INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Parkinson's Disease Detection from Speech using Single Frequency Filtering Cepstral Coefficients

Sudarsana Reddy Kadiri [1], *Rashmi Kethireddy* [2], *Paavo Alku* [1]

[1]Department of Signal Processing and Acoustics, Aalto University, Finland
[2]Speech Processing Laboratory, International Institute of Information Technology, Hyderabad, India
sudarsana.kadiri@aalto.fi; rashmi.kethireddy@research.iiit.ac.in; paavo.alku@aalto.fi

Abstract

Parkinson's disease (PD) is a progressive deterioration of the human central nervous system. Detection of PD (discriminating patients with PD from healthy subjects) from speech is a useful approach due to its non-invasive nature. This study proposes to use novel cepstral coefficients derived from the single frequency filtering (SFF) method, called as single frequency filtering cepstral coefficients (SFFCCs) for the detection of PD. SFF has been shown to provide higher spectro-temporal resolution compared to the short-time Fourier transform. The current study uses the PC-GITA database, which consists of speech from speakers with PD and healthy controls (50 males, 50 females). Our proposed detection system is based on the i-vectors derived from SFFCCs using SVM as a classifier. In the detection of PD, better performance was achieved when the i-vectors were computed from the proposed SFFCCs compared to the popular conventional MFCCs. Furthermore, we investigated the effect of temporal variations by deriving the shifted delta cepstral (SDC) coefficients using SFFCCs. These experiments revealed that the i-vectors derived from the proposed SFFCCs+SDC features gave an absolute improvement of 9% compared to the i-vectors derived from the baseline MFCCs+SDC features, indicating the importance of temporal variations in the detection of PD.

Index Terms: Speech Pathology, Single Frequency Filtering, Parkinson's Disease Detection.

1. Introduction

Disorders such as neurodegenerative diseases, psychiatric and developmental diseases have an significant impact on humans at personal, professional and social levels [1]. These diseases influence adversely the quality of life and life span of individuals. The early detection of the disease might reduce the effects physically and mentally, and allow a timely control of the disease evaluation. The current study focuses on the detection of one of the neurodegenerative diseases from speech, namely Parkinson's disease (PD). PD is the second most common neurodegenerative disease after Alzheimer's disease [2, 3].

The detection of PD from speech has been investigated in many studies [4–7]. These previous studies can be divided into four categories depending on the approach which has been used to analyze the speech signal [4, 6]. These four approaches are: (1) phonatory, (2) articulatory, (3) prosodic, and (4) linguistic. The studies based on the phonatory approach focus on the changes in the glottal source and in the larynx. There are only a few studies in this group due to the difficulties in estimating the glottal source from disordered speech signals. Studies based on the articulatory and prosodic approaches are much more prevalent as there exist several speech processing tools for deriving features such as Mel-frequency cepstral coefficients (MFCCs), pitch, duration, etc. [8,9]. In addition, the prevalence of the articulatory approach is explained by results shown in several studies indicating that articulation is greatly affected in PD [8–10]. Finally, the studies belonging to the linguistic approach examine the use of vocabulary, phrase construction and repetition of words by Parkinsonian speakers [11, 12]. For investigating the linguistic approach, previous studies have used automatic speech recognition (ASR) in representing linguistic units from speech. These studies have used classical features such as bag of words and term frequency-inverse document frequency [13].

