vicsi, Estimating the severity of parkinson's disease from speech using linear regression and database partitioning, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2015.
[120] A. Zlotnik, J.M. Montero, R. San-Segundo, A. Gallardo-Antolin, Random forest-based prediction of parkinson's disease progression using acoustic, ASR and intelligibility features, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2015.
[121] J.R. Orozco-Arroyave, J. Vasquez-Correa, F. Honig, J.D. Arias-Londono, J. Vargas-Bonilla, S. Skodda, J. Rusz, E. Noth, Towards an automatic monitoring of the neurological state of parkinson's patients from speech, in: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP, IEEE, 2016, pp. 6490–6494.
[122] T. Arias-Vergara, J.C. Vásquez-Correa, J.R. Orozco-Arroyave, J.F.V. Bonilla, E. Nöth, Parkinson's disease progression assessment from speech using gmm-ubm, INTERSPEECH (2016) 1933–1937.
[123] Z. Galaz, Z. Mzourek, J. Mekyska, Z. Smekal, T. Kiska, I. Rektorova, J.R. Orozco-Arroyave, K. Daoudi, Degree of parkinson's disease severity estimation based on speech signal processing, in: 2016 39th International Conference on Telecommunications and Signal Processing (TSP), IEEE, 2016, pp. 503–506.
[124] H. Ramezani, H. Khaki, E. Erzin, O.B. Akan, Speech features for telemonitoring of parkinson's disease symptoms, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017, pp. 3801–3805.
[125] T. Arias-Vergara, P. Klumpp, J.C. Vásquez-Correa, J.R. Orozco-Arroyave, E. Nöth, Parkinson's disease progression assessment from speech using a mobile device-based application, in: International Conference on Text, Speech, and Dialogue, Springer, 2017, pp. 371–379.
[126] S.U. Hahm, J.U. Wang, Parkinson's condition estimation using speech acoustic and inversely mapped articulatory data, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, volume 2015, International Speech and Communication Association, 2015.
[127] J. Orozco-Arroyave, J. Arias-Londoño, New Spanish speech corpus database for the analysis of people suffering from parkinson's disease, Proceedings on the International Conference on Language Resources and Evaluation (LREC) (2014).
[128] T. Khan, J. Westin, M. Dougherty, Cepstral separation difference: a novel approach for speech impairment quantification in parkinson's disease, Biocybern. Biomed. Eng. 34 (2014) 25–34.
[129] M. Cernak, J.R. Orozco-Arroyave, F. Rudzicz, H. Christensen, J.C. Vásquez-Correa, E. Nöth, Characterisation of voice quality of parkinson's disease using differential phonological posterior features, Comput. Speech Lang. 46 (2017) 196–208.
[130] H.-H. Zhang, L. Yang, Y. Liu, P. Wang, J. Yin, Y. Li, M. Qiu, X. Zhu, F. Yan, Classification of parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples, Biomed. Eng. Online 15 (2016) 122.
[131] W. Froelich, K. Wrobel, P. Porwik, Diagnosis of parkinson's disease using speech samples and threshold-based classification, J. Med. Imaging Health Inform. 5 (2015) 1358–1363.
[132] J.C. Vásquez-Correa, J.R. Orozco-Arroyave, E. Nöth, Convolutional neural network to model articulation impairments in patients with parkinson's disease, INTERSPEECH (2017) 314–318.
[133] J.C. Vásquez-Correa, T. Arias-Vergara, C.D. Rios-Urrego, M. Schuster, J. Rusz, J. R. Orozco-Arroyave, E. Nöth, Convolutional neural networks and a transfer learning strategy to classify parkinson's disease from speech in three different languages, in: Iberoamerican Congress on Pattern Recognition, Springer, 2019, pp. 697–706.
[134] L. Moro-Velazquez, J. Villalba, N. Dehak, Using x-vectors to automatically detect parkinson's disease from speech, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 1155–1159.
[135] C. Botelho, F. Teixeira, T. Rolland, A. Abad, I. Trancoso, Pathological Speech Detection Using X-Vector Embeddings, 2020 arXiv preprint arXiv:2003.00864.
[136] Y. Hauptman, R. Aloni-Lavi, I. Lapidot, T. Gurevich, Y. Manor, S. Naor, N. Diamant, I. Opher, Identifying distinctive acoustic and spectral features in parkinson's disease, Interspeech (2019) 2498–2502.
