

Analyzing Vocal Tremor as an Acoustic Biomarker for Parkinson's Disease Using Sustained Vowels

Luiz Alberto Silva, Edson Cataldo and Andrea Gomes de Oliveira Aguiar

Abstract—This study investigates the contribution of vocal tremor-related acoustic features to the automatic detection of Parkinson's disease (PD) from sustained vowel phonations. While most voice-based approaches rely on conventional perturbation and spectral descriptors, tremor-related parameters remain underexplored despite their physiological relevance to neuromotor impairment. A comprehensive set of acoustic features was extracted, including jitter, shimmer, noise-related measures, nonlinear descriptors, and tremor features derived from low-frequency modulations of fundamental frequency and amplitude. Experiments were conducted on multiple datasets with distinct recording conditions, and classification was performed using a multilayer perceptron under subject-independent cross-validation. Results indicate that conventional features remain highly discriminative, but tremor-related descriptors consistently contribute complementary information and improve classification performance, achieving accuracy up to 98% when combined with traditional features. Although not sufficient alone, tremor features enhance both performance and physiological interpretability. These findings support the inclusion of vocal tremor descriptors in PD detection systems and highlight their potential as interpretable acoustic biomarkers across diverse recording conditions.

Keywords—Vocal tremor; acoustic features; voice analysis; Parkinson's disease.

I. INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by the degeneration of dopaminergic neurons, leading to motor symptoms such as bradykinesia, rigidity, tremor, and postural instability, as well as non-motor impairments [6]. Given its progressive nature, the development of objective methods for early detection and monitoring is essential.

Speech is frequently affected in PD, often in early stages, making acoustic analysis a promising non-invasive approach. Typical alterations include reduced loudness, monopitch, and articulatory imprecision, reflecting impaired neuromotor control. While many studies focus on perturbation, spectral, and nonlinear features, the role of vocal tremor remains comparatively underexplored.

Vocal tremor manifests as low-frequency (< 12 Hz) quasi-periodic modulations in fundamental frequency and amplitude, arising from oscillatory neuromotor activity [2]. Unlike perturbation measures, which capture short-term irregularities,

tremor reflects slower rhythmic dynamics and may provide complementary information for PD detection.

This work investigates the contribution of tremor-related acoustic parameters for automatic PD detection using sustained vowel phonations. Experiments are conducted on three datasets: mPower [7], PC-GITA [5], and an Italian voice dataset [8], enabling evaluation under different recording conditions and populations.

The feature set includes traditional acoustic measures, nonlinear descriptors, and dedicated tremor parameters derived from frequency and amplitude contours. A multilayer perceptron classifier is used to assess their discriminative relevance, emphasizing feature contribution over model complexity.

The main contributions are:

- Systematic evaluation of tremor-related acoustic parameters for PD detection;
- Comparison with traditional perturbation and nonlinear features;
- Cross-dataset analysis to assess robustness and generalization;
- Identification of a compact and informative feature subset;
- An interpretable neural framework to assess feature relevance.

II. ACOUSTIC FEATURES

This study employs a set of acoustic descriptors extracted from sustained vowel recordings, with emphasis on vocal tremor. A total of 20 tremor-related parameters are considered, comprising 10 features derived from low-frequency modulations of the fundamental frequency and 10 from amplitude modulations. These descriptors capture tremor rate, magnitude, and cyclic behavior associated with neuromotor activity [3].

Conventional acoustic features are also included, such as perturbation measures (jitter and shimmer) and noise-related descriptors (e.g., HNR). In addition, nonlinear measures, including pitch period entropy (PPE), are used to characterize irregular temporal patterns in the voice signal [4].

The combination of tremor-related, conventional, and nonlinear features enables a comprehensive representation of both short-term phonatory perturbations and slow oscillatory dynamics relevant to PD detection.