For the analysis and detection of PD, many feature extraction methods of speech have been used to express Parkinsonian speech signals in parametric forms. For capturing phonatory aspects, features quantifying variations in speech periodicity have been investigated [14–16]. This parameterization approach is justified because the extent of variations in the vocal folds vibration in Parkinsonian speakers is more likely to deviate from healthy speakers [17]. Features such as jitter (which is defined as perturbation in fundamental frequency) and shimmer (which is defined as perturbation in amplitude) are the most commonly used measures for capturing variations in vocal fold vibrations [5, 18]. In [5, 18, 19], various nonlinear voice production -based features (such as the recurrence period density entropy) were investigated. Recently in [14], glottal source features (such as the quasi-open quotient, the normalized amplitude quotient, the harmonic richness factor) were investigated in the analysis of newly diagnosed Parkinsonian patients. To measure the amount of noise in voice due to incomplete glottal closure (with symptoms like breathiness and harshness), the harmonics-to-noise ratio and the noise-to-harmonics ratio have been used [5, 18]. For capturing articulatory variations, different feature extraction methods have been widely investigated in the areas of speech, speaker, and language recognition [20–22]. Among the popular features used in these areas, MFCCs, along with their first and second derivatives, have been found to be effective in the detection of PD [21, 23]. Examples of other popular feature extraction methods, which have been first used particularly in ASR but later also in the detection of PD, are the linear prediction coefficients (LPCs), linear prediction cepstral coefficients (LPCCs) and perceptual linear prediction cepstral coefficients (PLPCCs) [24–26]. To capture the prosodic aspects, features such as duration of voiced sounds, intonation, loudness, speaking (syllable) rate are typically investigated in the detection of PD. More details about the various types of features used in the literature can be found in [21, 27–29].

In [21,24,30], it was observed that cepstral coefficients such as MFCCs outperformed conventional features such as phonatory and prosodic features. Motivated by these previous studies, we propose in the current article to use cepstral coefficients derived using the recently proposed signal processing method, sin-

http://dx.doi.org/10.21437/Interspeech.2020-3197

gle frequency filtering (SFF). The SFF method has been shown in [31–34] to provide higher spectral and temporal resolution for deriving speech features compared to the short time Fourier transform, which is used in the computation of MFCCs.

The major contributions of this study are:

- Investigating single frequency filtering cepstral coefficients (SFFCCs) for the detection of PD from speech signals
- Exploring shifted delta cepstra (SDC) with SFFCCs for capturing temporal variations in speech information across frames

The organization of this paper is as follows. The single frequency filtering (SFF) method is described first in Section 2 to explain the basis for extracting SFFCCs. The PD detection system pipeline developed in this study by using the SFFCCs features is described in Section 3. The experimental details are discussed in Section 4, and results of the experiments are presented in Section 5. Finally, Section 6 presents a summary.

2. Single Frequency Filtering (SFF) and Extraction of SFFCCs

In this section, the SFF method is first described. After this, the extraction of cepstral coefficients, SFFCCs, from SFF is described.

2.1. The Single Frequency Filtering (SFF) method

SFF [31, 32] is a time-frequency analysis method that is used to compute an amplitude envelope as a function of time for each frequency. The amplitude envelope is derived by first frequency-shifting the pre-emphasized speech signal ($x[n]$) by multiplying with an exponential function ($\hat{x}_k[n] = x[n]e^{-j2\pi \hat{f}_k n/f_s}$, where f_s is the sampling frequency, $\hat{f}_k = \frac{f_s}{2} - f_k$, and f_k is the k^{th} desired frequency). The frequency-shifted signal is filtered using a single-pole filter. The transfer function of the single-pole filter is $H(z) = \frac{1}{1+rz^{-1}}$. The pole of the filter is located on the negative real axis at $z = -r$. The output of the filter is given by

$$y_k[n] = -ry_k[n-1] + \hat{x}_k[n]. \tag{1}$$

At frequency f_k, the amplitude envelope ($v_k[n]$) of $y_k[n]$ is computed as

$$v_k[n] = \sqrt{(y_{k_r}[n])^2 + (y_{k_i}[n])^2}, \tag{2}$$

where $y_{k_r}[n]$ is the real part of $y_k[n]$ and $y_{k_i}[n]$ is the imaginary parts of $y_k[n]$. The amplitude envelopes can be computed for several frequencies at intervals of Δf by defining f_k as follows:

$$f_k = k\Delta f, \qquad k = 1, 2, \ldots, K, \tag{3}$$

where $K = \frac{(f_s/2)}{\Delta f}$. In this study, 512 linearly spaced frequency components are considered (which results in $\Delta f = 31.25$). The SFF spectrum can be obtained for each instant of time from $v_k[n]$. The steps involved in deriving the SFF spectrum are shown in Fig. 1.

