INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

On Modeling Glottal Source Information for Phonation Assessment in Parkinson's Disease

J. C. Vásquez-Correa[1,2], *J. Fritsch*[3,4], *J. R. Orozco-Arroyave*[1,2], *E. Nöth*[1], *M. Magimai-Doss*[3]

[1]Pattern Recognition Lab. Friedrich-Alexander Universität, Erlangen-Nürnberg, Germany
[2] Faculty of Engineering. Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín, Colombia
[3] Idiap Research Institute, Martigny, Switzerland
[4] École polytechnique fédérale de Lausanne, Switzerland

juan.vasquez@fau.de, julian.fritsch@idiap.ch, rafael.orozco@udea.edu.co, elmar.noeth@fau.de, mathew@idiap.ch

Abstract

Parkinson's disease produces several motor symptoms, including different speech impairments that are known as hypokinetic dysarthria. Symptoms associated to dysarthria affect different dimensions of speech such as phonation, articulation, prosody, and intelligibility. Studies in the literature have mainly focused on the analysis of articulation and prosody because they seem to be the most prominent symptoms associated to dysarthria severity. However, phonation impairments also play a significant role to evaluate the global speech severity of Parkinson's patients. This paper proposes an extensive comparison of different methods to automatically evaluate the severity of specific phonation impairments in Parkinson's patients. The considered models include the computation of perturbation and glottal-based features, in addition to features extracted from a zero frequency filtered signals. We consider as well end-to-end models based on 1D CNNs, which are trained to learn features from the raw speech waveform, reconstructed glottal signals, and zero-frequency filtered signals. The results indicate that it is possible to automatically classify between speakers with low versus high phonation severity due to the presence of dysarthria and at the same time to evaluate the severity of the phonation impairments on a continuous scale, posed as a regression problem.

Index Terms: Parkinson's disease, Phonation, Glottal source modeling, Zero-frequency filtering

1. Introduction

Parkinson's disease (PD) is a neurological disorder characterized by the progressive loss of dopaminergic neurons in the midbrain. It affects approximately 10 million people worldwide, with a doubling of the global burden over the past 25 years because the increase in longevity of people thanks to modern medicine methods [1]. PD produces different motor and non-motor symptoms in the patients. Motor symptoms include tremor, slowed movement, rigidity, bradykinesia, lack of coordination, among others. Approximately 70-90% of PD patients develop a multidimensional speech impairment called hypokinetic dysarthria [2, 3], which manifests itself typically in the imprecise articulation of consonants and vowels, monoloudness, monopitch, inappropriate silences and rushes of speech, dysrhythmia, reduced vocal loudness, and harsh or breathy vocal quality. All these symptoms affect the phonation, articulation, prosody, and intelligibility aspects of the speech of PD patients [4, 5, 6].

Dysarthria severity is usually evaluated with perceptual scales such as the Frenchay dysarthria assessment [7], or the Radbound dysarthria assessment [5], which evaluate different speech dimensions such as phonation, articulation, prosody, resonance, among others. Different studies in the literature have focused on the automation of the evaluation process of these speech dimensions in order to assess the global dysarthria severity of patients. Most of those studies have mainly focused on the automatic analysis of articulation and prosody because they seem to be the most prominent symptoms associated to dysarthria severity. Articulation impairments have been modeled with speech features based on the vowel space area [8], formant frequencies, voiced onset time [9], the energy content in onset transitions [10], and recent models based on convolutional neural networks (CNNs) [11] and posterior probabilities of certain phonemic classes [12, 13]. Prosody deficits have been commonly evaluated with features related to pitch, intensity and duration [14, 15].