[137] 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 J. Biomed. Health Inform. 17 (2013) 828–834.
[138] T. Bocklet, S. Steidl, E. Nöth, S. Skodda, Automatic evaluation of parkinson's speech-acoustic, prosodic and voice related cues, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2013.
[139] J.R. Orozco-Arroyave, J.C. Vásquez-Correa, J.F. Vargas-Bonilla, R. Arora, N. Dehak, P.S. Nidadavolu, H. Christensen, F. Rudzicz, M. Yancheva, H. Chinaei, et al., Neurospeech: an open-source software for parkinson's speech analysis, Digit. Signal Process. (2017).
[140] L. Naranjo, C.J. Pérez, Y. Campos-Roca, J. Martín, Addressing voice recording replications for parkinson's disease detection, Expert Syst. Appl. 46 (2016) 286–292.
[141] Z. Galaz, J. Mekyska, V. Zvoncak, J. Mucha, T. Kiska, Z. Smekal, I. Eliasova, M. Mrackova, M. Kostalova, I. Rektorova, et al., Changes in phonation and their relations with progress of parkinson's disease, Appl. Sci. 8 (2018) 2339.
[142] A. Tsanas, M.a. Little, C. Fox, L.O. Ramig, Objective automatic assessment of rehabilitative speech treatment in parkinson's disease, IEEE Trans. Neural Syst. Rehabil. Eng. 22 (2014) 181–190.
[143] S. Arora, N.P. Visanji, T.A. Mestre, A. Tsanas, A. AlDakheel, B.S. Connolly, C. Gasca-Salas, D.S. Kern, J. Jain, E.J. Slow, et al., Investigating voice as a biomarker for leucine-rich repeat kinase 2-associated parkinson's disease, J. Parkinson's Dis. 8 (2018) 503–510.

[144] R. Viswanathan, S.P. Arjunan, A. Bingham, B. Jelfs, P. Kempster, S. Raghav, D. K. Kumar, Complexity measures of voice recordings as a discriminative tool for parkinson's disease, Biosensors 10 (2020) 1.
[145] L. Ramig, C. Bonitati, J. Lemke, Y. Horii, et al., Voice treatment for patients with parkinson disease: Development of an approach and preliminary efficacy data, J. Med. Speech Lang. Pathol. 2 (1994) 191–209.
[146] T. Villa-Canas, J.D. Arias-Londono, J.R. Orozco-Arroyave, J. Vargas-Bonilla, E. Noth, Low-frequency components analysis in running speech for the automatic detection of parkinson's disease, Sixteenth Annual Conference of the International Speech Communication Association (2015).
[147] M. Brückl, A. Ghio, F. Viallet, Measurement of tremor in the voices of speakers with parkinson's disease, Procedia Comput. Sci. 128 (2018) 47–54.
[148] S. Sapir, L.O. Ramig, J. Spielman, C. Fox, Acoustic metrics of vowel articulation in parkinson's disease: Vowel space area (vsa) vs. vowel articulation index (vai), Seventh International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (2011).
[149] A.R. Fletcher, M.J. McAuliffe, K.L. Lansford, J.M. Liss, Assessing vowel centralization in dysarthria: a comparison of methods, J. Speech Lang. Hear. Res. 60 (2017) 341–354.
[150] V. Martel Sauvageau, J. Macoir, M. Langlois, M. Prud'Homme, L. Cantin, J.-P. Roy, Changes in vowel articulation with subthalamic nucleus deep brain stimulation in dysarthric speakers with parkinson's disease, Parkinson's Dis. 2014 (2014).
[151] V. Martel Sauvageau, J.-P. Roy, L. Cantin, M. Prud'Homme, M. Langlois, J. Macoir, Articulatory changes in vowel production following stn dbs and levodopa intake in parkinson's disease, Parkinson's Dis. 2015 (2015).
[152] J.R. Orozco-Arroyave, F. Hönig, J.D. Arias-Londoño, J. Vargas-Bonilla, S. Skodda, J. Rusz, E. Nöth, Voiced/unvoiced transitions in speech as a potential bio-marker to detect parkinson's disease, Sixteenth Annual Conference of the International Speech Communication Association (2015).
[153] A. Benba, A. Jilbab, A. Hammouch, Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with parkinson's disease and healthy people, Int. J. Speech Technol. 19 (2016) 449–456.
