Statistical Modeling of Speech Spectral Coefficients in Patients with Parkinson's Disease

Ina Kodrasi, Hervé Bourlard

Idiap Research Institute, Speech and Audio Processing Group, Martigny, Switzerland
Email: {ina.kodrasi, herve.bourlard}@idiap.ch

Abstract

To automatically detect and monitor Parkinson's disease (PD) from speech, crafting features which robustly differentiate between speech of PD patients and healthy speakers is necessary. Since the voice of PD patients is typically breathy and semi-whispery and since their speech is characterized by significantly less pauses than healthy speech, it can be expected that PD speech spectral coefficients are less super-Gaussian than healthy speech spectral coefficients. In this paper we propose to use the distribution of speech spectral coefficients as a novel discriminatory feature between PD and healthy speech. Speech spectral magnitudes are modeled using the Weibull distribution, with the shape parameter controlling the super-Gaussianity of the complex spectral coefficients. Supported by empirical analysis on healthy and PD speech, it is shown that the shape parameter modeling PD spectral magnitudes is larger than the shape parameter modeling healthy spectral magnitudes, i.e., PD spectral coefficients are less super-Gaussian than healthy spectral coefficients. This result should be taken into account not only when discriminating between healthy and PD speech, but also when developing statistical signal processing techniques.

1 Introduction

Knowledge of the distribution of the speech spectral coefficients is crucial for many speech enhancement techniques [1–4], model-based voice activity detectors [5, 6], and several statistical signal processing algorithms [7]. Supported by empirical observations, e.g., in [8–10], it is widely accepted that the distribution of the complex spectral coefficients is super-Gaussian. This holds both locally, i.e., when observing the spectral coefficients in a single time-frequency bin [8, 9], as well as globally, i.e., when observing the distribution of the spectral coefficients in a single frequency bin [10]. Super-Gaussianity of the speech spectral coefficients arises due to pauses between phonemes and due to formant transitions in voiced sounds. Exploiting the super-Gaussianity of speech signals has proven to be beneficial for many speech enhancement techniques, such as e.g. single- or multi-channel dereverberation techniques [4, 11, 12]. To the best of our knowledge, the distribution of the speech spectral coefficients and the benefits of enhancement techniques exploiting super-Gaussianity have been demonstrated using utterances from speakers who do not manifest any speech disorder.

Speech disorders are common symptoms of several neurodegenerative diseases such as Parkinson's disease (PD). PD is among the most prevalent progressive neurodegenerative disorders, affecting nearly 1.5 % of the European population older than 60 [13]. Among other symptoms, the majority of PD patients develop hypokinetic dysarthria, which is a speech disorder characterized by imprecise articulation, abnormal speech rhythm, increased vocal tremor, and breathiness [14–16]. With the aim to assist the clinical diagnosis and treatment of PD patients, there has been a growing interest in the research community to develop discriminatory features which can be used for the automatic detection and monitoring of PD, e.g., in [17–26].

In [17–19] features such as vowel space area, vowel articulation index, consonant spectral trend, and consonant spectral moment were investigated to characterize the imprecise articulation in PD patients. In [20] the articulatory rate and pause time were investigated, where it was shown that PD patients make significantly less pauses between words and within polysyllabic words. Increased vocal tremor and breathiness were characterized using features such as jitter, shimmer, harmonics-to-noise ratio, or glottal-to-noise excitation ratio in [21, 22], whereas differential phonological posterior features were used in [26]. Although numerous contributions have been made in crafting robust features to discriminate between healthy and PD speech, the distribution of the spectral coefficients of healthy and PD speech has not been compared. Since PD patients make significantly less pauses than healthy speakers and since their speech is characterized by vocal tremor and breathiness [20, 21], it can expected that the distribution of PD speech spectral coefficients is less super-Gaussian than healthy speech.

