INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Which aspects of motor speech disorder are captured by Mel Frequency Cepstral Coefficients? Evidence from the change in STN-DBS conditions in Parkinson's disease

Vojtěch Illner[1], *Petr Krýže*[1], *Jan Švihlík*[1,2], *Mario Sousa*[3], *Paul Krack*[3], *Elina Tripoliti*[4], *Robert Jech*[5], *Jan Rusz*[1,3,5]

[1]Faculty of Electrical Engineering, Czech Technical University in Prague, Czechia
[2]Faculty of Chemical Engineering, University of Chemistry and Technology, Czechia
[3]Movement Disorders Center, Department of Neurology, University Hospital of Bern, Switzerland
[4]UCL Queen Square Institute of Neurology, University College London, United Kingdom
[5]Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University, Czechia

Abstract

One of the most popular speech parametrizations for dysarthria has been Mel Frequency Cepstral Coefficients (MFCCs). Although the MFCCs ability to capture vocal tract characteristics is known, the reflected dysarthria aspects are primarily undisclosed. Thus, we investigated the relationship between key acoustic variables in Parkinson's disease (PD) and the MFCCs. 23 PD patients were recruited with ON and OFF conditions of Deep Brain Stimulation of the Subthalamic Nucleus (STN-DBS) and examined via a reading passage. The changes in dysarthria aspects were compared to changes in a global MFCC measure and individual MFCCs. A similarity was found in 2nd to 3rd MFCCs changes and voice quality. Changes in 4th to 9th MFCCs reflected articulation clarity. The global MFCC parameter outperformed individual MFCCs and acoustical measures in capturing STN-DBS conditions changes. The findings may assist in interpreting outcomes from clinical trials and improve the monitoring of disease progression.

Index Terms: Mel Frequency Cepstral Coefficients, Parkinson's disease, speech disorder, dysarthria, acoustic analysis

1. Introduction

Speech represents the most complex quantitative marker of motor function, vastly sensitive to damage to the brain's neural structures [1]. Speech dysfunctions presence has been documented in a number of progressive neurological diseases, such as Parkinson's disease (PD) [2]. In recent years, due to technological and computational advances, there has been an increasing interest in the use of speech for monitoring disease progression, symptoms severity, and a potential diagnostic aid [1, 3]. Improved ease in obtaining voice recordings using either smartphones [4, 5, 6], or telemonitoring homecare systems [7] offers intriguing advances as speech evaluation is inexpensive, non-invasive, simple to administer, and scalable to a large population.

Analysis of the acquired speech data and potential pathology is primarily interpreted using physiological speech patterns describing vocal tract abilities, such as articulation, pitch variability, loudness, rhythm, and phonation [8]. However, speech can also be parametrized by sets with low interpretability, but high performance, such as Mel Frequency Cepstral Coefficients (MFCCs) and their derivatives [9, 10, 11], Relative Spectral Transform - Perceptual Linear Prediction parameters [12], and deep neural networks embeddings [13]. The undisclosed explanation poses no problem for complex frameworks such as speech recognition but limits the use in clinically related studies and, most notably, in clinical trials [14].

Nevertheless, as one of the most popular speech parametrizations, MFCCs remain highly relevant due to their capacity to capture considerable information from the speech waveform. While being a long-standing essential part of frameworks for speech recognition [15], speaker detection [16], speech synthesis [17], and many others, in the last decade, they also gained interest in studies focused on speech impairments in neurological diseases [5, 9, 10, 11, 18]. However, the complete relationship between the MFCCs and particular speech dysfunctions remains clouded. In [9], authors comment that the coefficients detect subtle changes in the motion of the articulators (tongue, lips). Nonetheless, such an assumption has never been validated, while MFCCs can be easily influenced by other factors such as age, gender, speaking style, or recording procedure/microphone quality [19]. Most recently, in Roche's PD Mobile application designed for clinical trial measures in PD [5, 20], the speech performance of the patients was analyzed on a sustained phonation task using only the second coefficient, MFCC2, representing a low-to-high frequency energy ratio [8].

