



OPEN

DATA DESCRIPTOR

NeuroVoz: a Castilian Spanish corpus of parkinsonian speech

Janaína Mendes-Laureano¹, Jorge A. Gómez-García¹, Alejandro Guerrero-López¹,
Elisa Luque-Buzo², Julián D. Arias-Londoño¹, Francisco J. Grandas-Pérez²
& Juan I. Godino-Llorente¹✉

The screening of Parkinson's Disease (PD) through speech is hindered by a notable lack of publicly available datasets in different languages. This fact limits the reproducibility and further exploration of existing research. To address this gap, this manuscript presents the NeuroVoz corpus consisting of 112 native Castilian-Spanish speakers, including 58 healthy controls and 54 individuals with PD, all recorded in ON state. The corpus showcases a diverse array of speech tasks: sustained vowels; diadochokinetic tests; 16 Listen-and-Repeat utterances; and spontaneous monologues. The dataset is also complemented with subjective assessments of voice quality performed by an expert according to the GRBAS scale (Grade/Roughness/Breathiness/Asthenia/Strain), as well as annotations with a thorough examination of phonation quality, intensity, speed, resonance, intelligibility, and prosody. The corpus offers a substantial resource for the exploration of the impact of PD on speech. This data set has already supported several studies, achieving a benchmark accuracy of 89% for the screening of PD. Despite these advances, the broader challenge of conducting a language-agnostic, cross-corpora analysis of Parkinsonian speech patterns remains open.

Background & Summary

Parkinson's Disease (PD) is a progressive neurodegenerative disorder primarily associated with ageing. Although it predominantly affects people over 65 year, it is a common and worldwide condition affecting approximately 61 million people in 2016¹. PD is characterised by neuronal degeneration in the substantia nigra, a region of the brain responsible for the production of dopamine². The resulting decrease in dopamine levels leads to various motor symptoms, including speech impairments, coordination difficulties, muscle rigidity, bradykinesia, among others, significantly affecting the quality of life of those affected³. In addition to motor problems, patients with PD often experience non-motor symptoms such as cognitive impairment, mood disturbances, and autonomic dysfunction⁴.

The diagnosis of PD is primarily based on the evaluation of motor and non-motor symptoms⁴, with a final confirmation that can only be provided by postmortem neuropathological examination⁵. Despite the clinical precision achieved, which is approximately 90%, reaching a definitive diagnosis can take an average of 2.75 years⁶. In addition to that, the varied symptomatology of PD and its overlap with other neurodegenerative diseases highlight the need for more sophisticated diagnostic tools to improve evaluation procedures⁷.

Recent advances⁸ underscore the potential of the integration of novel biomarkers and Machine Learning (ML) algorithms to improve diagnostic precision and accelerate disease identification, thus facilitating earlier intervention and potentially improving the quality of life of patients with PD. In this regard, speech, with its precise and coordinated movements, holds promise as a potential biomarker to detect and assess PD. Speech impairments are common in approximately 89% of patients with PD⁹, and are believed to be the result of a combination of broad motor deficits and specific difficulties related to the breathing and swallowing processes¹⁰. These impairments include deficits in voice production (e.g., dysponia), in articulation (e.g., dysarthria), and in prosody (e.g., dysprosody)^{11–18}. The literature has documented articulatory deficits in patients with PD, including reduced precision and variability in speech organ movements, leading to notable changes in speech patterns^{19,20}. These include variations in voice onset time²¹, consonant spirantization^{22,23}, altered formant frequencies²⁴, reduced area of the vowel space²⁵, and changes in speech rate²⁶, which collectively serve as potential indicators of the presence of PD. Regarding phonation deficits, PD patients' voice has been identified with hoarseness, tremor,

¹Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, 28040, Madrid, Spain. ²Department of Neurology, Hospital General Universitario Gregorio Marañón, 28007, Madrid, Spain. ✉e-mail: ignacio.godino@upm.es

roughness²⁷ or with breathiness and asthenia²⁸. There is also evidence of higher levels of noise and irregularity due to incomplete closure of the vocal folds, phase asymmetry, or vocal tremor, which have been verified instrumentally²⁹ and through noise measurements (jitter and Harmonics-to-Noise Ratio (HNR))^{30–33}. Similarly, variability in fundamental frequency has been identified to be significantly different in PD vs. controls^{28,32}. Regarding prosody, literature evidences that PD patients mainly present a reduction in Sound Pressure Level (SPL) or SPL variability³⁴. They also present a reduced frequency range^{18,34} as the most prominent feature.