Despite the fact that articulation and prosody are the most studied speech dimensions in hypokinetic dysarthria, phonation impairments also play a significant role to evaluate the global speech severity of PD patients. Phonation symptoms are related to the stability and periodicity of the vocal fold vibration, and with difficulties in the process of producing air in the lungs to make the vocal folds vibrate. Different phonation deficits appear in PD patients' speech, including differences in glottal noise compared to healthy speakers, incomplete vocal fold closure, and vocal folds bowing, which are typically characterized with measures such as noise to harmonics ratio, glottal to noise excitation ratio, and voice turbulent index, among others [16]. Additional phonation features include perturbation measures such as jitter, shimmer, amplitude perturbation quotient (APQ), pitch perturbation quotient (PPQ), and nonlinear dynamics measures [17, 18], as well as features extracted from the reconstruction of the glottal source signal such as the quasi open quotient, the normalized amplitude quotient, and the harmonic richness factor [19, 20, 21]. However, it is not clear whether these traditional features are able to properly characterize specific phonatory impairments that appear in the speech of PD patients because they are usually only considered to classify PD vs. healthy control (HC) speakers.

This paper proposes a comparison of a set of models to evaluate specific phonation symptoms related to the breathing capabilities of PD patients. The considered models include perturbation and glottal features such as the previously described, in addition to features extracted from signals obtained from a zero frequency filtering (ZFF) method [22], proposed originally to characterize glottal closure instants (GCIs). The considered methods also include the use of raw waveform CNNs [23, 24],

http://dx.doi.org/10.21437/Interspeech.2021-1084

which are designed to extract features from the speech waveforms, reconstructed glottal waves, and ZFF signals. To the best of our knowledge, this is the first study that performs a comparison regarding different methods to evaluate specific phonation impairments that appear in PD patients' speech.

The paper is organized as follows. Sections 2 and 3 present the methods and the data. Sections 4 and 5 present the results and conclusion, respectively.

2. Methods

The phonatory impairments in PD patients are evaluated using different feature extraction strategies, which are computed from raw speech, reconstructed glottal waves, and ZFF signals.

2.1. Glottal source reconstruction

We considered two different methods to reconstruct the glottal source signals. The first one being the classical Iterative and/or Adaptive Inverse Filtering (IAIF) [25], which is based on linear prediction (LP) filters that are computed in a two stage procedure. This method is based on an iterative refinement of both the vocal tract and the glottal components. The glottal excitation is obtained by cancelling the effects of the vocal tract and lip radiation by inverse filtering. The second method is the glottal closure/opening instant estimation forward-backward algorithm (GEFBA) [26], which is based on detecting instants of significant excitation (epochs) for high resolution glottal activity detection. GEFBA estimates the instants of glottal closures for determining the boundaries of glottal activity by assuming that two consecutive voiced regions differ by a distance greater than twice the maximum pitch period.

2.2. ZFF

ZFF is designed for epoch extraction, and aims to remove all the influence from the vocal tract system in the speech waveform. The core idea of ZFF is to exploit the fact that glottal closure produces an excitation similar to an impulse [27]. This information is present in all the frequencies, including 0 Hz. To obtain the information present at 0 Hz, the speech waveform is filtered through a cascade of two 0 Hz resonators followed by a trend removal operation. By passing the signal through the resonators, the effect of vocal tract resonance is minimized. The ZFF signal oscillates at local fundamental frequency, and the negative to positive zero crossings gives epoch locations.

Figure 1 shows the difference between the raw speech waveform, the IAIF and GEFBA methods used to reconstruct the glottal signal, and the ZFF signal. These four signals are used to evaluate the phonation impairments that appear in PD patients.

2.3. Perturbation features

Perturbation features are used to model abnormal patterns in the vocal fold vibrations. Perturbation features are extracted from the raw speech waveforms and from the ZFF signals. The feature set includes seven descriptors: (1-2) *Jitter* and *shimmer* to describe temporal perturbations in the fundamental frequency and amplitude of the speech signal, respectively [17]. (3) APQ, which aims to measure the long-term variability of the peak-to-peak amplitude of the speech signal, by using a smoothing factor of 11 voiced periods. (4) PPQ to measure the long-term variability of the fundamental frequency, with a smoothing factor of five periods. (5-6) The first and second derivatives of the fundamental frequency contour, and (7) the log-energy as a measure of loudness. Four statistical functionals are calculated per descriptor (mean, standard deviation, skewness, and kurtosis), forming a 28–dimensional feature vector per utterance.