Although the MFCCs are emerging as one of the principal features in assessing speech impairments in neurological diseases, their interpretability remains limited. Therefore, we tested the sensitivity of MFCCs in a scenario covering Parkinsonian patients with ON and OFF conditions of Deep Brain Stimulation of the Subthalamic Nucleus (STN-DBS). Since the STN-DBS might substantially alter the patient's speech abilities [21], we expect to discover changes in MFCCs that might correspond to changed acoustical patterns of hypokinetic dysarthria.

2. Methods

2.1. Participants

A total of 23 individuals with PD (four females), with a mean age of 61.7 years (SD = 5.0, range: 53–72), who were treated with bilateral STN-DBS in combination with dopaminergic medication, were recruited for the study. The examination in the PD group was held in two conditions, including STN-DBS switched OFF (hereafter, the DBS OFF condition) and STN-DBS switched ON (hereafter, the DBS ON condition). Detailed clinical characteristics (clinical scores and DBS settings across individuals with PD) and experimental procedure description can be found in previous study [22]. As a healthy control

10.21437/Interspeech.2023-1744

(HC) group, 23 age- and sex-matched (four females) volunteers, a mean age of 61.5 years (SD = 5.6, range: 52–72), with no history of neurological or communication disorders, were recruited. All participants were native Czech speakers. The study complied with the Helsinki Declaration and was approved by the Ethics Committee of the General University Hospital, Prague, Czech Republic. Each participant provided written informed consent.

2.2. Speech examination

The patients were recorded after the individual therapeutical setting and were asked to perform phonetically balanced reading passage task of a standardized text of 313 words with a familiar, up-to-date vocabulary and grammatical structure. The audio recordings were conducted in a quiet room with a low ambient noise level using a head-mounted condenser microphone (Beyerdynamic Opus 55, Heilbronn, Germany) placed approximately $5\,\mathrm{cm}$ from the subject's mouth. Speech signals were sampled at $48\,\mathrm{kHz}$ with 16-bit resolution.

2.3. MFCCs computation

After the trial testing, the following procedure was established to calculate the first 16 MFCCs. The number 16 was set, similarly to [11, 23], as a tradeoff between studies using fewer coefficients, such as 12 or 13 [9, 24], and longer coefficients vectors, such as 20 [10]. The computations were conducted in MATLAB, Natick, USA.

Similarly, as in [10, 25], the audio input was first downsampled to $16\,\mathrm{kHz}$ with lowpass pre-filtering to guard against aliasing. Next, a pre-emphasis filter was applied to the samples with $\alpha = -0.95$. MFCCs were computed using MATLAB Auditory Toolbox functions. The entire signal is processed in frames using a Hamming window of a length $25\,\mathrm{ms}$ with 5% overlap. The frame‘s FFT magnitude is converted into Mel filterbank outputs using 13 linearly spaced filters followed by 27 log-spaced filters ranging from approximately 133 Hz to 6864 Hz. Next, a cosine transform of the $\log_{10}$ of the output is computed. The result is a vector of c_0-c_{16} MFCCs standard deviations across frames. The 0th coefficient, c_0, representing signal energy, is discarded.

Necessarily, a voice activity detector has to be applied, so the coefficients are used only in the segments of speech. In this study, dynamical thresholding of the spectral distance of the computed coefficients is utilized to mark segments without speech presence [26]. Coefficients in such segments are discarded.

Apart from analyzing individual c_1-c_{16}, a global MFCC measure is established, inspired by [24], as a mean of the standard deviation (std) of c_1-c_{16}. It is designed to represent the overall dynamic movement ability of individual vocal tract elements, as the individual MFCCs overlap the partitions of the frequency domain.

2.4. Physiological acoustic markers

To link the MFCCs to key dysarthria elements of PD, five acoustic variables with well-known pathophysiological interpretation were extracted from the speech waveform using the framework developed in [8].