[154] Q.W. Oung, S.N. Basah, H. Muthusamy, V. Vijean, H. Lee, Evaluation of short-term cepstral based features for detection of parkinson's disease severity levels through speech signals, in: IOP Conference Series: Materials Science and Engineering, volume 318, IOP Publishing, 2018, 012039.
[155] B.T. Harel, M.S. Cannizzaro, H. Cohen, N. Reilly, P.J. Snyder, Acoustic characteristics of parkinsonian speech: a potential biomarker of early disease progression and treatment, J. Neurolinguist. 17 (2004) 439–453.
[156] T. Tykalova, J. Rusz, J. Klempir, R. Cmejla, E. Ruzicka, Distinct patterns of imprecise consonant articulation among parkinson's disease, progressive supranuclear palsy and multiple system atrophy, Brain Lang. 165 (2017) 1–9.
[157] Y. Kim, Y. Choi, A cross-language study of acoustic predictors of speech intelligibility in individuals with parkinson's disease, J. Speech Lang. Hear. Res. 60 (2017) 2506.
[158] L.v.d. Maaten, G. Hinton, Visualizing data using t-sne, J. Mach. Learn. Res. 9 (2008) 2579–2605.
[159] K.A. Bridges, D.V.L. Sidtis, J.J. Sidtis, The role of subcortical structures in recited speech: Studies in parkinson's disease, J. Neurolinguist. 26 (2013) 591–601.
[160] J.A. Whitfield, A.M. Goberman, Speech motor sequence learning: effect of parkinson disease and normal aging on dual-task performance, J. Speech Lang. Hear. Res. 60 (2017) 1752–1765.
[161] V. Biousse, B. Skibell, R. Watts, D. Loupe, C. Drews-Botsch, N. Newman, Ophthalmologic features of parkinson's disease, Neurology 62 (2004) 177–180.
[162] S. Abrevaya, L. Sedeño, S. Fitipaldi, D. Pineda, F. Lopera, O. Buritica, A. Villegas, C. Bustamante, D. Gomez, N. Trujillo, et al., The road less traveled: alternative pathways for action-verb processing in parkinson's disease, J. Alzheimer's Dis. 55 (2017) 1429–1435.
[163] P. Pérez-Toro, J. Vásquez-Correa, M. Strauss, J. Orozco-Arroyave, E. Nöth, Natural language analysis to detect parkinson's disease, in: International Conference on Text, Speech, and Dialogue, Springer, 2019, pp. 82–90.
[164] V. Boulenger, L. Mechtouff, S. Thobois, E. Broussolle, M. Jeannerod, T.A. Nazir, Word processing in parkinson's disease is impaired for action verbs but not for concrete nouns, Neuropsychologia 46 (2008) 743–756.
[165] L. Fernandino, L.L. Conant, J.R. Binder, K. Blindauer, B. Hiner, K. Spangler, R. H. Desai, Parkinson's disease disrupts both automatic and controlled processing of action verbs, Brain Lang. 127 (2013) 65–74.
[166] Y. Bocanegra, A.M. García, D. Pineda, O. Buriticá, A. Villegas, F. Lopera, D. Gómez, C. Gómez-Arias, J.F. Cardona, N. Trujillo, et al., Syntax, action verbs, action semantics, and object semantics in parkinson's disease: dissociability, progression, and executive influences, Cortex 69 (2015) 237–254.
[167] Y. Bocanegra, A.M. García, F. Lopera, D. Pineda, A. Baena, P. Ospina, D. Alzate, O. Buriticá, L. Moreno, A. Ibá nez, et al., Unspeakable motion: selective action-verb impairments in parkinson's disease patients without mild cognitive impairment, Brain Lang. 168 (2017) 37–46.
[168] D. Kemmerer, L. Miller, M.K. MacPherson, J. Huber, D. Tranel, An investigation of semantic similarity judgments about action and non-action verbs in parkinson's disease: implications for the embodied cognition framework, Front. Hum. Neurosci. 7 (2013) 146.
[169] S. Pinto, A. Chan, I. Guimarães, R. Rothe-Neves, J. Sadat, A cross-linguistic perspective to the study of dysarthria in parkinson's disease, J. Phonet. 64 (2017) 156–167.
[170] H. Hazan, D. Hilu, L. Manevitz, L.O. Ramig, S. Sapir, Early diagnosis of parkinson's disease via machine learning on speech data, in: 2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel, IEEE, 2012, pp. 1–4.