A. Vocal tremor parameters

A set of 20 tremor-related descriptors is considered, divided into frequency-based features extracted from the fundamental frequency contour $F_0(t)$ and amplitude-based features derived

Luiz Alberto Silva, Graduate Program in Electrical and Telecommunications Engineering, UFF, Niteroi-RJ, e-mail: luizalbertosilva@id.uff.br; Edson Cataldo, Graduate Program in Electrical and Telecommunications Engineering, UFF, Niteroi-RJ, e-mail: ecataldo@id.uff.br; Andrea Gomes de Oliveira Aguiar, Department of General and Specialized Surgery, UFF, Niteroi-RJ, e-mail: ag_oliveira@id.uff.br. This work was partially supported by CNPq (309803/2023-3 and 178711/2025-0).

from the signal envelope [2], [3]. These parameters characterize modulation rate, magnitude, cyclicity, and spectral properties of tremor.

Table I summarizes the descriptors [16] used in this work.

TABLE I

TREMOR-RELATED ACOUSTIC PARAMETERS CONSIDERED IN THIS STUDY.

Frequency-based tremor descriptors	Abbreviation
Frequency Tremor Intensity Index	FTRI
Pitch Frequency Tremor Rate	PFTR
Frequency Tremor Frequency	FTrF
Frequency Contour Magnitude	FCoM
Frequency Tremor Cyclicity	FTrC
Frequency Modulation Count	FMoN
Frequency Tremor Power	FTrP
Frequency Tremor Cyclicity Index Peak	FTrCIP
Frequency Tremor Power Spectrum	FTrPS
Frequency Contour Harmonics-to-Noise Ratio	FCoHNR
Amplitude-based tremor descriptors	Abbreviation
Amplitude Tremor Rate	AFTR
Pitch Amplitude Tremor Rate	PATR
Amplitude Tremor Intensity Index	ATRI
Amplitude Contour Magnitude	ACoM
Amplitude Tremor Cyclicity	ATrC
Amplitude Modulation Count	AMoN
Amplitude Tremor Power Index	ATrP
Amplitude Tremor Cyclicity Index Peak	ATrCIP
Amplitude Tremor Power Spectrum	ATrPS
Amplitude Contour Harmonics-to-Noise Ratio	ACoHNR

III. MATERIALS AND METHODS

This section describes the signal processing and classification pipeline for automatic PD detection from sustained vowel phonations. The methodology comprises four stages: (i) signal preprocessing, (ii) extraction of fundamental frequency and amplitude contours, (iii) computation of acoustic features, including tremor descriptors, and (iv) classification using a neural network model.

Special emphasis is placed on tremor-related features derived from low-frequency modulations of pitch and amplitude. To ensure their reliability, preprocessing and pitch estimation procedures are designed to preserve slow oscillatory components.

The resulting feature vectors are used as input to a multilayer perceptron (MLP) classifier, evaluated under a cross-validation protocol. Detailed descriptions of each stage are provided in the following subsections.

A. Datasets

This study uses three publicly available datasets containing sustained vowel phonations from individuals with PD and healthy controls: mPower [7], the Italian Parkinson's Voice dataset [8], and PC-GITA [5]. The use of multiple corpora enables evaluation under diverse recording conditions, ranging from real-world smartphone data to controlled clinical environments, supporting robustness and generalization analysis.

The mPower dataset consists of large-scale smartphone recordings collected in unconstrained conditions (44.1 kHz), capturing natural variability in voice production. The Italian dataset provides high-quality recordings acquired in controlled environments (16 kHz), while PC-GITA includes clinical

recordings from PD and HC subjects (44.1 kHz), increasing speaker and condition diversity.

Only sustained vowel phonations are considered, as they reduce linguistic variability and facilitate the analysis of phonatory stability and low-frequency modulations associated with vocal tremor [15].

A summary of the datasets is provided in Table II.

TABLE II

SUMMARY OF THE DATASETS USED IN THIS STUDY. PD: PARKINSON'S DISEASE; HC: HEALTHY CONTROLS. EACH PARTICIPANT CONTRIBUTED ONE SUSTAINED VOWEL RECORDING.

Dataset	PD subjects	HC subjects	Recordings
mPower	1779	2101	3880
Italian Parkinson's Voice	22	28	50
PC-GITA	50	50	100

A subject-independent cross-validation protocol is adopted, ensuring that all recordings from a given speaker are assigned exclusively to either training or test sets, preventing subject leakage and providing a realistic evaluation of generalization.