Speech impairments in PD can be, for the most part, characterized by decreased voice quality, imprecise articulation, monoloudness, monopitch, deficits in speech timing, and inappropriate pauses [27]. A decrease in voice quality can be reflected by a lower Cepstral Peak Prominence (CPP) measure [28]. Aspects of imprecise articulation can be represented by a decrease in resonant frequency attenuation (RFA) measure, defined as the ratio between local second formant region maxima and local valley region minima. RFA is mainly sensitive to articulatory decay but may also be partly influenced by abnormal nasal resonance [8]. Monoloudness corresponds to a lower std of the speech energy (stdPWR) and monopitch to a lower std of estimated pitch contour (stdF0). Deficits in speech timing and rhythm, such as slowing or accelerating tempo, are reflected by the net speech rate (NSR) measure. Since the information about the length and occurrence of pauses is uncapturable by MFCCs, the measure representative for the description of pauses was omitted.

2.5. Statistical analysis

The following two experiments were conducted to assess the physiological nature of MFCCs.

First, the differences in the variables between DBS ON and OFF, called $\Delta_{\mathrm{OFF}}^{\mathrm{ON}}$, were calculated, representing the change in speech characteristics:

$$\Delta_{\mathrm{OFF}}^{\mathrm{ON}} v_i = v_i^{\mathrm{ON}} - v_i^{\mathrm{OFF}}, \tag{1}$$

where v_i^{ON}, v_i^{OFF} are the variables (MFCCs, global MFCC parameter, acoustical features) from DBS ON and DBS OFF groups, respectively. Then, Spearman correlation was computed between $\Delta_{\mathrm{OFF}}^{\mathrm{ON}}$ MFCC variables and $\Delta_{\mathrm{OFF}}^{\mathrm{ON}}$ acoustical variables.

Subsequently, the individual variables were compared in the three groups (HC, DBS ON, DBS OFF) using repeated measures analysis of variance (RM-ANOVA) followed by Bonferroni post-hoc correction, where the HC group is age- and sex-aligned with the DBS subjects and treated as associated measurement.

3. Results

The results from the first experiment are shown in Figure 1. Change in CPP was correlated with changes in lower MFCCs (c_2-c_3), $\rho > 0.48$, $p < 0.05$. Change in RFA correlated with c_4-c_9 coefficients changes, $\rho > 0.45$, $p < 0.05$. Changes in the global MFCC parameter achieved significant correlations with changes in CPP and RFA, $\rho = 0.46$, $p < 0.05$, partly also reflecting changes in stdF0 and NSR, $\rho = 0.41$, $p = 0.06$, resp. $\rho = -0.40$, $p = 0.06$.

The results from RM-ANOVA for c_1-c_{16} and global MFCC are shown in Figure 2. According to $F(1, 22)$ statistics, the global MFCC parameter achieved the highest overall significance in between-group differences ($F(1, 22) = 53.1$, $p < 0.001$ for HC vs. DBS OFF, $F(1, 22) = 19.1$, $p < 0.001$ for HC vs. DBS ON, $F(1, 22) = 8.4$, $p < 0.05$ for DBS ON vs. DBS OFF). Lower coefficients (c_1-c_5) demonstrate significant differences between HC and DBS ON or DBS OFF groups ($F(1, 22) > 8.0$, $p < 0.05$). However, significant contrast between DBS ON and OFF is present in higher coefficients, c_5-c_8 ($F(1, 22) > 5.8$, $p < 0.05$).

Figure 3 shows boxplots for the global MFCC parameter and acoustical features. Only the global MFCC parameter achieved a significant difference between DBS ON and OFF ($F(1, 22) = 8.4$, $p < 0.05$). RFA and stdF0 demonstrated significant contrast between HC and both DBS ON and OFF ($F(1, 22) > 11.7$, $p < 0.01$). NSR and stdPWR showed significant differences only between HC and DBS OFF ($F(1, 22) > 6.8$, $p < 0.05$).