With the above-mentioned evidence on the characteristics of Parkinsonian speech, literature reports the use of ML methodologies to exploit the close relationship between speech and PD. To this respect, automatic speech analysis has shown promising results, providing objective tools to detect and assess the severity of the neurodegenerative effects of the disease³⁵ using traditional ML techniques³⁶ and advanced Deep Learning (DL) approaches. In this respect, models using classical voice quality measurements¹⁷ or Mel-frequency cepstral coefficients³⁷ have shown the potential to model the influence of PD on the speech, with recent advances in DL offering even greater accuracy by extracting abstract high-level characteristics^{38–40}. However, training such models requires a broad variety of public datasets of Parkinsonian speech to ensure robustness and generalizability.

In PD speech research, public datasets are notably scarce. Currently, to the author's knowledge, there exists only one public dataset of Italian speakers⁴¹, featuring 831 audio files of 28 patients with PD and 50 Healthy Control (HC) speakers. It includes two reading tasks, the repetition of the syllables /pa/ and /ta/, and the sustained phonation of vowels. Although this data set provides valuable information, it is limited by its small cohort of PD patients, the absence of spontaneous speech, and the lack of manually transcribed speech. Furthermore, its reliance on text-dependent utterances read by the participants (i.e., not uttered in an spontaneous way) may pose challenges for elderly participants due to the higher cognitive load required. Regarding Spanish speech and language research, the GITA⁴² dataset contains data from Colombian Spanish speakers, but its access is restricted and is available only upon request. Regarding the English language, two datasets are well-known. First, the mPower corpus⁴³, an American English resource, contains data from self-identified PD and HC individuals. This dataset includes a survey, the MDS-UPDRS questionnaire, various activities, and a voice component featuring 10-second recordings of sustained vowel /a/. With 5,826 participants and 65,022 recordings, mPower is extensive, but its voice data is limited to a single vocal task, which may restrict its use in diverse research contexts. Another relevant dataset is the Mobile Device Voice Recordings recorded by King's College London (MDVR-KCL)⁴⁴, which includes recordings from 21 HC individuals and 16 PD patients. Each of the 37 participants performed two tasks: reading a paragraph of text and engaging in spontaneous dialogue. The recordings, stored in WAV format, have an average duration of 150 seconds each. However, the small size of this dataset and its limited participant pool highlight the need for more extensive datasets.

Addressing these significant gaps, NeuroVoz (v.3.0.0) emerges as the most comprehensive publicly available dataset of Parkinsonian speech research to date. In addition, it is the first fully public dataset in Spanish, marking a crucial contribution to the field. NeuroVoz not only increases resources for analyzing PD speech, but also serves as a tool for examining the articulatory, phonatory, and prosodic aspects of Parkinsonian speech. Moreover, this dataset significantly expands the research possibilities within the Spanish-speaking world, embracing the distinct linguistic features of Castilian Spanish and thus enriching the diversity and representativeness of Spanish speech corpora. The NeuroVoz dataset is a specialized collection of speech recordings from individuals diagnosed with PD and HC, designed to support the development and validation of ML models for the diagnosis and monitoring of PD.

NeuroVoz was jointly recorded by the Bioengineering and Optoelectronics (ByO) group from Universidad Politécnica de Madrid (UPM) and the Otorhinolaryngology and Neurology Services of Hospital General Universitario Gregorio Marañón (HGUGM) and Hospital Universitario de Fuenlabrada (HUF), Madrid, Spain. The data set includes a diverse set of speech tasks, ranging from sustained phonation of vowels, text-dependent utterances, a diadochokinetic (DDK) test, and a spontaneous free monologue. The specific material recorded is detailed in the Protocol Section.