B. Signal preprocessing

Each audio recording was processed to extract a quasi-stationary segment suitable for tremor analysis. Instead of using energy-based silence detection, a fixed-duration segment was selected to ensure consistent processing and avoid segmentation errors.

For each signal, a segment of approximately 1.5 seconds was extracted starting at 20% of the total duration. This strategy avoids onset and offset effects, which are associated with non-stationary phonatory behavior [1], [9], and increases the likelihood of capturing a stable phonatory regime.

The selected duration is sufficient to capture multiple cycles of low-frequency tremor oscillations (typically 3–12 Hz), enabling reliable estimation of tremor-related descriptors while remaining adequate for conventional acoustic measures.

C. Tremor feature extraction

The fundamental frequency contour (F_0) was first estimated after segment selection, as it constitutes the basis for frequency tremor analysis. Pitch was extracted using the autocorrelation method implemented in Praat via Parselmouth, with a search range of 60–350 Hz. The resulting contour was uniformly sampled, and unreliable frames were discarded to avoid artifacts. To isolate tremor-related oscillations, the contour was detrended by removing a linear component and subsequently normalized by its mean, producing a dimensionless representation suitable for cross-speaker comparison.

Tremor descriptors were then computed from pitch and amplitude contours to characterize low-frequency oscillations associated with neuromotor control, including their rate, intensity, and regularity. Unlike classical perturbation measures (e.g., jitter and shimmer), which capture cycle-to-cycle irregularities, tremor descriptors reflect slow, quasi-periodic modulations (typically 3–12 Hz) and provide complementary information related to neuromotor activity.

1) *Frequency tremor*: Frequency-based descriptors were derived from the normalized F_0 contour using power spectral density (PSD) estimated via Welch's method. Features were computed within the tremor band (3–12 Hz), capturing dominant oscillation frequency and modulation strength.

2) *Amplitude tremor*: Amplitude-based descriptors were extracted from the signal envelope obtained via the Hilbert transform. The envelope was centered and bandpass filtered (3–12 Hz) to isolate tremor-related oscillations, followed by PSD estimation using Welch's method. These features characterize slow amplitude modulations and complement frequency-based tremor descriptors.

D. Summary of the features considered

The acoustic features are grouped into three categories: tremor-related, nonlinear, and traditional parameters, as summarized in Table III.

TABLE III
EXTRACTED ACOUSTIC FEATURES GROUPED BY CATEGORY.

Category	Features
Tremor (frequency)	FTRI, PFTR, FTrF, FCoM, FTrC, FMoN, FTrP, FTrCIP, FTrPS, FCoHNR
Tremor (amplitude)	ATRI, PATR, AFTR, ACoM, ATrC, AMoN, ATrP, ATrCIP, ATrPS, ACoHNR
Nonlinear	PPE, RPDE, DFA, Spread1, Spread2
Traditional	F_0 , F_{hi} , F_{lo} , Jitter, Jitter(Abs), Jitter RAP, Jitter PPQ, Jitter DDP, Shimmer, Shimmer(dB), Shimmer APQ3, Shimmer APQ5, Shimmer APQ11, Shimmer DDA, HNR, NHR

E. Feature selection

Feature selection was applied within each training fold to avoid information leakage and reduce redundancy. Feature relevance was estimated using a gradient-based criterion, where the magnitude of gradients reflects the sensitivity of the loss to each input feature.

Based on this ranking, subsets of top-10, top-15, and top-20 features were evaluated to analyze the trade-off between dimensionality and performance. This procedure also improves generalization by removing redundant or weakly informative descriptors.

F. Neural network classifier

Classification was performed using a multilayer perceptron (MLP) implemented in PyTorch. The network consists of an input layer, a single hidden layer with eight neurons (with batch normalization, ReLU, and dropout), and an output layer with two neurons (HC and PD), followed by softmax.

The model was trained using the Adam optimizer and weighted cross-entropy loss to handle class imbalance. Early stopping based on validation loss was applied to prevent overfitting. The final decision threshold was determined using the Equal Error Rate (EER) criterion on validation folds.

The architecture was intentionally kept simple to emphasize the contribution of the acoustic features, particularly tremor-related descriptors, rather than model complexity.