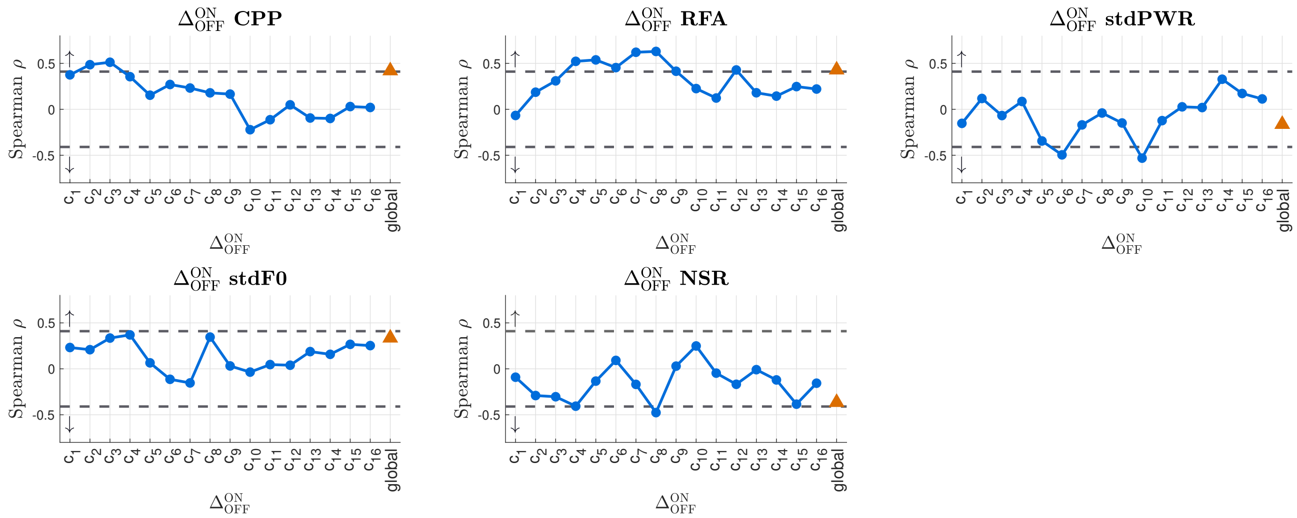

Figure 1: *Spearman correlation of $\Delta_{\mathrm{OFF}}^{\mathrm{ON}}$, differences in variables between DBS ON and DBS OFF, for individual c_1-c_{16} Mel Frequency Cepstral Coefficients (MFCC) and global MFCC parameter (mean standard deviation of c_1-c_{16}) to acoustical variables. Dashed lines represent the boundary of significant correlation with $p < 0.05$. Captions: CPP=Cepstral Peak Prominence, RFA=Resonant Frequency Attenuation, stdPWR=signal energy standard deviation, stdF0=pitch contour standard deviation, NSR=Net Speech Rate.*

4. Discussion

The present study explored the relationship between individual c_1-c_{16} MFCCs and physiologically interpretable acoustic features, including a global MFCC parameter composed of all the coefficients. Based on changes in speech characteristics between DBS ON and DBS OFF, we could relate the alterations of individual coefficients to specific acoustical markers. Moreover, the global parameter as well as some individual MFCCs were able to capture significant differences in speech production between the HC group and PD groups, and even between DBS ON and OFF, outperforming the traditional acoustical measures.

4.1. Relationship between MFCCs and acoustical measures

Changes in lower cepstral coefficients ($\Delta_{\mathrm{OFF}}^{\mathrm{ON}} c_2$, $\Delta_{\mathrm{OFF}}^{\mathrm{ON}} c_3$) significantly reflect changes in CPP, representing voice quality measure. CPP has been shown to strongly correlate with the increase in the severity of dysphonia and breathiness in various languages [28]. CPP integrates multiple acoustical measures related to lower speech frequencies, such as first rahmonic, pitch, waveform deviations, and noise perturbations [28]. Since both c_2 and c_3 are related to the signal energy, covering the corresponding range (approximately 200 - 500 Hz), the relationship with CPP and the ability to capture such characteristics become apparent.