The objective of this paper is to model and compare the global distribution of healthy and PD speech spectral coefficients as a function of the frequency and the frame size. To this end, the spectral magnitudes are modeled using the Weibull distribution [27], where the shape parameter controls the super-Gaussianity of the complex spectral coefficients [10]. Maximum likelihood (ML) estimates of the shape parameter are derived using a database of healthy and PD speech. It is shown that independently of the frequency and frame size, the shape parameter modeling the distribution of spectral magnitudes is larger for PD speech than for healthy speech, i.e., the spectral coefficients of PD speech are less super-Gaussian than healthy speech. The importance of this result is two-fold. First, the distribution of spectral coefficients can be used as an additional discriminatory feature in existing automatic PD classification techniques such as in [21, 26]. Second, this difference in the distribution of spectral coefficients should be taken into account when developing statistical signal processing techniques. As demonstrated by experimental results in Section 3.4, enhancement techniques promoting super-Gaussianity yield a lower performance for PD speech than for healthy speech.

2 Distribution Modeling

In this section, the Weibull distribution is presented and the ML estimation of the shape parameter is discussed. In the following, speech spectral coefficients are denoted by

$S_k(l)$, where k denotes the frequency bin index, l the frame index, and λ_k^2 the average speech power spectral density (PSD), i.e.,

$$\lambda_k^2 = \mathscr{E}\{|S(k)|^2\}, \tag{1}$$

with $\mathscr{E}$ the expected value operator.

2.1 Weibull distribution

Similarly to [10], the distribution of the spectral magnitudes $|S_k|$ in each frequency bin k is modeled using the Weibull distribution with probability density function [27]

$$p(|S_k|) = \frac{\beta_k}{\alpha_k}\left(\frac{|S_k|}{\alpha_k}\right)^{\beta_k - 1} e^{-\left(\frac{|S_k|}{\alpha_k}\right)^{\beta_k}}, \tag{2}$$

where β_k denotes the shape parameter and α_k denotes the scale parameter. The mean μ_k and the variance σ_k^2 of the Weibull distribution are given by

$$\mu_k = \alpha_k \Gamma\left(1 + \frac{1}{\beta_k}\right), \tag{3}$$

$$\sigma_k^2 = \alpha_k^2\left[\Gamma\left(1 + \frac{2}{\beta_k}\right) - \Gamma^2\left(1 + \frac{1}{\beta_k}\right)\right], \tag{4}$$

with $\Gamma(\cdot)$ being the gamma function [28]. Figure 1 depicts the probability density function of the Weibull distribution with $\sigma_k^2 = 1$ and different values of the shape parameter β_k. For $\beta_k = 2$, the Weibull distribution resembles the Rayleigh distribution, whereas for $\beta_k = 1$, the Weibull distribution resembles the exponential distribution. In addition, for $\beta_k < 2$, the Weibull distribution models the magnitude of super-Gaussian distributed complex spectral coefficients [10], with lower values of β_k corresponding to more super-Gaussian distributed complex spectral coefficients.

2.2 Shape parameter estimation

To model the spectral magnitudes using a Weibull distribution, the scale parameter α_k and the shape parameter β_k need to be estimated, cf. (2). Since the variance is given by $\sigma_k^2 = \mathscr{E}\{|S_k|^2\} - \mathscr{E}^2\{|S_k|\}$, the scale parameter α_k can be expressed in terms of the average PSD λ_k^2 and shape parameter β_k as (cf. (1), (3), and (4))

$$\alpha_k = \frac{\lambda_k}{\sqrt{\Gamma\left(1 + \frac{2}{\beta}\right)}}. \tag{5}$$

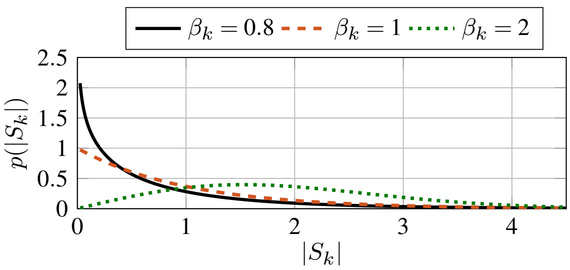

Figure 1: Probability density function of the Weibull distribution for $\sigma_k^2 = 1$ and different values of the shape parameter β_k.