2.2. Extraction of SFFCCs

SFFCCs are extracted by computing the cepstrum ($C_k[n]$) of the SFF spectrum as follows:

$$C_k[n] = \text{IFFT}(\log_{10}(v_k[n])). \tag{4}$$

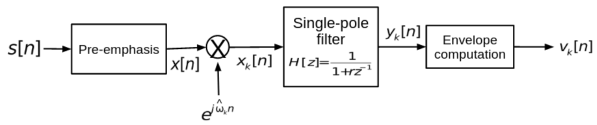

Figure 1: *Block diagram describing the steps involved in deriving the single frequency filtering (SFF) spectrum [34].*

From $C_k[n]$, the first 13 cepstral coefficients are considered, which are referred to as SFFCCs. From static coefficients, delta (Δ) and double-delta ($\Delta\Delta$) coefficients are also derived, which results in 39-dimensional SFFCCs. The schematic block diagram describing the steps involved in the extraction of SFFCCs is shown in Fig. 2.

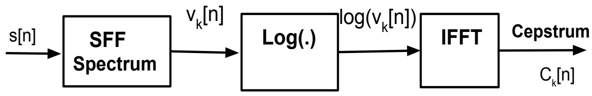

Figure 2: *Block diagram of the extraction of single frequency filtering cepstral coefficients (SFFCCs) [33, 35]*

3. Detection System

This section describes the details of the proposed system to detect PD from speech signals. The block diagram describing the steps involved in the proposed system is shown in Fig. 3. The system consists of three main parts: front-end feature extraction, back-end processing, and classification. The feature extraction stage includes the extraction of SFFCCs from the SFF spectrum, followed by the extraction of SDCs from SFFCCs. Back-end processing involves the extraction of fixed-length i-vectors from the variable-length features. The last part involves classifying the speech signals either as healthy or Pakinsonian using the support vector machine (SVM) classifier.

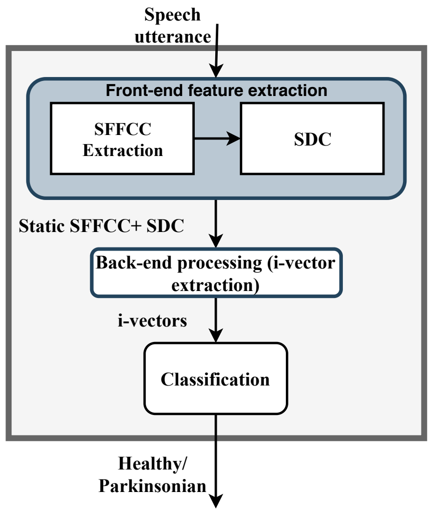

Figure 3: *Block diagram of the proposed PD detection system.*

3.1. Parameters used for SFFCCs feature extraction

For the SFF spectrum estimation, $r = 0.99$ and $\Delta f = 31.25$ Hz (resulting in 512 amplitude envelopes) are used. SFFCCs are extracted with an interval of 10 ms from $v_k[n]$ rather than considering at every time instant, and first 13 cepstral coefficients are extracted.