G. Evaluation metrics

Performance was evaluated using standard binary classification metrics, including accuracy, precision, recall, and AUC-ROC.

AUC-ROC was used as a threshold-independent measure of discriminative performance, particularly under class imbalance.

All metrics were computed using 10-fold cross-validation, and results are reported as mean and standard deviation across folds.

IV. RESULTS AND DISCUSSION

This section presents classification results on the Italian Parkinson's Voice, mPower, and PC-GITA datasets, focusing on the contribution of tremor-related features to PD detection. Performance is evaluated across different feature configurations, including tremor-only, traditional acoustic features, and reduced subsets obtained via feature selection, enabling comparison of their discriminative relevance under different conditions. The analysis also examines cross-validation stability and the relative contribution of each feature group, with emphasis on the complementary role of tremor descriptors.

The neural network was trained using mini-batch gradient descent with dropout regularization, and performance was monitored using validation loss and standard metrics (accuracy, precision, recall, and AUC-ROC). Hyperparameters were selected empirically to balance performance and simplicity, ensuring that results primarily reflect the discriminative power of the acoustic features rather than model complexity. The final configuration, summarized in Table IV, was kept fixed across all experiments.

TABLE IV
HYPERPARAMETERS USED FOR TRAINING THE NEURAL NETWORK CLASSIFIER.

Hyperparameter	Value
Model type	MLP
Input features	Selected subset
Hidden layers	1
Hidden neurons	8
Activation function	ReLU
Batch normalization	Yes
Dropout rate	0.3
Output neurons	2
Output activation	Softmax
Loss function	Weighted cross-entropy
Optimizer	Adam
Learning rate	0.003
Batch size	32
Maximum epochs	100
Early stopping	Yes
Validation	10-fold CV

A. Classification performance across datasets

Classification was evaluated using tremor-only, traditional, combined, and reduced feature sets (Top-20, Top-15, Top-10), with results reported as mean $\pm$ standard deviation over 10-fold cross-validation (Tables V–VII).

Across datasets, combining tremor and traditional features consistently improves or maintains performance compared to

TABLE V

CLASSIFICATION PERFORMANCE (MEAN $\pm$ STANDARD DEVIATION) ACROSS ITALIAN PARKINSON'S VOICE AND SPEECH DATASET USING SUBJECT-INDEPENDENT 10-FOLD CROSS-VALIDATION. BEST RESULTS DATASET ARE HIGHLIGHTED IN BOLD.

Feature set	Accuracy	AUC-ROC	Precision	Recall
Traditional	0.88 $\pm$ 0.05	0.93 $\pm$ 0.04	0.88 $\pm$ 0.05	0.88 $\pm$ 0.05
Tremor	0.86 $\pm$ 0.06	0.95 $\pm$ 0.03	0.86 $\pm$ 0.06	0.87 $\pm$ 0.06
Traditional + Tremor	0.98 $\pm$ 0.06	0.97 $\pm$ 0.11	0.98 $\pm$ 0.05	0.98 $\pm$ 0.05
Traditional + Nonlinear	0.96 $\pm$ 0.08	0.93 $\pm$ 0.14	0.97 $\pm$ 0.07	0.97 $\pm$ 0.07
All features	0.96 $\pm$ 0.12	0.98 $\pm$ 0.05	0.96 $\pm$ 0.13	0.96 $\pm$ 0.13
Top 20	0.94 $\pm$ 0.03	0.98 $\pm$ 0.02	0.94 $\pm$ 0.03	0.94 $\pm$ 0.03
Top 15	0.92 $\pm$ 0.03	0.98 $\pm$ 0.02	0.91 $\pm$ 0.03	0.92 $\pm$ 0.03
Top 10	0.93 $\pm$ 0.05	0.97 $\pm$ 0.03	0.93 $\pm$ 0.06	0.93 $\pm$ 0.05

TABLE VI

CLASSIFICATION PERFORMANCE (MEAN $\pm$ STANDARD DEVIATION) ACROSS MPOWER DATASET USING SUBJECT-INDEPENDENT 10-FOLD CROSS-VALIDATION. BEST RESULTS DATASET ARE HIGHLIGHTED IN BOLD.