To estimate the shape parameter β_k, an ML estimator is used in this paper. Given the speech spectral coefficients at frequency bin k, i.e., $S_k(1)$, $S_k(2)$, ..., $S_k(L)$, with L the total number of frames, the likelihood function of the shape parameter is given by

$$\mathscr{L}(\beta_k) = \prod_{l=1}^{L} \frac{\beta_k}{\alpha_k}\left(\frac{|S_k(l)|}{\alpha_k}\right)^{\beta_k - 1} e^{-\left(\frac{|S_k(l)|}{\alpha_k}\right)^{\beta_k}}. \tag{6}$$

An ML estimate of the shape parameter β_k can be obtained by minimizing the negative of the log-likelihood function, i.e., by solving the optimization problem

$$\begin{aligned}\min_{\beta_k} \quad & \Big[L\log\beta_k - L\beta_k \log\alpha_k \\ & + (\beta_k - 1)\sum_{l=1}^{L}\log|S_k(l)| - \sum_{l=1}^{L}\left(\frac{|S_k(l)|}{\alpha_k}\right)^{\beta_k}\Big], \end{aligned} \tag{7}$$

with α_k given by (5). Since no analytical solution to (7) exists, an iterative optimization technique should be used to find the ML estimate of the shape parameter β_k. In this paper, the one-dimensional quasi-Newton method is used.

3 Experimental Results

In this section, empirical analysis of the distribution of healthy and PD spectral magnitudes for different frame sizes are presented. In addition, experimental results are presented to demonstrate that successful enhancement techniques promoting super-Gaussianity yield a lower performance for PD speech than for healthy speech. For this purpose, we use the multi-channel linear prediction (MCLP)-based dereverberation technique from [4], whose success is based upon promoting super-Gaussianity of the output speech spectral coefficients.

Section 3.1 describes the used databases of healthy and PD speech. Section 3.2 discusses the methodology for estimating the shape parameter, whereas the obtained shape parameter values are presented in Section 3.3. Section 3.4 discusses the MCLP technique, algorithmic settings, performance measures, and presents dereverberation performance results for healthy and PD speech for several reverberation times and system configurations.

3.1 Databases

French recordings of 47 healthy speakers [29] and 39 PD patients are considered [30][1], with all speakers being French native speakers. The age of the healthy speakers ranges from 22 to 88 years old, with the average age being 52.8 years old. The age of the PD patients ranges from 45 to 78 years old, with the average age being 61.6 years old. The sampling frequency of the recordings is 44.1 kHz. The speakers are asked to read samples from a list of 54 pseudo-words, i.e., strings of characters resembling real words but having no meaning. The number of available recordings for each pseudo-word is different, however, the number of healthy and PD recordings for individual pseudo-words is the same. The minimum number of available recordings for a pseudo-word is 9 healthy and 9 PD recordings, whereas the maximum number of available recordings is 39 healthy and 39 PD recordings.

[1] Approval from Swissethics, number 2015-00028 – (15-258).

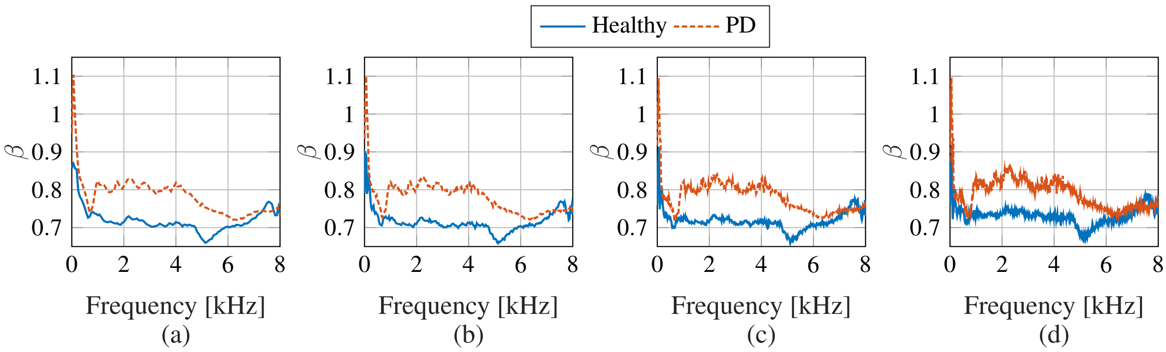

Figure 2: Frequency-dependent shape parameter estimated using all available recordings of healthy and PD speech for several frame sizes: (a) 256, (b) 512, (c) 1024, and (d) 2048.