The NeuroVoz data set has contributed significantly to research on the screening and detection of PD through speech analysis, marked by a series of studies that evolved in both methodology and focus. Moro *et al.*⁴⁵ achieved 82% accuracy using Acoustic Landmarks and a GMM-UBM-Blend classification strategy, emphasizing the utility of burst segments for the screening of PD. In a later work, Moro *et al.*⁴⁶ introduced a novel approach that combined speaker recognition and allophonic grouping, achieving an accuracy of 86%. Further studies, such as Moro *et al.*⁴⁷, focused on phonatory features that achieved up to 89% accuracy, while others^{48,49} placed the focus on an articulation analysis, reaching accuracies between 81%–89%. Arias *et al.*³⁹ explored the use of DL and transfer learning for Unified Parkinson's Disease Rating Scale (UPDRS) score prediction, highlighting the integration of feature engineering and learning approaches. Furthermore, Godino *et al.*⁵⁰ underscored artificial intelligence's role in the screening of PD using various methods, reporting accuracies over 85%. Lastly, Ibarra *et al.*⁵¹ improved the model generalisability using domain adversarial training, achieving 83% accuracy in a cross-corpora scenario and Guerrero *et al.*⁵² achieved a benchmarking 91 ± 9 discrimination power. Collectively, these studies establish the NeuroVoz dataset as a valuable benchmark for PD detection models, with the current leading performance achieving 89% accuracy within this dataset, and 83% for cross-corpora evaluations.

This database, NeuroVoz, marks a significant contribution to the scientific community by offering a comprehensive resource for the systematic exploration of Parkinsonian speech characteristics of Castilian Spanish speakers. It fills a notable gap in the available resources, as currently the only fully public and freely accessible dataset for the analysis of PD speech is in Italian⁴¹, leaving a void for research in Spanish, and more specifically, in Castilian Spanish. By making NeuroVoz publicly available in adherence to the principles of FAIR (Findable, Accessible, Interoperable, and Reusable), we aim to broaden the scope of understanding PD's impact on speech patterns.

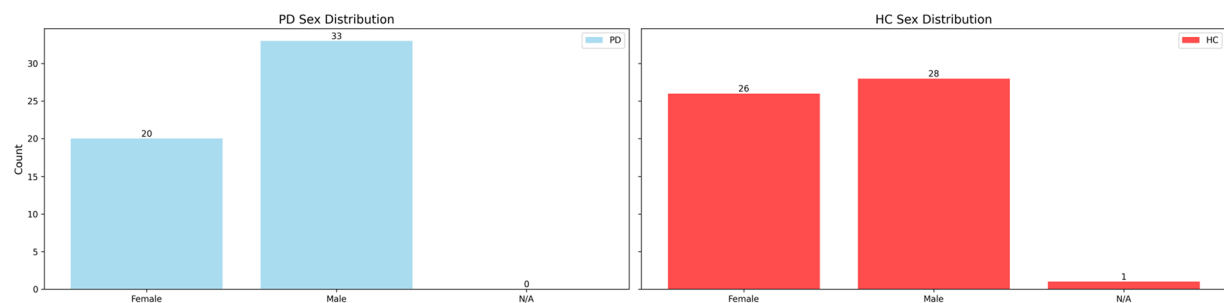


Fig. 1 Sex distribution of participants.

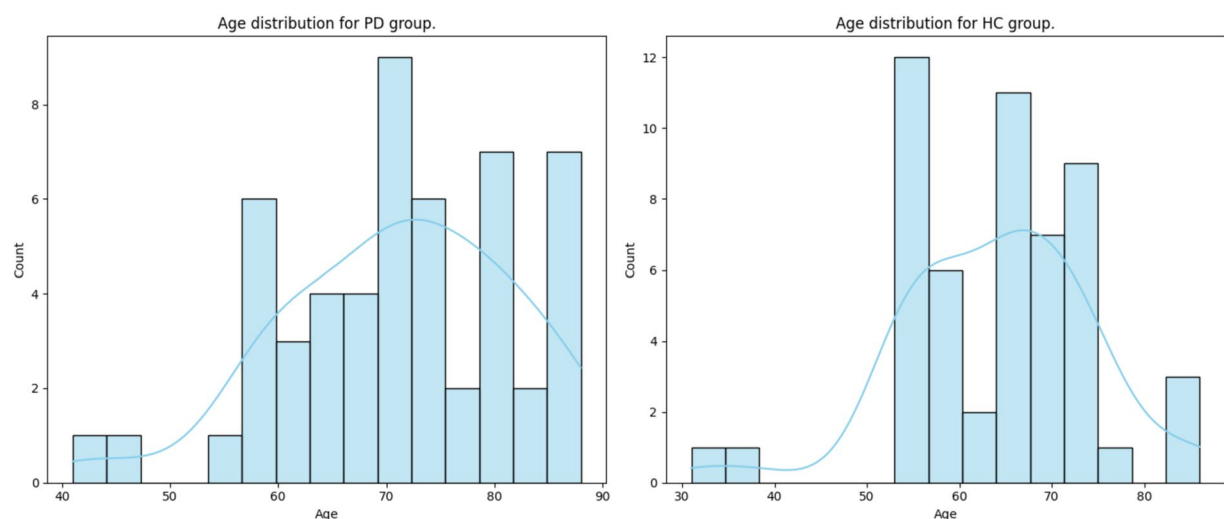


Fig. 2 Age distribution of participants.