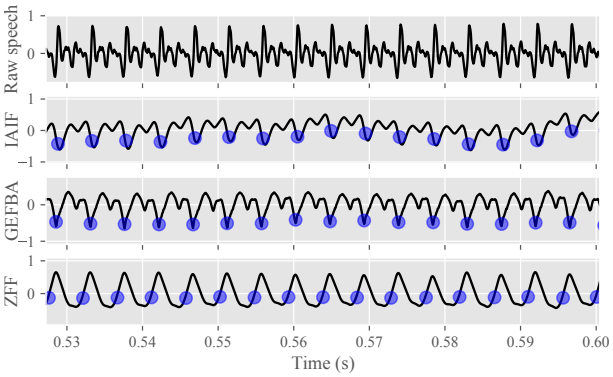

Figure 1: *Different signals extracted from the phonation of a sustained vowel /ah/, and considered to evaluate the phonation impairments from PD patients. Blue dots indicate the detected GCIs.*

2.4. Glottal features

Glottal features are computed over the reconstructed glottal signals using the IAIF and the GEFBA methods. Glottal features are focused on specific parts of the glottal cycle such as the opening and closing phases. The proposed feature vector comprises nine descriptors: (1) the temporal variability between consecutive GCIs, (2-3) the average and variance of the *Open Quotient (OQ)*, which is the ratio of the duration of the opening phase and the duration of the glottal cycle. (4-5) the average and variance of the *Normalized Amplitude Quotient (NAQ)*, which is defined as the ratio of the maximum of the glottal flow and the minimum of its derivative. (6-7) the average and variance of *H1H2*, which is the difference between the first two harmonics of the glottal flow signal. Finally (8-9) are average and variance of the *Harmonic Richness Factor (HRF)*, which is the ratio of the sum of the harmonics amplitude and the amplitude of the fundamental frequency. These features are computed for every glottal cycle in segments with 200 ms length in order to measure short-term perturbations of the glottal flow. Finally, similar to the perturbation features, four statistical functionals are calculated per descriptor (mean, standard deviation, skewness, and kurtosis), forming a 36–dimensional feature vector per utterance. The source code to extract the IAIF-based glottal signals and to compute the perturbation and glottal features is available online for the research community via the DisVoice toolkit[1].

2.5. Raw waveform CNNs

Raw-waveform CNNs directly model raw signals by applying 1D filters on the raw waveform. From an input window of 1 second, the architecture applies four 1D convolutional layers, followed by a hidden layer and an output-layer. In order to guide the learning procedure, the first layer's filters' kernel length is relevant. In previous work we distinguished sub-segmental (filter length $<$ 1 pitch period) and segmental (filter length $>$ 1 pitch period) filtering, though in this work we only deployed the sub-segmental filtering (cf. Table 1 in [24]). On raw signals, a sub-segmental architecture tends to focus on vocal-tract related information [23], which is not desired in this work. However, on signals that were filtered to enhance voice-source related characteristics, a sub-segmental filtering is more suitable [28], which is why it was consistently considered for this task. As pointed earlier, we used the same architecture as in [24]. The

[1] https://github.com/jcvasquezc/DisVoice

CNNs were trained using Keras-Tensorflow framework. The classification task was trained with binary cross-entropy loss function and a sigmoid function at the output; the regression task with mean-squared-error loss and a linear output function. In both cases, the starting learning rate is $1e-3$, which is halved after an epoch in which the validation loss did not reduce. Early stopping method was used to stop the training.