The average number of available recordings for a pseudo-word is 27 healthy and 27 PD recordings. The total number of available recordings for the healthy and PD patients is 1465. The average length of the available recordings is 791 ms for the healthy speakers and 829 ms for the PD patients. Concatenating all recordings yields a 19.3 minutes long signal for the healthy speakers and a 20.2 minutes long signal for the PD patients. Manual voice activity detection has been performed for each recording to discard the silent (or noise-only) segments.

3.2 Processing

The signals are downsampled to 16 kHz and are processed using a weighted overlap-add short-time Fourier transform (STFT) framework. The considered frame sizes are 256, 512, 1024, and 2048, i.e., the considered frame durations range from 16 ms to 128 ms. The overlap between successive frames is 50 %. For each frequency bin k, the following processing is performed:

- Extract the spectral magnitudes $|S_k(1)|, \ldots, |S_k(L)|$.
- Compute the average speech PSD λ_k^2 and express the scale parameter α_k in terms of the shape parameter β_k as in (5).
- Determine the shape parameter β_k by solving the optimization problem in (7). In order to initialize the iterative optimization method, $\beta_k = 2$ has been used. However, it should be noted that the optimization procedure is not sensitive to the shape parameter initialization value.

3.3 Distribution of healthy and PD speech

Using all recordings for each group (i.e., the signals created by concatenating all pseudo-word recordings), the frequency-dependent shape parameter is estimated. Figure 2 depicts the obtained shape parameter values for healthy and PD speech for all considered frame sizes. It can be observed that as expected, the distribution of speech spectral magnitudes is closer to an exponential distribution than to a Rayleigh distribution, independently of the frame size and independently of whether healthy or PD speech is considered. In addition, it can be observed that independently of the considered frame size, the shape parameter modeling PD spectral magnitudes is typically larger than the shape parameter modeling healthy spectral magnitudes, i.e., PD speech is less super-Gaussian than healthy speech. For frequencies larger than approximately 7 kHz, the shape parameter modeling PD spectral magnitudes is similar or slightly lower than for healthy spectral magnitudes. This occurs due to the lack of speech in the higher frequencies, with the spectral coefficients mainly reflecting recording noise.

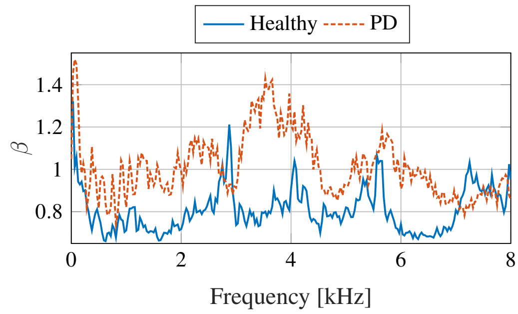

Figure 3: Frequency-dependent shape parameter estimated using a single recording for an exemplary pseudo-word from a healthy speaker and a PD patient (frame size: 512).

This difference in the shape parameter value for healthy and PD spectral magnitudes can be observed not only when considering all recordings from all speakers (as in Fig. 2), but also when considering a single pseudo-word from a single speaker. Figure 3 depicts the frequency-dependent shape parameter values for an exemplary pseudo-word from a healthy speaker and a PD patient. Both speakers are females, with the healthy speaker being 79 years old and the PD patient being 49 years old. The length of the healthy speaker recording is 697 ms, whereas the length of the PD patient recording is 778 ms. Similarly to before, it can be observed that the obtained shape parameter values for PD speech are typically larger than for healthy speech. Hence, it can be said that the difference in the super-Gaussianity of the spectral coefficients of healthy and PD speech is clearly observable also when considering short utterances in the order of hundreds of ms, making the super-Gaussianity of the speech signal a robust discriminatory feature between healthy and PD speech.

In summary, the presented results show that independently of the frequency, frame size, or length of the utterance, PD speech is less super-Gaussian than healthy speech. Furthermore, additional experiments (not presented here for privacy reasons) have shown that the presented results and conclusions generalize well to other databases, i.e., to languages other than French and to dif-

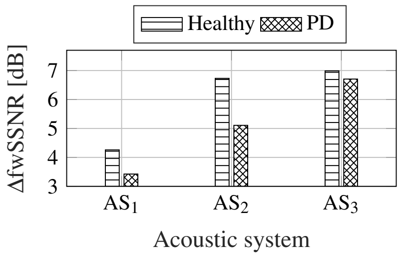

Figure 4: Dereverberation performance in terms of ΔfwSSNR when using the MCLP technique on healthy and PD reverberant speech.