Methods

A description of several aspects of the corpus is presented next, including the ethics declaration, a description of the participants, the speech protocol tasks, and the audio registration process.

Ethics declaration. The study was approved by the Ethics Review Board of the HUF and HGUGM with codes 18/11-ENM1 and 11/2015 respectively, and in accordance with the Spanish Ethical Review Act. All patients and controls were native Spanish speakers and followed the same experimental protocol.

All participants were informed about the project objectives and, if they agreed with the study conditions, they were recruited for participation and recording at HGUGM facilities. Participants received a document containing details about the project goals prior to recording. Subsequently, they were asked to sign a consent form. Participants did not receive any compensation for participating in the study, agreeing to share their voices for research purposes. Patients were informed of their rights and the possibility of leaving the study at any time.

Patients were individually identified with a code, which is different from the one used in the Hospital for their clinical histories. No personal data was exchanged with external researchers who had access to the corpus. Only one specialist got in contact with each patient, being also in charge of collecting the clinical data.

Participants. The dataset comprises 112 individuals, evenly distributed in terms of age and sex, with 46 females, 61 males, and one participant whose sex information was not provided. All participants are native Castilian Spanish speakers.

NeuroVoz includes two cohorts: a PD group, consisting of patients clinically diagnosed with PD; and a HC group, comprising normophonic individuals. In total, 58 HC speakers and 54 patients with PD were recorded. The mean age of the participants was 71.13 ± 10.62 for the PD group and 64.04 ± 10.26 for the HC group. The total number of subjects is limited by the availability of patients received in the Neurology Service of HGUGM and HUF.

Figure 1 shows the sex distribution for the PD and HC groups, while Fig. 2 presents the age distribution for both cohorts. The exclusion criteria for PD patients and HC participants is included in Table 1.

All participants were evaluated according to the Grade, Roughness, Breathiness, Asthenia, and Strain (GRBAS) perceptual scale⁵³, and PD patients were also evaluated according to the UPDRS⁵⁴ and Hoehn and Yahr (H-Y)⁵⁵ clinical scales.

Common Exclusion Criteria	
A history of voice or speech alterations or laryngeal or vocal cords surgery.	
Regular consumer of alcohol.	
Smoker.	
Speaker whose mother tongue is not Spanish.	
Speaker suffering from any disease that affects speech intelligibility or affects speech or voice.	
PD Patients	HC Participants
Diagnosis of another neurological disease.	Participants with neurological disorders and/or PD.

Table 1. Exclusion Criteria for PD and HC participants.

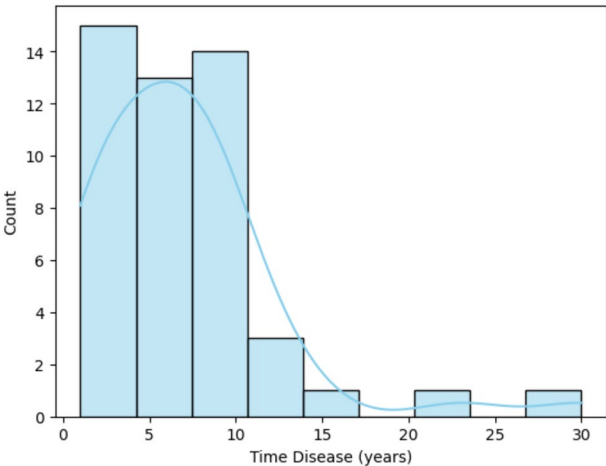


Fig. 3 Distribution of years since the diagnosis of PD.

To ensure the exclusion of organic vocal pathologies (i.e. lesions, scar, reflux) and facilitate a comprehensive clinical diagnosis, an otolaryngologist also performed an evaluation of symptoms such as vocal tremor, cephalic tremor, mandibular tremor, siallorrhea, dysphagia, and hipophonic voice.