3.2. Shifted delta cepstra (SDC)

Studies in [36] have shown that the shifted delta cepstra (SDC) features capture temporal variations in speech across several frames. In this study, the SDC features are computed from the SFFCCs for each frame. The configuration used in the computation of the SDC features, denoted as N-d-p-K, involves four parameters, which are explained as follows. For time instant t, delta computations are conducted between the cepstral coefficients of the $(t+ip-d)^{th}$ frame and the $(t+ip+d)^{th}$ frame. By varying i from 1 to K and by stacking the obtained values, the delta coefficients are obtained. The SDC features, denoted by $\boldsymbol{\Delta}c(t,i)$, are computed for the cepstral coefficients at time t and shift i as follows

$$\boldsymbol{\Delta}\mathbf{c}(\mathbf{t},\mathbf{i}) = \mathbf{c}(\mathbf{t}+\mathbf{ip}+\mathbf{d}) - \mathbf{c}(\mathbf{t}+\mathbf{ip}-\mathbf{d}), \qquad (5)$$

where N denotes the dimension of the static cepstral coefficients; d denotes the delay/advance from the current frame; p is the shift between consecutive delta computations; and K such delta computations are concatenated to form N×K-dimensional SDC coefficients. In many previous studies, SDCs are derived from 13-dimensional MFCCs, using the 1-3-7 (d-p-K) configuration. In the present study, the configuration for SDC was initially set to 13-1-3-7 which resulted in 91 SDCs. Combining both the static 13-dimensional SFFCCs and the 91-dimensional SDCs results in a 104-dimensional feature vector for each frame. In addition, we investigate in this study the effect of using different values for d and K which changes the amount of temporal context (see Tables 3 and 4).

3.3. Extraction of i-vectors

The process of extracting i-vectors [37] involves factor analysis for representing the feature vectors in terms of uncorrelated latent components. Given variable-length features derived from utterances, GMM models the dependencies along the utterances by representing them using Gaussian components. Factor analysis is applied over the Gaussian supervectors (stacked means of GMMs). Each utterance adapted by Gaussian supervectors $\mathbf{M}$ can be represented by a mean component ($\mathbf{m}$) and an offset ($\mathbf{Tw}$).

$$\mathbf{M} = \mathbf{m} + \mathbf{Tw}, \qquad (6)$$

where mean component $\mathbf{m}$ is obtained from the model trained on all class data (GMM-UBM supervectors), $\mathbf{T}$ is the total variability matrix, and $\mathbf{w}$ represents the i-vectors that are used for classification. The initial means and variances of the GMM-UBM model were learnt using k-means clustering. Various configurations for the GMM and i-vector dimensions were investigated and it was found that the system performed best using 8 Gaussian components and 50-dimensional i-vectors.

3.4. Classification

Detection experiments were carried out using the SVM classifier. SVM was chosen because it is known to be an effective classifier, when the amount of training data is limited. In the experiments, 80% of the data is used for training and 20% of data is used for testing. The SVM was trained with the radial basis function (RBF) kernel in the one-vs-rest fashion.

4. Experimental setup

This section describes the PD database used, the evaluation metrics selected and the baseline features considered for comparison.

4.1. Database

In this study, the PC-GITA Parkinsonian speech database is used [38]. The database consists of speech recordings of 50 patients with PD (25 male and 25 female) and 50 healthy controls (25 male and 25 female). Recordings were carried out in a sound proof booth and all the speakers are Colombian Spanish native speakers. The database is balanced by gender and age. The age ranges between 33 and 77 years (mean 62.26 years) for the male patients, and between 44 and 75 years (mean 60.16 years) for the female patients. For the healthy speakers, the age ranges between 31 and 86 years (mean 61.26 years) for the males, and between 43 and 76 years (mean 60.76 years) for the females. The data consists of vowels, isolated words, diadochokinetic words, sentences, and reading text. In this study, three repetitions of the five Spanish sustained vowels are considered as in [3,18]. Further details of the database can be found in [38, 39].

4.2. Evaluation metrics

The primary evaluation metric is the detection accuracy. As an additional metric, the F1-score is also reported. The F1-score balances between false positives and false negatives which makes it unbiased to the majority class.

4.3. Baseline features for comparison

The most popular speech features, namely MFCCs [19, 21, 24], are used as the baseline features. The baseline system configuration is similar to the proposed system. In the baseline system, 13-dimensional MFCCs are extracted using a hamming window size of 20 ms and a shift of 10 ms. The mean normalized 13 static MFCC features are used to obtain 91 SDC features using the 13-1-3-7 configuration which is similar to the configuration used in our proposed system.