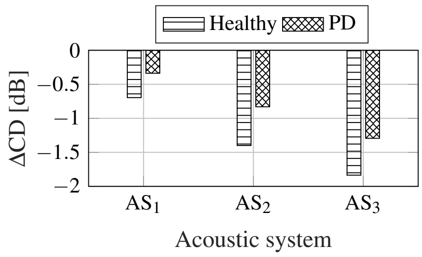

Figure 5: Dereverberation performance in terms of ΔCD when using the MCLP technique on healthy and PD reverberant speech.

ferent recording conditions.

3.4 Dereverberation of healthy and PD speech

To motivate the necessity of taking this different distribution of PD speech into account when developing statistical signal processing algorithms, in this section we use the multi-channel linear prediction (MCLP)-based speech dereverberation technique from [4] to dereverberate healthy and PD speech. The MCLP technique achieves speech dereverberation by exploiting the super-Gaussianity of speech spectral coefficients and it has been shown to result in a high performance improvement. Given a reverberant acoustic system, MCLP dereverberation filters are obtained in an iterative fashion by solving an l_p-norm minimization problem, with $0 \leq p \leq 1$ promoting super-Gaussianity of the enhanced speech spectral coefficients[2]. The results presented in the following are generated using $p = 0$, i.e., an l_0-norm is used. However, similar results are obtained also for other values of the parameter p.

We consider three reverberant acoustic systems with a single source and $M = 4$ microphones. The first acoustic system AS_1 consists of a linear microphone array with an inter-sensor distance of 8 cm [31], the second acoustic system AS_2 consists of a circular microphone array with a radius of 10 cm [32], and the third acoustic system AS_3 consists of a linear microphone array with an inter-sensor distance of 4 cm [33]. The reverberation times are $T_{60} \approx 0.61$ s for AS_1, $T_{60} \approx 0.73$ s for AS_2, and $T_{60} \approx 1.25$ s for AS_3. The speech components are generated by convolving the healthy and PD recordings described in Section 3.1 with measured room impulse responses for each acoustic system. For the results presented in the following, the signals are processed in the STFT domain with a frame size of 512.

To evaluate the performance, we use the improvement in frequency-weighted segmental signal-to-noise ratio (ΔfwSSNR) [34] and in cepstral distance (ΔCD) [35] between the enhanced output signal and the reverberant input signal. The first microphone is arbitrarily selected as the reference microphone. The reference signal used to compute the instrumental measures is the clean speech signal. It should be noted that a positive ΔfwSSNR and a negative ΔCD indicate a performance improvement.

Figures 4 and 5 depict the ΔfwSSNR and ΔCD values obtained using the MCLP technique for healthy and PD reverberant speech for all considered acoustic systems. As expected, it can be observed that MCLP yields a high performance improvement for all considered acoustic systems when dereverberating healthy speech. Furthermore, it can be observed that the performance improvement in terms of both performance measures and for all considered acoustic systems is higher for healthy speech than for PD speech. Specifically, using MCLP for PD speech yields a ΔfwSSNR up to 1.6 dB lower than for healthy speech. In addition, using MCLP for PD speech yields a ΔCD up to 0.6 dB higher than for healthy speech. The presented results confirm that state-of-the-art enhancement techniques which have proven successful in enhancing healthy speech by promoting super-Gaussianity of the spectral coefficients yield a lower performance when enhancing PD speech.

[2]For details on the MCLP technique, the reader is referred to [4].