3. Data

The proposed systems are evaluated on the PC-GITA corpus [29]. The data comprises of utterances from 50 PD patients and 50 HC subjects, Colombian Spanish native speakers. The participants were asked to pronounce 10 sentences, six diadochokinetic (DDK) exercises, one text with 36 words, the sustained phonation of vowels, and a monologue. All patients were evaluated by a neurologist expert according to the MDS-UPDRS-III scale [30], and they were recorded in ON state. The dysarthria severity of the participants was evaluated according to the m-FDA scale [6], which consists of 13 items and evaluates seven aspects of the speech including breathing, lips movement, palate/velum movement, laryngeal movement, intelligibility, and monotonicity. Each item ranges from 0 to 4 (integer values), thus the total score ranges from 0 (healthy speech) to 52 (completely dysarthric). Two items of the m-FDA scale are related to phonation impairments of the patients and include breathing duration (BD) and breathing capacity (BC) when participants pronounce sustained phonation of vowels and DDK tasks. The ratings of such items are used to evaluate the proposed models. We consider as well the global m-FDA breathing impairment score, which combines information about BD and BC (it ranges from 0 to 8). For this study, we only considered data from the phonations of sustained vowels and DDK tasks, which were the recordings used by the phoniatrician to label the phonation severity of the participants. Table 1 shows clinical and demographic information from the participants of this study.

Table 1: *Demographic information from the participants in this study.* ***BD****: breathing duration,* ***BC****: breathing capacity.*

	PD (n=50)	**HC (n=50)**	**PD vs. HC**	**F vs. M**
Sex (F/M)	25/25	25/25	–	–
Age	61.0 (9.3)	61.0 (9.4)	0.49^a	0.29
Years since diagnosis	10.6 (9.1)	–	–	–
MDS-UPDRS-III	37.7 (18.1)	–	–	–
MDS-UPDRS-speech	1.3 (0.8)	–	–	–
Total m-FDA	28.8 (8.3)	8.5 (7.4)	$\ll 0.005^a$	0.28^a
m-FDA-BD	2.6 (1.0)	1.0 (0.9)	$\ll 0.005^a$	0.21^a
m-FDA-BD (high/low)	37/13	8/42	$\ll 0.005^b$	0.71^b
m-FDA-BC	2.5 (0.9)	0.7 (0.7)	$\ll 0.005^a$	0.25^a
m-FDA-BC (high/low)	37/13	2/48	$\ll 0.005^b$	0.12^b
m-FDA breath	5.1 (1.7)	1.7 (1.4)	$\ll 0.005^a$	0.18^a
m-FDA breath (high/low)	40/10	8/42	$\ll 0.005^b$	0.84^b

[a] p-values calculated using Mann-Whitney U tests
[b] p-values calculated using Chi-squared tests

The m-FDA labels for BD and BC are converted into high/low scores based on a threshold (median value of the scores assigned to the patients). Those subjects with scores lower than two are assigned with low phonation severity. Conversely, subjects with the item higher or equal to two are labeled as patients with high phonation impairments. Hence, we decided to solve either a regression problem on the full range of the m-FDA sub-scores or a classification problem to evaluate low vs. high phonation impairment. The distribution between PD and HC subjects and the assigned m-FDA labels are gender-balanced (all p-values>0.05) and age-balanced (Spearman's correlation between age and m-FDA scores are lower than 0.2 with all p-values>0.05). Hence, the influence produced by demographic data in our problem can be discarded.

4. Experiments and Results

The extracted perturbation and glottal features were used to train a support vector machine (SVM) classifier with a Gaussian kernel. The model is validated following a nested 10-fold speaker independent stratified cross-validation strategy. The hyper-parameters C and γ were optimized in a randomized-search strategy [31] based on the development set accuracy. Similarly, the raw waveform CNNs were validated on the same 10-fold cross-validation strategy so that the results are comparable. All systems are trained to solve either the classification problem (low vs. high phonation impairments) or the regression problem (severity of the phonation impairment). All systems are applied to the three problems described in Section 3, namely breathing duration, breathing capacity, and global breathing impairment. The latter being the combination of the breathing duration and capacity scores.

The results obtained classifying high vs. low phonation impairments are shown in Table 2. In general, the best results are observed using perturbation features computed either from the raw speech waveform or from the ZFF signals. Regarding the two methods for glottal source estimation, higher accuracies are observed with the glottal signals computed using the GEFBA method. The accuracies obtained with the raw waveform CNNs are not as high as expected. However, note that moderate results are observed when the CNNs are trained with the ZFF signals.