Feature set	Accuracy	AUC-ROC	Precision	Recall
Traditional	0.76 $\pm$ 0.04	0.82 $\pm$ 0.01	0.75 $\pm$ 0.04	0.76 $\pm$ 0.03
Tremor	0.65 $\pm$ 0.03	0.69 $\pm$ 0.02	0.64 $\pm$ 0.02	0.65 $\pm$ 0.02
Traditional + Tremor	0.72 $\pm$ 0.02	0.80 $\pm$ 0.02	0.72 $\pm$ 0.02	0.72 $\pm$ 0.02
Traditional + Nonlinear	0.75 $\pm$ 0.02	0.82 $\pm$ 0.01	0.75 $\pm$ 0.02	0.75 $\pm$ 0.02
All features	0.76 $\pm$ 0.01	0.83 $\pm$ 0.02	0.76 $\pm$ 0.01	0.76 $\pm$ 0.02
Top 20	0.81 $\pm$ 0.03	0.85 $\pm$ 0.01	0.81 $\pm$ 0.03	0.81 $\pm$ 0.03
Top 15	0.77 $\pm$ 0.03	0.85 $\pm$ 0.02	0.76 $\pm$ 0.02	0.77 $\pm$ 0.04
Top 10	0.75 $\pm$ 0.02	0.83 $\pm$ 0.01	0.75 $\pm$ 0.01	0.75 $\pm$ 0.01

using individual feature groups. On the Italian dataset, the combined set achieved the best performance (Accuracy = 0.98), indicating strong complementarity. On mPower and PC-GITA, reduced feature sets (particularly Top-20) achieved the best or comparable results, suggesting that compact representations are sufficient for robust classification.

Overall, results indicate that tremor-related descriptors contribute complementary information and that competitive performance can be achieved with reduced feature subsets, enabling lower-dimensional and more efficient models.

B. Comparison with previous studies

Tables VIII and IX compare the proposed method with prior work using the mPower and Italian datasets.

The proposed method outperforms previous approaches on mPower and achieves competitive performance on the Italian dataset. Unlike deep learning models based on high-dimensional spectral inputs, the proposed approach relies on interpretable acoustic features, explicitly incorporating tremor-

TABLE VII

CLASSIFICATION PERFORMANCE (MEAN $\pm$ STANDARD DEVIATION) ACROSS PC-GITA DATASET USING SUBJECT-INDEPENDENT 10-FOLD CROSS-VALIDATION. BEST RESULTS DATASET ARE HIGHLIGHTED IN BOLD.

Feature set	Accuracy	AUC-ROC	Precision	Recall
Traditional	0.59 $\pm$ 0.11	0.72 $\pm$ 0.09	0.59 $\pm$ 0.12	0.59 $\pm$ 0.10
Tremor	0.71 $\pm$ 0.16	0.70 $\pm$ 0.11	0.72 $\pm$ 0.16	0.71 $\pm$ 0.16
Traditional + Tremor	0.76 $\pm$ 0.10	0.80 $\pm$ 0.15	0.76 $\pm$ 0.14	0.76 $\pm$ 0.12
Traditional + Nonlinear	0.76 $\pm$ 0.14	0.78 $\pm$ 0.09	0.77 $\pm$ 0.15	0.77 $\pm$ 0.15
All features	0.76 $\pm$ 0.02	0.83 $\pm$ 0.05	0.76 $\pm$ 0.01	0.77 $\pm$ 0.01
Top 20	0.77 $\pm$ 0.13	0.83 $\pm$ 0.14	0.78 $\pm$ 0.13	0.77 $\pm$ 0.13
Top 15	0.75 $\pm$ 0.02	0.81 $\pm$ 0.03	0.75 $\pm$ 0.02	0.76 $\pm$ 0.02
Top 10	0.68 $\pm$ 0.02	0.74 $\pm$ 0.01	0.69 $\pm$ 0.02	0.68 $\pm$ 0.02

TABLE VIII

COMPARISON WITH PREVIOUS STUDIES USING THE MPOWER DATASET AND SPEECH FEATURES.