4 Conclusion

In this paper, we have presented a model for the global distribution of PD speech spectral magnitudes and compared it to the distribution of healthy speech spectral magnitudes. Supported by empirical analysis, it has been shown that due to increased vocal tremor and breathiness and due to a decrease in the number of pauses, PD spectral coefficients are less super-Gaussian than healthy spectral coefficients. In addition, experimental results have shown that due to this difference in the distribution of healthy and PD speech, successful state-of-the-art speech enhancement techniques exploiting super-Gaussianity yield a lower performance when used on PD speech. Potential future steps are in two directions. The first direction consists in developing novel PD automatic detection and monitoring techniques exploiting only the super-Gaussianity of the spectral coefficients as a discriminatory feature between healthy and PD speech. In addition, current automatic PD detection and monitoring systems can be extended to include the super-Gaussianity of the spectral coefficients as an additional discriminatory feature. The second direction consists in using this knowledge of the distribution of PD speech spectral coefficients in statistical signal processing techniques targeting PD speech.

Acknowledgment

The authors would like to acknowledge the support of the Swiss NSF project no CRSII5_173711 "MoSpeeDi". They would also like to thank the project partners from University of Paris III: Sorbonne Nouvelle, Geneva University Hospitals, and University of Geneva.

References

[1] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[2] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, “Minimum mean-square error estimation of discrete Fourier coefficients with generalized Gamma priors,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 6, pp. 1741–1752, Aug. 2007.

[3] K. Kumatani, J. McDonough, B. Rauch, D. Klakow, P. N. Garner, and L. Weifeng, “Beamforming with a maximum negentropy criterion,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 5, pp. 994–1008, Jul. 2009.

[4] A. Jukić, T. Van Waterschoot, T. Gerkmann, and S. Doclo, “Multichannel linear prediction-based speech dereverberation with sparse priors,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1509–1520, Sep. 2015.

[5] J. Sohn, N. S. Kim, and W. Sung, “A statistical model-based voice activity detection,” *IEEE Signal Processing Letters*, vol. 6, no. 1, pp. 1–3, Jan. 1999.

[6] J.-H. Chang, N. S. Kim, and S. K. Mitra, “Voice activity detection based on multiple statistical models,” *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 1965–1976, Jun. 2006.

[7] P. Bofill and M. Zibulevsky, “Blind separation of more sources than mixtures using sparsity of their short-time Fourier transform,” in *Proc. International Workshop on Independent Component Analysis and Blind Signal Separation*, Helsinki, Finland, Jun. 2000, pp. 87–92.

[8] R. Martin, “Speech enhancement using MMSE short time spectral estimation with Gamma distributed speech priors,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, USA, May 2002, pp. 253–256.

[9] T. Gerkmann and R. Martin, “Empirical distributions of DFT-domain speech coefficients based on estimated speech variances,” in *Proc. International Workshop on Acoustic, Echo, and Noise Control*, Tel Aviv, Israel, Sep. 2010.

[10] I. Tashev and A. Acero, “Statistical modeling of the speech signal,” in *Proc. International Workshop on Acoustic, Echo, and Noise Control*, Tel Aviv, Israel, Sep. 2010.

[11] I. Kodrasi and S. Doclo, “Sparsity-promoting acoustic multichannel equalization techniques,” *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 25, no. 7, pp. 1512–1525, Jul. 2017.

[12] H. Kameoka, T. Nakatani, and T. T. Yoshioka, “Robust speech dereverberation based on non-negativity and sparse nature of speech spectrograms,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taipei, Taiwan, Apr. 2009, pp. 45–48.

[13] S. von Campenhausen, B. Bornschein, R. Wick, K. Bötzel, C. Sampaio, W. Poewe, W. Oertel, U. Siebert, K. Berger, and R. Dodel, “Prevalence and incidence of Parkinson’s disease in Europe,” *European Neuropsychopharmacology*, vol. 15, no. 4, pp. 473–490, Aug. 2005.

[14] G. J. Canter, “Speech characteristics of patients with Parkinson’s disease: I. Intensity, pitch, and duration,” *Journal of Speech and Hearing Disorders*, vol. 28, pp. 221–229, Aug. 1963.

[15] ——, “Speech characteristics of patients with Parkinson’s disease: III. Articulation, diadochokinesis, and over-all speech adequacy,” *Journal of Speech and Hearing Disorders*, vol. 30, pp. 217–224, Aug. 1965.

[16] C. Stewart, L. Winfield, A. Hunt, S. B. Bressman, S. Fahn, A. Blitzer, and M. F. Brin, “Speech dysfunction in early Parkinson’s disease,” *Movement Disorders*, vol. 10, no. 5, pp. 562–565, Sep. 1995.