Higher MFCCs, starting from $\Delta_{\mathrm{OFF}}^{\mathrm{ON}} c_4$ to $\Delta_{\mathrm{OFF}}^{\mathrm{ON}} c_9$, significantly correlate with the changes in RFA measure. RFA represents the second formant to anti-formant based system [8], i.e., special case of MFCC limited to frequency regions around the second formant. Therefore, RFA obviously provides comparable results to the MFCC system, although MFCCs cover more wide frequency range and are not dependent on the correct estimation of the position of the second formant. Both these MFCCs and RFA metrics (at least considering the frequency range between 4th to 9th cepstral coefficients) might thus provide a measure of the dynamical ability of articulatory movement. Such a measure might supplement the traditional formant-based approaches reflecting the range of movement of articulators, which particularly vary with tongue placement position.

The changes in the global MFCC parameter, designed to represent the overall dynamic movement ability of individual vocal tract elements, are significantly correlated to changes in CPP and RFA as well, thus incorporating the captured speech characteristics from corresponding individual MFCCs. The pitch variability is mildly related and would likely significantly contribute to observed results with increasing sample size. Interestingly, NSR is negatively correlated with the global measure meaning that with an increased articulation rate, the articulation ability and the quality of voice decrease; however, the trend is not significant.

4.2. Capacity of individual MFCCs to capture within-group differences

Individual MFCCs expressed within-group speech characteristics differences with a high significance. Especially lower coefficients (c_1-c_5) achieved an excellent score in distinguishing HC and PD cohorts (including DBS ON and OFF groups). The results might be explained by the close connection between the coefficients and measures of CPP and stdF0, explored in the previous section. On the other hand, higher c_5-c_9 MFCCs, related to the RFA measure, outperformed the lower ones in terms of capturing changes between DBS ON and DBS OFF. The evidence is that distinctive and eminently recognizable speech changes between speech in HC and PD are represented by lower MFCCs, whereas higher (approximately c_5-c_9) reflect subtle changes in articulation ability and formant structure, present between DBS ON and DBS OFF conditions. High coefficients, c_{10}-c_{16}, do not appear to have a significant effect on the group differences between PD and HC groups. However, the statistics are much more powerful between DBS ON and OFF than their comparison to HC. Sporadic significant differences between DBS ON and DBS OFF in coefficients c_{10}, c_{14}, and c_{15} might be due to correspondence with particular high formant structures but also due to noise which is more present in higher frequencies.

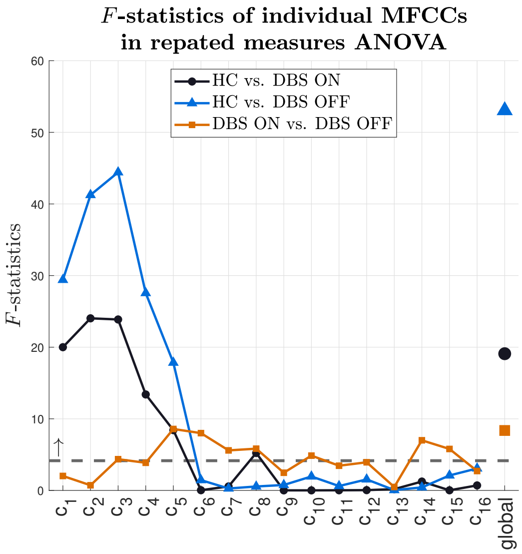

Figure 2: *Computed $F(1,22)$ statistics for individual c_1-c_{16} Mel Frequency Cepstral Coefficients (MFCC) and global MFCC parameter (mean standard deviation of c_1-c_{16}) according to repeated measures ANOVA. Dashed line represent boundary of significant difference of $p < 0.05$ based on Bonferroni post-hoc correction. Captions: ANOVA=analysis of variance, HC=healthy controls, DBS=deep brain stimulation.*

4.3. The use of the global MFCC parameter

Interestingly, the global MFCC measure demonstrated the most significant overall within-group difference. It achieved the highest score in separating HC and DBS OFF and a comparable score between HC and DBS ON and DBS ON and DBS OFF with the best-achieving coefficients. The fact that the global parameter comprehends the properties of the individual coefficients while maintaining high robustness might prove beneficial for its use in practice. For example, the c_2 coefficient demonstrated a significant, comparable score to the global measure between HC and both DBS states. However, it achieved poor results distinguishing between DBS ON and DBS OFF. The same can be analogously applied to, for example, c_6.