Table 2: *Results classifying the different low vs. high phonation impairments in PD patients.*

Signal	Features	ACC [%]	F-score	SENS [%]	SPEC [%]
m-FDA Breathing duration					
Raw	Perturbation	78	0.779	80	76
Raw	CNN	65	0.521	47	80
IAIF	Glottal	71	0.703	62	78
IAIF	CNN	60	0.289	26	89
GEFBA	Glottal	76	0.752	64	85
GEFBA	CNN	56	0.370	44	66
ZFF	**Perturbation**	79	0.786	73	84
ZFF	**CNN**	70	0.577	56	83
m-FDA Breathing capacity					
Raw	**Perturbation**	84	0.830	77	89
Raw	CNN	65	0.437	35	83
IAIF	Glottal	65	0.627	51	74
IAIF	CNN	43	0.344	54	40
GEFBA	Glottal	72	0.714	74	70
GEFBA	CNN	44	0.194	29	54
ZFF	Perturbation	80	0.793	79	80
ZFF	**CNN**	69	0.425	37	89
m-FDA Global Breathing impairments					
Raw	Perturbation	76	0.758	69	83
Raw	CNN	56	0.534	56	55
IAIF	Glottal	60	0.592	48	71
IAIF	CNN	54	0.542	77	32
GEFBA	Glottal	71	0.708	65	77
GEFBA	CNN	49	0.476	57	43
ZFF	**Perturbation**	76	0.760	77	75
ZFF	**CNN**	66	0.555	52	79

The accuracy to assess breathing duration ranges from 56 to 79% depending on the considered method. The highest accuracy is obtained with the computation of perturbation features

over the ZFF signals. Similar accuracies are observed with the raw speech waveform. The highest accuracy for the breathing capacity (84%) is obtained as well with the perturbation features, but in this case computed upon the raw speech waveform, followed by the perturbation features computed upon the ZFF signals. Finally, the accuracy for the assessment of the global breathing impairments ranges from 54 to 76%. Similar accuracies are observed with the perturbation features computed upon the raw speech waveform and the ZFF signals.

The results about the continuous evaluation of the phonation impairments of the participants using a regression approach are presented in Table 3 for the three addressed problems. The results are presented in terms of Pearson's correlation coefficient (r), Spearman's correlation coefficient (ρ), and mean absolute error (MAE). Strong correlations are obtained for the three addressed problems, especially using the perturbation features computed upon the raw speech waveforms and the ZFF signals. Similar to the classification results, the correlations observed with the raw waveform CNNs are not as high as expected; however, this can be explained by the little amount of data and the reduced variability of the labels to solve the regression problems. In addition, the results observed with the glottal signals estimated with the GEFBA method outperformed the ones obtained with the glottal signals estimated with the classic IAIF algorithm. Particularly, the best result is observed for the assessment of the global motor performance (ρ=0.741) probably because this is the scale with more variability in the labels (it ranges from 0 to 8), as compared to the breathing duration and breathing capacity, which only range from 0 to 4.

Table 3: *Results evaluating the severity of the different phonation impairments in PD patients.*

Signal	Features	r	ρ	MAE
m-FDA Breathing duration				
Raw	**Perturbation**	0.659	0.662	0.86
Raw	**CNN**	0.379	0.427	1.12
IAIF	Glottal	0.436	0.444	1.00
IAIF	CNN	0.016	-0.034	1.46
GEFBA	Glottal	0.426	0.457	1.00
GEFBA	CNN	0.214	0.182	1.43
ZFF	Perturbation	0.591	0.603	1.00
ZFF	CNN	0.075	0.076	1.86
m-FDA Breathing capacity				
Raw	**Perturbation**	0.660	0.659	0.86
Raw	**CNN**	0.354	0.315	1.31
IAIF	Glottal	0.308	0.429	1.00
IAIF	CNN	0.003	-0.039	1.44
GEFBA	Glottal	0.460	0.510	1.00
GEFBA	CNN	-0.121	-0.109	1.45
ZFF	Perturbation	0.659	0.683	0.89
ZFF	CNN	0.125	0.102	1.67
m-FDA Global Breathing impairments				
Raw	**Perturbation**	0.732	0.741	1.40
Raw	**CNN**	0.352	0.341	1.27
IAIF	Glottal	0.129	0.474	2.00
IAIF	CNN	0.065	0.098	1.58
GEFBA	Glottal	0.528	0.620	1.91
GEFBA	CNN	-0.029	0.024	1.34
ZFF	Perturbation	0.673	0.714	1.54
ZFF	CNN	0.260	0.250	1.62