The inclusion criteria for Parkinsonian speakers included patients diagnosed with PD by the HGUGM and HUF Neurology Services, no matter the start date of clinical signs. Participants with PD were recorded in ON state, having taken the prescribed medication -when applicable- 2 to 5 hours before the recording session. Most patients have been diagnosed with PD for less than 10 years, as seen in Fig. 3.

For the HC group, volunteers were recruited based on the following criteria: they were not professional voice users (i.e., their voice is not essential for their job); had no neurological diseases; and they had no diagnosis of PD. In addition, the HC group was selected to match Parkinsonian patients in terms of sex and age. The HC group was also subjected to a neurological state assessment by a survey.

The distribution of GRBAS total scores, i.e. sum up all GRBAS values for the same patient for each audio, are shown in Fig. 4.

Recording protocol. PD is a neurodegenerative disease causing movement disorders and, thus, the effect of PD is initially expected to be mother tongue independent. But articulations during speech production are slightly different for each language (or even dialect), due to differences mainly attributed to phonetic, phonological and linguistic variances. Thus, the effect of PD in the speech is expected to introduce small differences that are language and/or dialect dependent. Besides, certain manner classes are considered more relevant than others for a good characterization of the effect of the disease⁴⁹. Thus, the recording protocol should pay attention to be phonetically balanced, but also to include a wide variety of the most affected phonemes by the disease. To this respect, literature reports that plosives are very relevant for the detection of PD, but also fricatives⁴⁹.

In view of the aforementioned, an extensive recording protocol was designed. The protocol was developed to facilitate a further detailed analysis of known deviations introduced by the disease with respect to the expected normality.

The protocol includes four different speech tasks:

- **Sustained phonation of vowels:** Participants were instructed to phonate the five Spanish vowels /a/, /e/, /i/, /o/, and /u/ at a comfortable pitch and volume, taking brief pauses to breathe between each phonation. This procedure was repeated three times. The mean duration of sustained vowel phonation was 3.89 ± 0.73 s. for the PD group and 4.38 ± 1.54 s. for the HC group.

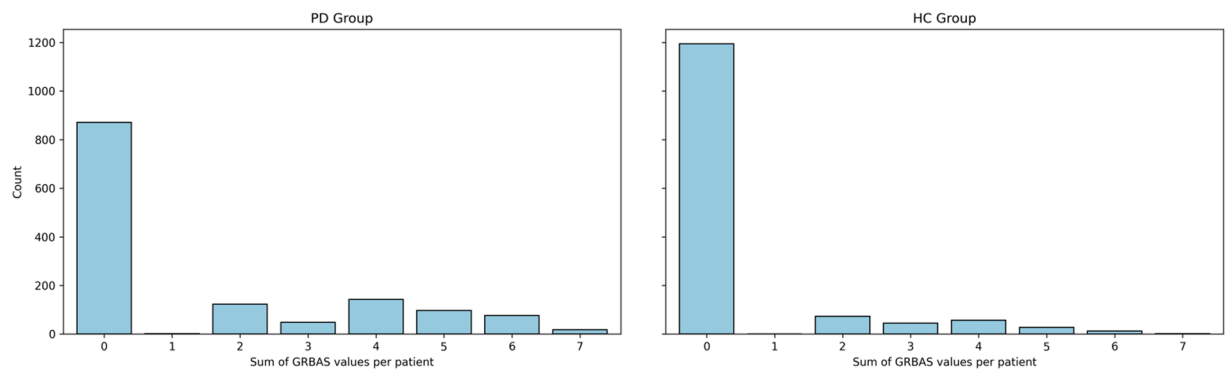


Fig. 4 GRBAS total sum distribution of participants.

Order	Task	Description
1	Phonation	Repeating sustained vowels to evaluate phonation.
2	Velopharyngeal Closure	LR phrases 1, 2, 6, 9, 10, 11, 12 and 13 from Table 3 for velopharyngeal closure assessment.
3	Articulation	DDK test to evaluate articulation capabilities.
4	Prosody	LR phrases 3 and 7 from Table 3 for prosody evaluation.
5	Intonation-Emotion	LR phrases 5, 14, and 15 from Table 3 with emphasis on capitalised words for intonation-emotion assessment.
6	Other tasks	LR phrases 4, 8, and 10 from Table 3.

Table 2. Speech tasks protocol followed in order.