5. Results and discussion

Detection experiments are carried out with i-vectors extracted from six cepstral representations as follows: (1) 13-dimensional MFCCs, (2) 39-dimensional MFCCs+Δ + $\Delta\Delta$, (3) MFCCs+SDC (i.e., the combination of 13-dimensional MFCCs and the corresponding 91-dimensional SDC), (4) 13-dimensional SFFCCs, (5) 39-dimensional SFFCCs+Δ + $\Delta\Delta$, and (6) SFFCCs+SDC (i.e., the combination of 13-dimensional SFFCCs and the corresponding 91-dimensional SDC). These systems are referred to, respectively, as: MFCCs, MFCCs+Δ + $\Delta\Delta$, MFCCs+SDC, SFFCCs, SFFCCs+Δ + $\Delta\Delta$ and SFFCCs+SDC in this paper.

The detection results of the six systems described above are reported in Table 1. From the table, it can be clearly seen that all the systems using the proposed SFFCCs features (rows 4-6) perform clearly better than the baseline systems (rows 1-3). Among the baseline systems, the MFCCs+SDC system performed better than the other systems. Similarly, among the proposed systems, the SFFCCs+SDC system performed better than the other ones.

By comparing the results obtained with the static cepstral coefficients, it was found that the proposed SFFCCs system outperformed the baseline MFCCs system with an absolute improvement of 12%. Adding the first and second derivatives to the static cepstral coefficients does not improve the performance of the baseline MFCC-based system. However, adding the first and second derivatives improved the performance of the

Table 1: *Detection performance (accuracy in % and F1-score in %) for the baseline (MFCCs-based) and proposed (SFFCCs-based) systems.*

Features used for deriving i-vectors	**Acc.**	**F1**
MFCCs	56.66	56.54
MFCCs+$\Delta + \Delta\Delta$	55.00	54.44
MFCCs+SDC	**64.66**	**64.26**
SFFCCs	68.66	68.59
SFFCCs+$\Delta + \Delta\Delta$	72.00	71.93
SFFCCs+SDC	**73.33**	**73.32**

proposed SFFCC-based system. Furthermore, it can be clearly seen that by introducing the SDC features with the static cepstral coefficients improved the detection performance for both the MFCC-based system (i.e., MFCCs+SDC) and the proposed SFFCC-based system (i.e., SFFCCs+SDC). This shows, importantly, that using speech information extracted over a longer context indeed helps improving the detection of PD. Between the SDC-based systems, the proposed SFFCCs+SDC system outperformed the baseline MFCCs+SDC system with an absolute improvement of 9%. Overall, it is observed that the proposed SFFCCs+SDC system performed better than any other system. The better results obtained with the SFFCC-based systems indicate the capability of SFF to capture the spectro-temporal variations of speech.

The class-wise detection performance in terms of accuracy is given in Table 2 for healthy and Parkinsonian speech. From the results, it can be observed that the baseline systems (MFCCs, MFCCs+$\Delta + \Delta\Delta$ and MFCCs+SDC) are more accurate in detecting healthy speech than Parkinsonian speech. However, the proposed systems (SFFCCs, SFFCCs+$\Delta + \Delta\Delta$ and SFFCCs+SDC) detect Parkinsonan speech better than the baseline systems. Among the baseline systems, the MFCCs+SDC system is better than the other baselines. Similarly, among the proposed systems, the SFFCCs+SDC is found to be more accurate than the other systems.