Figure 2 shows in more detail the best results obtained evaluating the global breathing impairments of the participants. The predictions are obtained using perturbation features computed over the raw speech waveform.

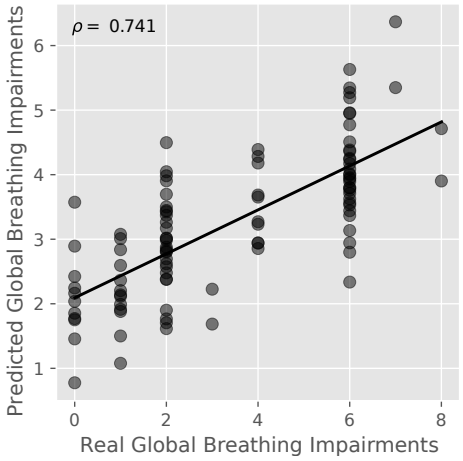

Figure 2: *Best result obtained evaluating the severity of the phonation impairments of PD patients using perturbation features computed over the raw speech waveform.*

5. Conclusion

This paper addressed the evaluation of the severity of different phonation impairments that appear in PD patients due to the presence of hypokinetic dysarthria. An extensive comparison among classical and novel methods is performed in order to accurately estimate the level of phonatory impairments in the patients. The considered methods include the computation of pertubation and glottal-based features, and the use of raw waveform CNNs that extract features directly from the raw signals. The described methods are applied over different signals, including the raw speech waveform, two versions of the glottal source signal, and the signals obtained after the application of a ZFF. Overall, utterance-level functionals of perturbation features give more robust estimates of the addressed problems.

The results indicate that it is possible to discriminate between low vs. high phonatory impairments with accuracies ranging from 76 to 84%. The most accurate results were observed with the use of perturbation features computed upon the raw speech waveform and the ZFF signals. The continuous evaluation of the phonation impairments of the participants was posed as a regression problem, where strong correlations were observed (ρ up to 0.741), using perturbation features computed over the raw speech waveforms. The results obtained using the raw waveform CNNs do not match that performance, probably because of the reduced amount of data and the low variability of the labels. Our future work aims at transferring the knowledge from this study into the assessment of phonation impairments in other diseases that affect the phonatory system of the patients such as different laryngeal pathologies, other neurodegenerative diseases, COVID-19 patients, among others.

6. Acknowledgements

This project received funding from the EU Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Action ITN-ETN project TAPAS, Grant Agreement No. 766287. This work was partially funded by CODI from the University of Antioquia, grant # PRG2017-15530.

7. References

[1] E. R. Dorsey *et al.*, "Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016," *The Lancet Neurology*, vol. 17, no. 11, pp. 939–953, 2018.

[2] J. A. Logemann *et al.*, "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients," *Journal of Speech and hearing Disorders*, vol. 43, no. 1, pp. 47–57, 1978.

[3] A. K. Ho *et al.*, "Speech impairment in a large sample of patients with Parkinson's disease," *Behavioural neurology*, vol. 11, no. 3, pp. 131–137, 1998.

[4] J. Rusz, R. Cmejla, H. Ruzickova, and E. Ruzicka, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease," *The journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 350–367, 2011.

[5] S. Knuijt, J. G. Kalf, B. G. van Engelen, B. J. de Swart, and A. C. Geurts, "The radboud dysarthria assessment: development and clinimetric evaluation," *Folia Phoniatrica et Logopaedica*, vol. 69, no. 4, pp. 143–153, 2017.