[17] S. Skodda, W. Visser, and U. Schlegel, “Vowel articulation in Parkinson’s disease,” *Journal of Voice*, vol. 25, no. 4, pp. 467–472, Jul. 2011.

[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: Effect of speaking task,” *Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 2171–2181, Sep. 2013.

[19] M. Novotný, J. Rusz, R. Čmejla, and E. Růžička, “Automatic evaluation of articulatory disorders in Parkinson’s disease,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 9, pp. 1366–1378, Sep. 2014.

[20] S. Skodda and U. Schlegel, “Speech rate and rhythm in Parkinson’s disease,” *Movement Disorders*, vol. 23, no. 7, pp. 985–992, May 2008.

[21] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, “Novel speech signal processing algorithms for high-accuracy classification of parkinson’s disease,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, May 2012.

[22] D. Sztahó, G. Kiss, and K. Vicsi, “Estimating the severity of Parkinson’s disease from speech using linear regression and database partitioning,” in *Proc. 16th Annual Conference of the International Speech Communication Association*, Dresden, Germany, Sep. 2015, pp. 498–502.

[23] A. Zlotnik, J. M. Montero-Martínez, R. San-Segundo-Hernández, and A. Gallardo-Antolín, “Random forest-based prediction of Parkinson’s disease progression using acoustic, ASR and intelligibility features,” in *Proc. 16th Annual Conference of the International Speech Communication Association*, Dresden, Germany, Sep. 2015, pp. 503–507.

[24] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. F. V. Bonilla, S. Skodda, J. Rusz, and E. Nöth, “Voiced/unvoiced transitions in speech as a potential bio-marker to detect Parkinson’s disease,” in *Proc. 16th Annual Conference of the International Speech Communication Association*, Dresden, Germany, Sep. 2015, pp. 95–99.

[25] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. F. V. Bonilla, K. Daqrouq, S. Skodda, J. Rusz, and E. Nöth, “Automatic detection of Parkinson’s disease in running speech spoken in three different languages,” *Journal of the Acoustical Society of America*, vol. 139, no. 1, pp. 481–500, Jan. 2016.

[26] M. Cernak, J. R. Orozco-Arroyave, F. Rudzicz, H. Christensen, J. C. Vásquez-Correa, and E. Nöth, “Characterisation of voice quality of Parkinson’s disease using differential phonological posterior features,” *Computer Speech and Language*, vol. 46, pp. 196–208, Nov. 2017.

[27] W. Weibull, “A statistical distribution function of wide applicability,” *Journal of Applied Mechanics*, vol. 18, pp. 293–297, Sep. 1951.

[28] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 6th ed. San Diego, CA, USA: Academic Press, 2000.

[29] C. Fougeron, V. Delvaux, L. Ménard, and M. Laganaro, “The MonPaGe_HA Database for the documentation of spoken French throughout adulthood,” in *Proc. 11th International Conference on Language Resources and Evaluation*, Miyazaki, Japan, May 2018.

[30] M. Fournet, S. C. Chiuvé, S. Momjian, P. R. Burkhard, and M. Laganaro, “Modulation of voice and speech parameters after subthalamic nucleus deep brain stimulation in Parkinson’s disease,” *Brain Stimulation*, vol. 10, no. 2, p. 470, Mar. 2017.

[31] E. Hadad, F. Heese, P. Vary, and S. Gannot, “Multichannel audio database in various acoustic environments,” in *Proc. International Workshop on Acoustic Echo and Noise Control*, Antibes, France, Sep. 2014, pp. 313–317.

[32] K. Kinoshita, M. Delcroix, S. Gannot, E. A. P. Habets, R. Haeb-Umbach, W. Kellermann, V. Leutnant, R. Maas, T. Nakatani, B. Raj, A. Sehr, and T. Yoshioka, “A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research,” *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, Jan. 2016.

[33] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, “The ACE challenge - Corpus description and performance evaluation,” in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, USA, Oct. 2015.

[34] Y. Hu and P. C. Loizou, “Evaluation of objective quality measures for speech enhancement,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 229–238, Jan. 2008.

[35] S. Quackenbush, T. Barnwell, and M. Clements, *Objective measures of speech quality*. New Jersey, USA: Prentice-Hall, 1988.