Additionally, compared to acoustical variables used in this study, the global MFCC demonstrates the highest overall significance between the DBS ON and DBS OFF conditions. The evidence might be due to the ability to reflect CPP and RFA, and partly stdF0 and NSR, altogether with capturing additional information about the individual vocal tract elements.

4.4. Limitations of the study

Only Czech-speaking subjects in a small cohort were part of the study. Further investigations should include other languages and larger sample sizes to confirm the findings. Additionally, it has been found that microphone quality and position highly influence amplitude-based features such as RFA [4]. Since we showed that MFCCs work on the same principle, the sensitivity of MFCCs to different experimental recording settings should be recognized and considered for large-scale use [29].

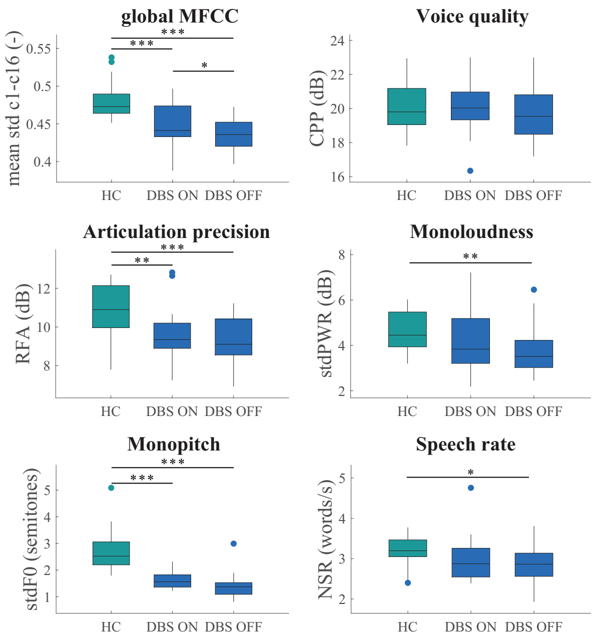

Figure 3: *Statistically significant group differences for the global Mel Frequency Cepstral Coefficients parameter (mean standard deviation of c_1-c_{16}) and acoustical variables with with ***, **, * referring to $p < 0.001$, $p < 0.01$, and $p < 0.05$ according to repeated measures analysis of variance with Bonferroni post-hoc correction. Middle bars represent median, and rectangles represent the interquartile range. Maximum and minimum values are by error bars. Outliers are marked as dots. Captions: DBS=deeb brain stimulation, CPP=Cepstral Peak Prominence, RFA=Resonant Frequency Attenuation, stdPWR=signal energy standard deviation, stdF0=pitch contour standard deviation, NSR=Net Speech Rate.*

5. Conclusions

The present study investigated the relationship between c_1-c_{16} MFCCs and five physiologically interpretable acoustical variables of hypokinetic dysarthria. In addition, a global MFCC parameter was established as mean std of c_1-c_{16}. A high correlation was shown between changes in low c_1-c_3 coefficients and changes in voice quality and signal envelope. Changes in coefficients from approximately c_4-c_9 reflect subtle changes in articulation ability and lower formants structures. The global MFCC measure comprehended the properties of the individual coefficients while maintaining high robustness and achieving significant between-group differences, outperforming all the single coefficients and acoustical measures. The findings may shed light on interpreting outcomes from speech assessment for future clinical trials.

6. Acknowledgments

This study was supported by the Czech Science Foundation (grant no. 21-14216L), SNF (grant no. 32003BL_197709), Czech Ministry of Education: National Institute for Neurological Research Programme EXCELES (ID Project no. LX22NPO5107) - Funded by the European Union – Next Generation EU, and Czech Technical University in Prague (grant no. SGS23/170/OHK3/3T/13).

7. References

[1] J. Duffy, *Motor speech disorders: substrates, differential diagnosis, and management*, 4th ed. Elsevier, 2019.