Table 2: *Class-wise accuracies (in %) for the baseline (MFCCs-based) and proposed (SFFCCs-based) systems. HS: Healthy speech, PS: Parkinsonian speech.*

Feature representation	**HS**	**PS**
MFCCs	62.00	51.33
MFCCs+$\Delta + \Delta\Delta$	66.00	44.00
MFCCs+SDC	**75.33**	**56.67**
SFFCCs	73.33	64.00
SFFCCs+$\Delta + \Delta\Delta$	76.66	67.33
SFFCCs+SDC	**75.33**	**71.33**

From the results in Tables 1 and 2, it was observed that by computing SDCs from the static cepstral coefficients (MFCCs or SFFCCs) gave a clear improvement in detection performance, and also lesser bias towards the healthy speech class. Due to this observation, we investigated further the configurations used in the SDC computation. More specifically, two important parameters d (the number of delays/advances from the current frame) and K (the number of delta computations for concatenation) were varied to study the effect of these parameters on detection performance. The remaining two parameters N (the dimension of the static cepstral coefficients) and p (the shift between consecutive delta computations) were fixed to 13 and 3, respectively.

Tables 3 and 4 show the results obtained for varying the selected two parameters in configuring the SDC computation. The d value was varied from 1 to 3 (see Tables 3), and it was found that the baseline MFCCs+SDC system performed best with $d = 1$ and that the proposed SFFCCs+SDC system performed best with $d = 2$. By fixing the d value, K value was varied from 5 to 8, and the results are given in Tables 4. From the results shown in table, it can be observed that both the baseline MFCCs+SDC system and the proposed SFFCCs+SDC system performed best with $K = 6$. In summary, the best performance for the MFCCs+SDC baseline system was achieved using the configuration 13-1-3-6 (65%) and the best configuration for the proposed SFFCCs+SDC system was 13-2-3-6 (76%). When comparing the best configurations for the baseline and proposed systems, it is found that the proposed system gave an absolute improvement of 11%.

Table 3: *Detection performance (accuracy in % and F1-score in %) obtained by varying the delay/advance parameter d (from 1 to 3) in the computation of SDC for the MFCCs+SDC and SFFCCs+SDC systems.*

Feature representation	**Acc.**	**F1**
MFCCs+SDC (13-**1**-3-7)	**64.66**	**64.26**
MFCCs+SDC (13-**2**-3-7)	60.00	59.78
MFCCs+SDC (13-**3**-3-7)	61.00	60.52
SFFCCs+SDC (13-**1**-3-7)	73.33	73.32
SFFCCs+SDC (13-**2**-3-7)	**75.33**	**75.29**
SFFCCs+SDC (13-**3**-3-7)	70.33	70.25

Table 4: *Detection performance (accuracy in % and F1-score in %) obtained by varying the delta computation parameter K (from 5 to 8) in the computation of SDC for the MFCCs+SDC and SFFCCs+SDC systems.*

Feature representation	**Acc.**	**F1**
MFCCs+SDC (13-1-3-**5**)	60.66	60.36
MFCCs+SDC (13-1-3-**6**)	**65.00**	**64.91**
MFCCs+SDC (13-1-3-**7**)	64.66	64.26
MFCCs+SDC (13-1-3-**8**)	62.00	61.71
SFFCCs+SDC (13-2-3-**5**)	69.66	69.59
SFFCCs+SDC (13-2-3-**6**)	**76.00**	**75.93**
SFFCCs+SDC (13-2-3-**7**)	75.33	75.29
SFFCCs+SDC (13-2-3-**8**)	72.66	72.56

6. Summary

In this study, we proposed to use single frequency filtering cepstral coefficients (SFFCCs) for the detection of Parkinson's disease from speech. The detection experiments were carried out with the sustained vowels from well-known PC-GITA database. Experiments with the SVM classifier revealed that the i-vectors derived from the proposed SFFCCs features outperformed the i-vectors derived from the conventional MFCCs features. Furthermore, by computing SDCs from the SFFCCs (SFFCCs+SDC) resulted in an absolute improvement of 9% compared to the baseline MFCCs+SDC features, indicating the importance of temporal variations captured by SDC.