[171] N. Garcia, J.C. Vasquez-Correa, J.R. Orozco-Arroyave, N. Dehak, E. Nöth, Language independent assessment of motor impairments of patients with parkinson's disease using i-vectors, in: International Conference on Text, Speech, and Dialogue, Springer, 2017, pp. 147–155.
[172] R.D. Kent, J.C. Rosenbek, Prosodic disturbance and neurologic lesion, Brain Lang. 15 (1982) 259–291.
[173] C.L. Ludlow, N.P. Connor, C.J. Bassich, Speech timing in parkinson's and huntington's disease, Brain Lang. 32 (1987) 195–214.
[174] J. Illes, Neurolinguistic features of spontaneous language production dissociate three forms of neurodegenerative disease: alzheimer's, huntington's, and parkinson's, Brain Lang. 37 (1989) 628–642.
[175] I. Hertrich, H. Ackermann, Acoustic analysis of speech prosody in Huntington's and Parkinson's disease: a preliminary report, Clin. Linguist. Phonet. 7 (1993) 285–297.
[176] E.D. Louis, L.A. Klatka, Y. Liu, S. Fahn, Comparison of extrapyramidal features in 31 pathologically confirmed cases of diffuse lewy body disease and 34 pathologically confirmed cases of parkinson's disease, Neurology 48 (1997) 376–380.
[177] G. Weismer, J. Wildermuth, Formant trajectory characteristics in persons with parkinson, cerebellar, and upper motor neuron disease, J. Acoust. Soc. Am. 103 (1998) 2892.
[178] G. Le Dorze, J. Ryalls, C. Brassard, N. Boulanger, D. Ratte, A comparison of the prosodic characteristics of the speech of people with parkinson's disease and friedreich's ataxia with neurologically normal speakers, Folia Phoniatr. Logop. 50 (1998) 1–9.
[179] C. Kuo, K. Tjaden, J.E. Sussman, Acoustic and perceptual correlates of faster-than-habitual speech produced by speakers with parkinson's disease and multiple sclerosis, J. Commun. Disord. 52 (2014) 156–169.
[180] Y.E. Huh, J. Park, M.K. Suh, S.E. Lee, J. Kim, Y. Jeong, H.-T. Kim, J.W. Cho, Differences in early speech patterns between parkinson variant of multiple system atrophy and parkinson's disease, Brain Lang. 147 (2015) 14–20.
[181] J. Rusz, J. Hlavnicka, R. Cmejla, E. Ruzicka, Automatic evaluation of speech rhythm instability and acceleration in dysarthrias associated with basal ganglia dysfunction, Front. Bioeng. Biotechnol. 3 (2015).
[182] J.C. Jimenez-Monsalve, J.C. Vasquez-Correa, J.R. Orozco-Arroyave, P. Gomez-Vilda, Phonation and articulation analyses in laryngeal pathologies, cleft lip and palate, and parkinson's disease, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 10338 LNCS (2017) 424–434.
[183] B. Das, K. Daoudi, J. Klempir, J. Rusz, Towards disease-specific speech markers for differential diagnosis in parkinsonism, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 5846–5850.
[184] G. Weismer, Y. Kim, B. Maassen, P. van Lieshout, Classification and taxonomy of motor speech disorders: what are the issues, Speech motor control: New developments in basic and applied research (2010) 229–241.
[185] M.A. Little, R. Badawy, Causal Bootstrapping, 2019 arXiv preprint arXiv: 1910.09648.
[186] J.G. Richens, C.M. Lee, S. Johri, Improving the accuracy of medical diagnosis with causal machine learning, Nat. Commun. 11 (2020) 1–9.
[187] J.R. Orozco-Arroyave, N. García, J.F. Vargas-Bonilla, E. Nöth, Automatic detection of parkinson's disease from compressed speech recordings, in: International Conference on Text, Speech, and Dialogue, Springer, 2015, pp. 88–95.
[188] D. Ireland, C. Atay, J. Liddle, D. Bradford, H. Lee, O. Rushin, T. Mullins, D. Angus, J. Wiles, S. McBride, et al., Hello harlie: enabling speech monitoring through chat-bot conversations, HIC (2016) 55–60.
[189] F. Schaeffler, S. Jannetts, J.M. Beck, Reliability of clinical voice parameters captured with smartphones-measurements of added noise and spectral tilt, in: Proceedings of the 20th Annual Conference of the International Speech Communication Association INTERSPEECH, Graz, Austria, 15-19 September 2019, ISCA, 2019.