[2] J. Rusz, C. Bonnet, J. Klempíř, T. Tykalová, E. Baborová, M. Novotný, A. Rulseh, and E. Růžička, "Speech disorders reflect differing pathophysiology in parkinson's disease, progressive supranuclear palsy and multiple system atrophy," *Journal of Neurology*, vol. 262, pp. 992–1001, 2015.

[3] J. Rusz, J. Hlavnička, M. Novotný, T. Tykalová, A. Pelletier, J. Montplaisir, J. Gagnon, P. Dušek, A. Galbiati, S. Marelli, P. C. Timm, L. N. Teigen, A. Janzen, M. Habibi, A. Stefani, E. Holzknecht, K. Seppi, E. Evangelista, A. L. Rassu, Y. Dauvilliers, B. Högl, W. Oertel, E. K. S. Louis, L. Ferini-Strambi, E. Růžička, R. B. Postuma, and K. Šonka, "Speech biomarkers in rapid eye movement sleep behavior disorder and parkinson disease," *Annals of Neurology*, vol. 90, pp. 62–75, 2021.

[4] J. Rusz, J. Hlavnicka, T. Tykalova, M. Novotny, P. Dusek, K. Sonka, and E. Ruzicka, "Smartphone allows capture of speech abnormalities associated with high risk of developing parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, pp. 1495–1507, 2018.

[5] F. Lipsmeier, K. I. Taylor, T. Kilchenmann, D. Wolf, A. Scotland, J. Schjodt-Eriksen, W. Y. Cheng, I. Fernandez-Garcia, J. Siebourg-Polster, L. Jin, J. Soto, L. Verselis, F. Boess, M. Koller, M. Grundman, A. U. Monsch, R. B. Postuma, A. Ghosh, T. Kremer, C. Czech, C. Gossens, and M. Lindemann, "Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 parkinson's disease clinical trial," *Movement Disorders*, vol. 33, pp. 1287–1297, 2018.

[6] T. Kouba, V. Illner, and J. Rusz, "Study protocol for using a smartphone application to investigate speech biomarkers of parkinson's disease and other synucleinopathies," *BMJ Open*, vol. 12, 2022.

[7] C. L. Payten, D. D. Nguyen, D. Novakovic, J. O'Neill, A. M. Chacon, K. A. Weir, and C. J. Madill, "Telehealth voice assessment by speech language pathologists during a global pandemic using principles of a primary contact model: an observational cohort study protocol," *BMJ Open*, vol. 12, 2022.

[8] J. Hlavnička, R. Čmejla, T. Tykalová, K. Šonka, E. Růžička, and J. Rusz, "Automated analysis of connected speech reveals early biomarkers of parkinson's disease in patients with rapid eye movement sleep behaviour disorder," *Scientific Reports*, vol. 7, 2017.

[9] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinsons disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, pp. 1264–1271, 2012.

[10] A. Benba, A. Jilbab, and A. Hammouch, "Detecting patients with parkinson's disease using mel frequency cepstral coefficients and support vector machines," *International Journal on Electrical Engineering and Informatics*, vol. 7, pp. 297–307, 2015.

[11] T. Kapoor and R. K. Sharma, "Parkinson's disease diagnosis using mel-frequency cepstral coefficients and vector quantization," *International Journal of Computer Applications*, vol. 14, pp. 975–8887, 2011.

[12] 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, pp. 1100–1108, 2016.

[13] R. Algayres, M. S. Zaiem, B. Sagot, and E. Dupoux, "Evaluating the reliability of acoustic speech embeddings," *Interspeech 2020, 21st Annual Conference of the International Speech Communication Association, Virtual Event, Shanghai, China, 25-29 October 2020*, pp. 4621–4625, 2020.

[14] H. P. Rowe, S. Shellikeri, Y. Yunusova, K. V. Chenausky, and J. R. Green, "Quantifying articulatory impairments in neurodegenerative motor diseases: A scoping review and meta-analysis of interpretable acoustic features," *International Journal of Speech-Language Pathology*, 2022.

[15] S. Lokesh and M. R. Devi, "Speech recognition system using enhanced mel frequency cepstral coefficient with windowing and framing method," *Cluster Computing*, vol. 22, pp. 11 669–11 679, 2019.