7. References

[1] L. Marsh, "Depression and parkinson's disease: current knowledge," *Current neurology and neuroscience reports*, vol. 13, no. 12, p. 409, 2013.

[2] L. V. Kalia and A. E. Lang, "Parkinson's disease." *Current neurology and neuroscience reports.*, Aug 2015.

[3] S. Arora, L. Baghai-Ravary, and A. Tsanas, "Developing a large

scale population screening tool for the assessment of parkinson's disease using telephone-quality voice," *The Journal of the Acoustical Society of America*, vol. 145, no. 5, pp. 2871–2884, 2019.

[4] M. Cernak *et al.*, "Characterisation of voice quality of parkinsonâs disease using differential phonological posterior features," *Computer Speech Language*, vol. 46, pp. 196 – 208, 2017.

[5] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average parkinson's disease symptom severity," *Journal of the royal society interface*, vol. 8, no. 59, pp. 842–855, 2011.

[6] J. R. Orozco-Arroyave *et al.*, "Characterization methods for the detection of multiple voice disorders: neurological, functional, and laryngeal diseases," *IEEE jour. of biomed. and health infor.*, vol. 19, no. 6, pp. 1820–1828, 2015.

[7] L. M. A., "Suitability of dysphonia measurements for telemonitoring of parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, p. 1015, 2009.

[8] M. Novotnỳ, J. Rusz *et al.*, "Automatic evaluation of articulatory disorders in Parkinson's disease," *IEEE/ACM Trans. on Audio, Speech and Lang. Process.*, vol. 22, no. 9, pp. 1366–1378, 2014.

[9] T. Tykalova, J. Rusz, J. Klempir, R. Cmejla, and E. Ruzicka, "Distinct patterns of imprecise consonant articulation among parkinson's disease, progressive supranuclear palsy and multiple system atrophy," *Brain and language*, vol. 165, pp. 1–9, 2017.

[10] H. B., "Variability in fundamental frequency during speech in prodromal and incipient parkinson's disease: A longitudinal case study," *Brain Cogn.*, vol. 56, p. 24, 2004.

[11] A. M. García *et al.*, "How language flows when movements don't: an automated analysis of spontaneous discourse in parkinson's disease," *Brain and language*, vol. 162, pp. 19–28, 2016.

[12] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, "Linguistic features identify alzheimer's disease in narrative speech," *Journal of Alzheimer's Disease*, vol. 49, no. 2, pp. 407–422, 2016.

[13] P. A. Pérez-Toro, J. C. Vásquez-Correa, M. Strauss, J. R. Orozco-Arroyave, and E. Nöth, "Natural language analysis to detect parkinson's disease," in *Text, Speech, and Dialogue*, 2019, pp. 82–90.

[14] M. Novotnỳ, P. Dušek, I. Daly, E. Ržička, and 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. Sign. Process. and Cont.*, vol. 57, p. 101818, 2020.

[15] A. Tsanas *et al.*, "Objective automatic assessment of rehabilitative speech treatment in parkinson's disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, p. 181, 2014.

[16] B. E. Sakar *et al.*, "Collection and analysis of a parkinson speech dataset with multiple types of sound recordings," *IEEE Journal of Biomed. and Health Infor.*, vol. 17, no. 4, pp. 828–834, 2013.

[17] C. M. Travieso, J. B. Alonso, J. Orozco-Arroyave, J. Vargas-Bonilla, E. Nöth, and A. G. Ravelo-García, "Detection of different voice diseases based on the nonlinear characterization of speech signals," *Exp. Syst. with App.*, vol. 82, pp. 184 – 195, 2017.

[18] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884–893, April 2010.

[19] P. Henríquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, J. I. Godino-Llorente, and F. Díaz-de-María, "Characterization of healthy and pathological voice through measures based on nonlinear dynamics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1186–1195, 2009.

[20] L. Moro-Velázquez *et al.*, "Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect parkinson's disease," *Appl. Soft Comput.*, vol. 62, pp. 649–666, 2018.