Sent. #	ID	Spanish transcription	IPA transcription	English translation
1	ABLANDADA	<i>La patata no está bien ablandada</i>	[la pa'tata no 'esta βjen aβlan'dada]	"The potato is not soft enough"
2	ACAMPADA	<i>Mañana vamos de acampada</i>	[ma'ʎana 'βamos de akam'pada]	"Tomorrow we are going camping"
3	BARBAS	<i>Cuando las barbas de tu vecino veas pelar, pon las tuyas a remojar</i>	[kwando las βarβas de tu βeθino βeas pelar pon las tuβas a remoβar]	"When your neighbor's beard you see peeling, put yours to soak"
4	BURRO	<i>Burro grande ande o no ande</i>	['bur.o 'gran.de 'an.de o no 'an.de]	"Big donkey, walk or not walk"
5	CALLE	<i>De la calle VENDRÁ quien de tu casa te ECHARÁ</i>	[de la kaβe βendra kjen de tu kasa te etβara]	"From outside will come who will kick you out from your house"
6	CARMEN	<i>Carmen baila el mambo</i>	['kar.men 'bai.la el 'mam.bo]	"Carmen dances the mambo"
7	DIABLO	<i>Cuando el diablo no sabe qué hacer, con el rabo mata moscas</i>	[kwando el βaβlo no saβe ke aθer kon el raβo mata moskas]	"When the devil does not know what to do, it kills flies with its tail"
8	GANGA	<i>Esto es una ganga</i>	['es.to es 'u.na 'gan.ga]	"This is a bargain"
9	MANGA	<i>Juan tira de la manga</i>	['xwan 'ti.ra de la 'man.ga]	"Juan pulls the sleeve"
10	PERRO	<i>Dame pan y llámame perro</i>	['da.me pan i ja.ma.me 'pero]	"Give me bread and call me dog"
11	PAN	<i>Al pan pan y al vino vino</i>	[al pan pan i al βino βino]	"To the bread, bread, and to the wine, wine"
12	PATATA	<i>La patata blanda es buena</i>	[la pa'tata 'blanda es 'bwena]	"The soft potato is good"
13	PETACA	<i>La petaca blanca es mía</i>	[la pe'taka 'blanka es 'mi.a]	"The white flask is mine"
14	PIDIO	<i>No pidas a quien PIDIÓ, ni sirvas a quien SIRVIÓ</i>	[no piðas a kjen piðo ni sirβas a kjen sirβjo]	"Do not beg the one who begged, nor serve the person who served"
15	SOMBRA	<i>El que a BUEN árbol se arrima, BUENA sombra le cobija</i>	[el ke a βwen 'arβol se a'rima, 'bwena 'sombra le ko'βixa]	"Who leans close to a good tree is sheltered by good shade"
16	TOMAS	<i>Tomás tira de la manta</i>	[to'mas 'ti.ra de la 'man.ta]	"Tomás pulls the blanket"

Table 3. Transcriptions in the IPA and translations of selected sentences.

- **Listen and Repeat (LR) sentences:** Participants were instructed to listen to 16 predefined phrases and repeat them aloud. This method, chosen over reading a text, aimed to lessen the cognitive load of the process for the patients. Sentences are familiar sayings that are easy to recall due to their recognizability. Additionally, this method mitigates potential biases in speech that could arise during reading due to vision impairments common among the elderly population, thereby facilitating a more natural and impartial articulation process. The predefined phrases are detailed in Table 3. The mean duration of the sentences was 11.11 ± 4.64 s. for the PD group, and 13.06 ± 5.46 s. for the HC group. These sentences were chosen to be acoustically balanced in conjunction, but also to ensure a wide variety of plosive and fricative sounds, and to force frequent velum openings and closures.



Fig. 5 The free monologue is guided by asking the patient to describe the scenes presented in this illustration.

- **DDK test:** Participants were instructed to perform the repetition of the sequence /pa-ta-ka/ during, at least, 5 s., and as fast as possible. The mean duration of the DDK tests were 3.26 ± 2.45 s. for the PD group and 3.70 ± 2.74 s. for the HC group.
- **Monologues:** Participants were also asked to deliver a free monologue, describing a predetermined illustration that depicts various activities of daily living, as illustrated in Fig. 5. This task was designed to gather spontaneous speech, further enriching the dataset with more natural speech patterns. The average duration of these monologues was 31.15 ± 12.43 s. for the PD group, and 46.53 ± 21.41 s. for the HC group.