[16] J. Martinez, H. Perez, E. Escamilla, and M. M. Suzuki, "Speaker recognition using mel frequency cepstral coefficients (mfcc) and vector quantization (vq) techniques," *CONIELECOMP 2012, 22nd International Conference on Electrical Communications and Computers*, pp. 248–251, 2012.

[17] D. Chazan, R. Hoory, G. Cohen, and M. Zibulski, "Speech reconstruction from mel frequency cepstral coefficients and pitch frequency," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 3, pp. 1299–1302, 2000.

[18] J. Rusz, R. Cmejla, T. Tykalova, H. Ruzickova, J. Klempir, V. Majerova, J. Picmausova, J. Roth, and E. Ruzicka, "Imprecise vowel articulation as a potential early marker of parkinson's disease," *The Journal of the Acoustical Society of America*, vol. 134, pp. 2171–2181, 2013.

[19] J. Rusz, M. Novotny, J. Hlavnicka, T. Tykalova, and E. Ruzicka, "High-accuracy voice-based classification between patients with parkinson's disease and other neurological diseases may be an easy task with inappropriate experimental design," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, pp. 1319–1321, 2017.

[20] F. Lipsmeier, K. I. Taylor, R. B. Postuma, E. Volkova-Volkmar, T. Kilchenmann, B. Mollenhauer, A. Bamdadian, W. L. Popp, W. Y. Cheng, Y. P. Zhang, D. Wolf, J. Schjodt-Eriksen, A. Boulay, H. Svoboda, W. Zago, G. Pagano, and M. Lindemann, "Reliability and validity of the roche pd mobile application for remote monitoring of early parkinson's disease," *Scientific Reports*, vol. 12, 2022.

[21] E. Tripoliti, M. Zrinzo, I. Martinez-Torres, E. Frost, S. Pinto, T. Foltynie, E. Holl, E. Petersen, M. Roughton, M. I. Hariz, and P. Limousin, "Effects of subthalamic stimulation on speech of consecutive patients with parkinson disease," *Neurology*, vol. 76, pp. 80–86, 2011.

[22] J. Svihlik, M. Novotny, T. Tykalova, K. Polakova, H. Brozova, P. Kryze, M. Sousa, P. Krack, E. Tripoliti, E. Ruzicka, R. Jech, and J. Rusz, "Long-term averaged spectrum descriptors of dysarthria in patients with parkinson's disease treated with subthalamic nucleus deep brain stimulation," *Journal of Speech, Language, and Hearing Research*, vol. 65, pp. 4690–4699, 2022.

[23] R. Fraile, N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, and C. Fredouille, "Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex," *Folia Phoniatrica et Logopaedica*, vol. 61, pp. 146–152, 2009.

[24] J. Rusz, J. Klempíř, E. Baborová, T. Tykalová, V. Majerová, R. Čmejla, E. Růžička, and J. Roth, "Objective acoustic quantification of phonatory dysfunction in huntington's disease," *PLoS ONE*, vol. 8, 2013.

[25] A. Slis, N. Lévêque, C. Fougeron, M. Pernon, F. Assal, and L. Lancia, "Analysing spectral changes over time to identify articulatory impairments in dysarthria," *The Journal of the Acoustical Society of America*, vol. 149, pp. 758–769, 2021.

[26] J. Haigh and J. Mason, "Robust voice activity detection using cepstral features," in *Proceedings of TENCON '93. IEEE Region 10 International Conference on Computers, Communications and Automation*, vol. 3, 1993, pp. 321–324.

[27] F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria," *Journal of Speech and Hearing Research*, vol. 12, pp. 246–269, 1969.

[28] R. Fraile and J. I. Godino-Llorente, "Cepstral peak prominence: A comprehensive analysis," *Biomedical Signal Processing and Control*, vol. 14, pp. 42–54, 2014.

[29] J. Rusz, T. Tykalova, L. O. Ramig, and E. Tripoliti, "Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders," *Movement Disorders*, vol. 36, pp. 803–814, 2021.