[6] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, T. Bocklet, and E. Nöth, "Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease," *Journal of communication disorders*, vol. 76, pp. 21–36, 2018.

[7] P. M. Enderby and R. Palmer, *FDA-2: Frenchay Dysarthria Assessment: Examiner's Manual*. Pro-ed, 2008.

[8] J. Rusz *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.

[9] D. Montaña, Y. Campos-Roca, and C. J. Pérez, "A Diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease," *Computer methods and programs in biomedicine*, vol. 154, pp. 89–97, 2018.

[10] J. R. Orozco-Arroyave, *Analysis of speech of people with Parkinson's disease*. Logos Verlag Berlin GmbH, 2016, vol. 41.

[11] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, and E. Nöth, "Convolutional Neural Network to Model Articulation Impairments in Patients with Parkinson's Disease," in *INTERSPEECH*, 2017, pp. 314–318.

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

[13] L. Moro-Velazquez *et al.*, "Phonetic relevance and phonemic grouping of speech in the automatic detection of Parkinson's Disease," *Scientific reports*, vol. 9, no. 1, pp. 1–16, 2019.

[14] T. Bocklet, S. Steidl, E. Nöth, and S. Skodda, "Automatic evaluation of Parkinson's speech-acoustic, prosodic and voice related cues." in *Interspeech*, 2013, pp. 1149–1153.

[15] R. Norel *et al.*, "Speech-based characterization of dopamine replacement therapy in people with Parkinson's disease," *NPJ Parkinson's disease*, vol. 6, no. 1, pp. 1–8, 2020.

[16] Y. Tanaka, M. Nishio, and S. Niimi, "Vocal acoustic characteristics of patients with Parkinson's disease," *Folia Phoniatrica et logopaedica*, vol. 63, no. 5, pp. 223–230, 2011.

[17] T. Arias-Vergara *et al.*, "Parkinson's disease and aging: analysis of their effect in phonation and articulation of speech," *Cognitive Computation*, vol. 9, no. 6, pp. 731–748, 2017.

[18] C. M. Travieso *et al.*, "Detection of different voice diseases based on the nonlinear characterization of speech signals," *Expert Systems with Applications*, vol. 82, pp. 184–195, 2017.

[19] S. R. Kadiri and P. Alku, "Analysis and detection of pathological voice using glottal source features," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 367–379, 2019.

[20] M. Novotnỳ *et al.*, "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," *Biomedical Signal Processing and Control*, vol. 57, p. 101818, 2020.

[21] N. P. Narendra and P. Alku, "Glottal source information for pathological voice detection," *IEEE Access*, vol. 8, pp. 67 745–67 755, 2020.

[22] K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1602–1613, 2008.

[23] H. Muckenhirn, M. Magimai-Doss, and S. Marcell, "Towards directly modeling raw speech signal for speaker verification using cnns," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 4884–4888.

[24] J. Fritsch, S. P. Dubagunta, and M. M. Doss, "Estimating the Degree of Sleepiness by Integrating Articulatory Feature Knowledge in Raw Waveform Based CNNS," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 6534–6538.

[25] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech communication*, vol. 11, no. 2-3, pp. 109–118, 1992.

[26] A. I. Koutrouvelis, G. P. Kafentzis, N. D. Gaubitch, and R. Heusdens, "A fast method for high-resolution voiced/unvoiced detection and glottal closure/opening instant estimation of speech," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 2, pp. 316–328, 2015.

[27] P. Gangamohan and B. Yegnanarayana, "A robust and alternative approach to zero frequency filtering method for epoch extraction," in *Proc. Interspeech 2017*, 2017, pp. 2297–2300.

[28] S. P. Dubagunta, B. Vlasenko, and M. Magimai.-Doss, "Learning voice source related information for depression detection," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2019.

[29] J. R. Orozco-Arroyave *et al.*, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)*, 2014, pp. 342–347.

[30] C. G. Goetz *et al.*, "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Movement disorders*, vol. 23, no. 15, pp. 2129–2170, 2008.

[31] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization." *Journal of machine learning research*, vol. 13, no. 2, 2012.