The order followed to perform the speech recordings is detailed in Table 2. Initially, for a further evaluation of the phonation, each sustained vowel was recorded independently. In the second place, eight different LR phrases were recorded to assess the velopharyngeal closure, which is often associated with Parkinsonian speech⁵⁶. Subsequently, the patient underwent the DDK test, which is typically used to evaluate the articulation ability. Next, three additional LR phrases were recorded for a potential analysis of prosodic aspects. Intonation and emotion were addressed by asking the patient to emphasize certain words in a more emotional manner (represented with capital letters in Table 3). Finally, three more LR phrases were recorded.

Table 3 specifies the phrases used for the LR tasks along with their translation to English and their International Phonetic Alphabet (IPA) transcriptions.

The audio was manually post-processed and edited to remove background noise, coughing, and long silences. Moreover, if the audio quality was poor, it was removed from the dataset, Fig. 6 shows the completeness of each audio material for all patients.

Recording setup. The audio recording setup was standardised to ensure high-quality data suitable for further analysis.

All participants were recorded under the same environmental conditions and using the same equipment. The technical setup for the audio recordings consisted of an AKG® C420 headset microphone. This microphone was connected to a preamplifier equipped with phantom power, ensuring optimal signal quality. The preamplifier was attached to a SoundBlaster® Live sound card, which operates at a sampling rate of 44,100 Hz. The recording system achieved a mean Signal-to-Noise (SNR) of 24.3 dB, indicating a high-quality audio capture. The audio device was connected to a personal computer running MedivozCaptura® software⁵⁷. The recording was carried out in a quiet room during the clinical routine, but in a non-acoustically isolated environment.

Throughout the recording sessions, the patient's comfort was prioritized. The participants were instructed to speak in their natural pitch and with a comfortable loudness. Pristine attention to the recording conditions not only ensures the comfort of the participants, but also contributes to the consistency and reliability of the collected data.

The audio files were also carefully post-processed and edited to exclude interjections, coughing, and background noise during silence periods.

Metadata associated to the speaker. In addition to speech recordings, demographic and clinical data were systematically collected for each participant at the beginning of each session. These data are summarized in Table 4. The date of the evaluation and recording was rounded to the first day of the corresponding month to avoid a potential identification of the speaker by crossing information.

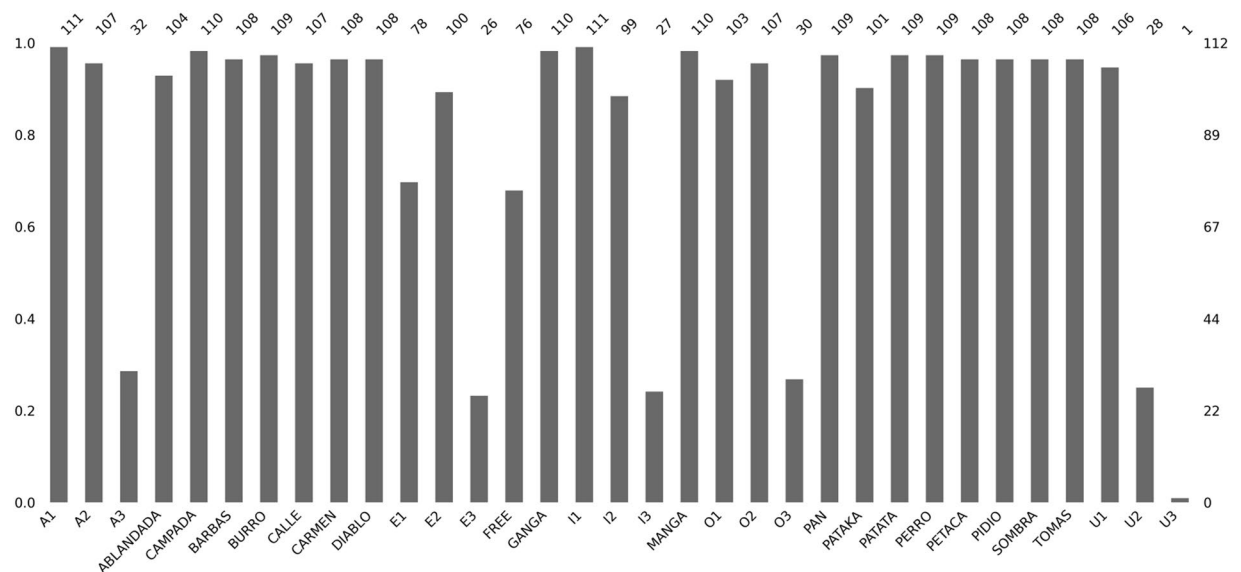


Fig. 6 Percentage of completeness of each audio material for all patients.