[190] S. Jannetts, F. Schaeffler, J. Beck, S. Cowen, Assessing voice health using smartphones: bias and random error of acoustic voice parameters captured by different smartphone types, Int. J. Lang. Commun. Disord. 54 (2019) 292–305.
[191] S.E. Linville, E.W. Korabic, M. Rosera, Intraproduction variability in jitter measures from elderly speakers, J. Voice 4 (1990) 45–51.
[192] R.L. Ringel, W.J. Chodzko-Zajko, Vocal indices of biological age, J. Voice 1 (1987) 31–37.
[193] S.E. Linville, The sound of senescence, J. Voice 10 (1996) 190–200.
[194] S. Arora, L. Baghai-Ravary, A. Tsanas, Developing a large scale population screening tool for the assessment of parkinson's disease using telephone-quality voice, J. Acoust. Soc. Am. 145 (2019) 2871–2884.
[195] C. Vitale, V. Marcelli, T. Abate, A. Pianese, R. Allocca, M. Moccia, E. Spina, P. Barone, G. Santangelo, M. Cavaliere, Speech discrimination is impaired in parkinsonian patients: expanding the audiologic findings of parkinson's disease, Parkinsonism Relat. Disord. 22 (2016) S138–S143.
[196] S. Gräber, I. Hertrich, I. Daum, S. Spieker, H. Ackermann, Speech perception deficits in parkinson's disease: underestimation of time intervals compromises identification of durational phonetic contrasts, Brain Lang. 82 (2002) 65–74.
[197] L. Jeancolas, D. Petrovska-Delacrétaz, G. Mangone, B.-E. Benkelfat, J.-C. Corvol, M. Vidailhet, S. Lehéricy, H. Benali, X-vectors: New Quantitative Biomarkers for

Early Parkinson's Disease Detection from Speech, 2020 arXiv preprint arXiv: 2007.03599.
[198] J. Hanratty, C. Deegan, M. Walsh, B. Kirkpatrick, Analysis of glottal source parameters in parkinsonian speech, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2016, pp. 3666–3669.
[199] D. Braga, A.M. Madureira, L. Coelho, R. Ajith, Automatic detection of parkinson's disease based on acoustic analysis of speech, Eng. Appl. Artif. Intell. 77 (2019) 148–158.
[200] R. Lenain, J. Weston, A. Shivkumar, E. Fristed, Surfboard: audio feature extraction for modern machine learning, Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH (2020).
[201] M.M. Mouradian, J. Junecos, G. Fabbrini, J. Schlegel, J. Bartko, T. Chase, Motor fluctuations in parkinson's disease: central pathophysiological mechanisms, part ii, Ann. Neurol. 24 (1988) 372–378.
[202] J. Jankovic, Motor fluctuations and dyskinesias in parkinson's disease: clinical manifestations, Mov. Disord. 20 (2005) S11–S16.
[203] A. Leentjens, K. Dujardin, L. Marsh, P. Martinez-Martin, I. Richard, S. E. Starkstein, Anxiety and motor fluctuations in parkinson's disease: a cross-sectional observational study, Parkinsonism Relat. Disord. 18 (2012) 1084–1088.
[204] R. Martínez-Fernández, E. Schmitt, P. Martinez-Martin, P. Krack, The hidden sister of motor fluctuations in parkinson's disease: a review on nonmotor fluctuations, Mov. Disord. 31 (2016) 1080–1094.
[205] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, F. Girardi, Assessment of speech intelligibility in parkinson's disease using a speech-to-text system, IEEE Access 5 (2017) 22199–22208.
[206] Y. Jadoul, B. Thompson, B. de Boer, Introducing parselmouth: a python interface to praat, J. Phonet. 71 (2018) 1–15, https://doi.org/10.1016/j.wocn.2018.07.001.
[207] A. Tsanas, Accurate Telemonitoring of Parkinson's Disease Symptom Severity Using Nonlinear Speech Signal Processing and Statistical Machine Learning, Ph.D. thesis, Oxford University, UK, 2012.
[208] P. Zwirner, G.J. Barnes, Vocal tract steadiness: a measure of phonatory and upper airway motor control during phonation in dysarthria, J. Speech Lang. Hear. Res. 35 (1992) 761–768.