Reference	Method	AUC-ROC	Accuracy
Rahmatallah et al., 2025 [14]	Log. Reg.	0.70	-
Tougui et al., 2022 [11]	Grad. Boost.	-	0.66
Carrón et al., 2021 [10]	Grad. Boost.	0.76	0.71
This work	MLP	0.85	0.81

related descriptors. This provides a favorable trade-off between performance and interpretability while highlighting the relevance of tremor information for PD detection.

V. ROBUST SUBSET OF ACOUSTIC PARAMETERS ACROSS DATASETS

Feature relevance was first analyzed using a gradient-based sensitivity approach, where the importance of each descriptor is quantified by the average magnitude of the partial derivatives of the network output with respect to the input features. Across all datasets, traditional acoustic parameters (e.g., jitter, shimmer, HNR) and nonlinear descriptors (e.g., PPE, DFA) consistently ranked among the most informative features.

Tremor-related descriptors were also frequently selected, confirming that slow oscillatory modulations in pitch and amplitude provide complementary information for PD detection. In particular, features such as FCoM, ATrI, and ATrPS were recurrently identified among the most relevant. While PPE showed high importance in the mPower and PC-GITA datasets, its relevance was reduced in the Italian dataset, likely due to differences in recording conditions and pitch estimation characteristics.

In addition, feature selection experiments showed that competitive performance can be achieved using reduced feature sets (10–20 descriptors), with minimal loss compared to the full feature space. Tremor-related features were consistently retained, indicating that a compact combination of traditional

TABLE IX

COMPARISON WITH PREVIOUS STUDIES USING SUSTAINED VOWEL PHONATIONS FROM THE ITALIAN DATASET.

Reference	Method	AUC-ROC	Accuracy
Mazinani et al., 2024 [12]	MLP	0.98	0.93
Simone et al., 2025 [13]	SVM	-	0.94
This work	MLP	0.97	0.98

and tremor descriptors is sufficient to capture the most discriminative information while reducing computational cost.

Motivated by these observations, a compact and interpretable subset of acoustic features was defined to assess cross-dataset generalization. The selected descriptors include tremor-related, traditional, and nonlinear features:

- Tremor: FCoM, FCoHNR, ATrI, ATrPS, AMoN, ACoM, ACoHNR;
- Traditional: F_{hi} , Jitter (Abs, %, DDP, RAP, PPQ), Shimmer (APQ11, DDA), NHR, HNR;
- Nonlinear: PPE, DFA, Spread2.

Table X summarizes the classification performance obtained using this subset across datasets. The results show strong performance on the Italian dataset ($AUC \approx 0.95$) and consistent results on mPower and PC-GITA, indicating robustness under varying recording conditions.

TABLE X

PERFORMANCE USING THE SELECTED FEATURE SUBSET ACROSS DATASETS.

Dataset	Accuracy	AUC-ROC
mPower	0.75 ± 0.02	0.83 ± 0.01
Italian	0.94 ± 0.13	0.95 ± 0.11
PC-GITA	0.73 ± 0.15	0.81 ± 0.12

Overall, these results indicate that a reduced and interpretable feature set—guided by feature importance and dimensionality reduction—can achieve robust performance across datasets. The consistent presence of tremor-related descriptors further supports their role as complementary and physiologically meaningful biomarkers for PD detection.

VI. CONCLUSIONS AND FUTURE WORK

This study investigated the contribution of vocal tremor-related acoustic features for Parkinson’s disease (PD) detection from sustained vowel phonations. The results show that conventional acoustic parameters remain highly discriminative, while tremor-related descriptors provide complementary information that improves classification performance and enhances physiological interpretability.

Feature importance and selection analyses indicate that perturbation measures (e.g., jitter and shimmer) and tremor descriptors capture distinct temporal aspects of phonatory variability, jointly contributing to robust discrimination. Moreover, competitive performance can be achieved using compact feature subsets, supporting the development of efficient and interpretable models for voice-based PD screening.

Some limitations should be noted. The datasets differ in size and recording conditions, which may affect performance

variability. In addition, the analysis was restricted to sustained vowels, whereas connected speech may reveal additional neuromotor characteristics. The use of a simple MLP, while intentional to emphasize feature relevance, may also limit performance compared to more complex models. Furthermore, broader validation on larger and more diverse datasets is required to confirm generalization.