[21] J. Orozco-Arroyave *et al.*, "Automatic detection of parkinson's disease in running speech spoken in three different languages," *The Journal of the Acoustical Society of America*, vol. 139, no. 1, pp. 481–500, 2016.

[22] L. Moro-Velazquez *et al.*, "A forced gaussians based methodology for the differential evaluation of Parkinson's Disease by means of speech processing," *Biomed. Signal Process. and Contr.*, vol. 48, pp. 205–220, 2019.

[23] C. O. Sakar and O. Kursun, "Telediagnosis of parkinson's disease using measurements of dysphonia," *Journal of Medical Systems*, vol. 34, no. 4, pp. 591–599, Aug 2010.

[24] J. I. Godino-Llorente, S. Aguilera-Navarro, and P. G. Vilda, "Lpc, LPCC and MFCC parameterisation applied to the detection of voice impairments," in *INTERSPEECH*, 2000, pp. 965–968.

[25] J. A. G. García, L. Moro-Velázquez, and J. I. Godino-Llorente, "On the design of automatic voice condition analysis systems. part I: review of concepts and an insight to the state of the art," *Biomed. Sign. Process. and Cont.*, vol. 51, pp. 181 – 199, 2019.

[26] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "Spectral and cepstral analyses for parkinson's disease detection in spanish vowels and words," *Expert Systems*, vol. 32, no. 6, pp. 688–697, 2015.

[27] A. Tsanas, M. A. Little, P. E. McSharry, J. L. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," *IEEE Trans. Biomed. Engineering*, vol. 59, no. 5, pp. 1264–1271, 2012.

[28] J. Mekyska *et al.*, "Robust and complex approach of pathological speech signal analysis," *Neurocomputing*, vol. 167, pp. 94 – 111, 2015.

[29] N. Cummins, A. Baird, and B. Schuller, "Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning," *Methods*, vol. 151, pp. 41–54, 2018.

[30] A. Benba, A. Jilbab, and A. Hammouch, "Discriminating between patients with parkinson's and neurological diseases using cepstral analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 10, pp. 1100–1108, Oct 2016.

[31] G. Aneeja and B. Yegnanarayana, "Single Frequency Filtering Approach for Discriminating Speech and Nonspeech," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 4, pp. 705–717, Apr 2015.

[32] S. R. Kadiri and B. Yegnanarayana, "Epoch extraction from emotional speech using single frequency filtering approach," *Speech Communication*, vol. 86, pp. 52–63, 2017.

[33] ——, "Analysis and detection of phonation modes in singing voice using excitation source features and single frequency filtering cepstral coefficients (SFFCC)," in *Proc. INTERSPEECH*, 2018, pp. 441–445.

[34] ——, "Analysis of aperiodicity in artistic noh singing voice using an impulse sequence representation of excitation source," *J. Acoust. Soc. Am.*, vol. 146, no. 6, pp. 4446–4457, 2019.

[35] K. N. R. K. R. Alluri *et al.*, "SFF anti-spoofer: IIIT-H submission for automatic speaker verification spoofing and countermeasures challenge 2017," in *Proc. INTERSPEECH*, 2017, pp. 107–111.

[36] P. A. Torres-Carrasquillo, E. Singer, M. A. Kohler, R. J. Greene, D. A. Reynolds, and J. R. Deller Jr., "Approaches to language identification using Gaussian mixture models and shifted delta cepstral features," in *Proc. ICSLP*, 2002.

[37] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.

[38] J. R. Orozco-Arroyave et al., "New Spanish Speech Corpus Database for the Analysis of People Suffering from Parkinson's Disease." in *LREC*, 2014, pp. 342–347.

[39] J. Rusz, R. Cmejla *et al.*, "Imprecise vowel articulation as a potential early marker of parkinson's disease: Effect of speaking task," *The Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 2171–2181, 2013.