Future work will focus on extending the analysis to larger and more heterogeneous datasets, incorporating connected speech tasks, improving tremor modeling techniques, and exploring advanced learning and interpretability methods to further enhance PD detection systems.

REFERENCES

- [1] I. R. Titze, *Principles of Voice Production*, Prentice Hall, Englewood Cliffs, NJ, USA, 1994.
- [2] G. Deuschl, P. Bain, and M. Brin, “Consensus statement of the Movement Disorder Society on tremor,” *Movement Disorders*, vol. 13, suppl. 3, pp. 2–23, 1998.
- [3] R. A. Lester and J. Barkmeier-Kraemer, “Speech and voice characteristics of persons with essential vocal tremor,” *Journal of Voice*, vol. 21, no. 4, pp. 423–436, 2007.
- [4] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, “Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection,” *BioMedical Engineering OnLine*, vol. 8, no. 23, 2009.
- [5] J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, and E. Nöth, “A new database for Parkinson’s disease speech analysis,” in *Proc. International Conference on Text, Speech and Dialogue (TSD)*, Brno, Czech Republic, pp. 400–407, 2014.
- [6] L. V. Kalia and A. E. Lang, “Parkinson’s disease,” *The Lancet*, vol. 386, no. 9996, pp. 896–912, 2015.
- [7] B. M. Bot, C. Suver, E. C. Neto, M. Kellen, A. Klein, C. Bare, M. Doerr, A. Pratap, J. Wilbanks, E. R. Dorsey, S. H. Friend, and A. D. Trister, “The mPower study, Parkinson disease mobile data collected using ResearchKit,” *Scientific Data*, vol. 3, no. 1, 2016.
- [8] G. Dimmauro, V. Di Nicola, V. Bevilacqua, P. Caivano, and F. Girardi, “Assessment of speech intelligibility in Parkinson’s disease using a speech-to-text system,” *IEEE Access*, vol. 5, pp. 22199–22208, 2017.
- [9] R. R. Patel, S. N. Awan, J. Barkmeier-Kraemer, M. Courey, D. Deliyski, T. Eadie, D. Paul, J. G. Švec, and R. Hillman, “Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function,” *American Journal of Speech-Language Pathology*, vol. 27, no. 3, pp. 887–905, 2018.
- [10] J. Carrón, Y. Campos-Roca, M. Madruga, and C. J. Pérez, “A mobile-assisted voice condition analysis system for Parkinson’s disease: assessment of usability conditions,” *BioMedical Engineering OnLine*, vol. 20, no. 1, pp. 1–23, 2021.
- [11] I. Tougui, A. Jilbab, and J. El Mhamdi, “Machine Learning Smart System for Parkinson Disease Classification Using the Voice as a Biomarker,” *Healthcare Informatics Research*, vol. 28, no. 3, pp. 210–221, 2022.
- [12] A. Mazinani, D. P. Pau, L. Davoli, and G. Ferrari, “Deep Neural Quantization for Speech Detection of Parkinson Disease,” *2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI)*, Milano, Italy, pp. 178–183, 2024.
- [13] L. Simone, M. G. Camporeale, V. M. Rubino, V. Gervasi, and G. Dimmauro, “Interpretable Early Detection of Parkinson’s Disease through Speech Analysis,” *Artificial Intelligence in Medicine*, 2025.
- [14] Y. Rahmatallah, A. S. Kemp, A. Iyer, et al., “Pre-trained convolutional neural networks identify Parkinson’s disease from spectrogram images of voice samples,” *Scientific Reports*, vol. 15, 2025.
- [15] B. R. Burk and C. R. Watts, “The effect of Parkinson disease tremor phenotype on cepstral peak prominence and transglottal airflow in vowels and speech,” *Journal of Voice*, vol. 33, no. 4, pp. 580.e11–580.e19, 2019.
- [16] E. Cataldo, A. G. de Oliveira Aguiar, and L. A. Santos Silva, “A Methodological Review of Voice Acoustic Parameters, with Emphasis on Vocal Tremor: Definitions, Computation, and Software Tools,” *Journal of Voice*, 2026. doi: <https://doi.org/10.1016/j.jvoice.2026.03.013>