Data Type	Description
Date of Recording	Date
Sex	Free text string
Age	Integer value
Diagnosis	Free text field
Symptoms	Binary variables (True/False):
	– Vocal tremor
	– Cephalic tremor
	– Mandibular tremor
	– Sialorrhea
	– Dysphagia
	– Hypophonic voice
Duration of disease	Integer value (years)
URPDS	UPDRS scale
H-Y	H-Y scales
Date of evaluations	Date
Medication	Free text field
Medical treatment during recording	Annotated “ON”/“OFF”
Occupation	Free text string
Vocal folds analysis	String field (Normal/Not performed/free text)
Observations	Free text field for additional notes
Doctor	Categorical value to indicate who evaluated the speaker

Table 4. Summary of the metadata collected for each participant.

The sex is stored as a float where “1” indicates male, “0” female and “NaN” for non-declared cases. Age was recorded as an integer. Moreover, the participants’ employment status was also documented, revealing that 63.88% of the total 112 individuals were retired, which is in consonance with the demographic composition of the dataset.

The diagnostic data identified by the physician included two different diagnoses, being Idiopathic PD the most prevalent (affecting 50 patients) and Parkinsonian Syndrome the less frequent (affecting the remaining 5 patients). An extra field contains medical annotations about the diagnosis, such as the most affected body part.

Binary indicators were used to annotate symptoms such as vocal tremor, cephalic tremor, mandibular tremor, sialorrhea, dysphagia, and hypophonic voice, with “1” representing presence and “0” absence, as shown in Fig. 7. In case of missing data, a float “NaN” is found. Medication details, when applicable, were recorded in a free text field. The time since diagnosis (in years) was also recorded.

Clinical evaluations were carried out following the UPDRS scale⁵⁴, and H-Y staging system⁵⁵. UPDRS was developed by the International Parkinson and Movement Disorders Society (MDS) to evaluate various aspects

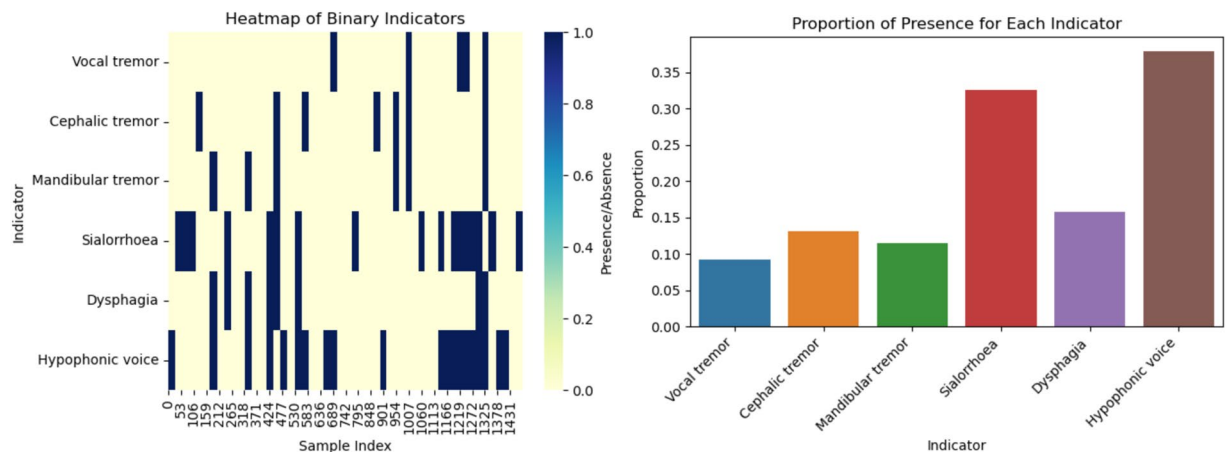


Fig. 7 Binary indicators regarding parkinsonian symptoms. Left: Binary indicators of the presence of symptoms for each PD patient. The horizontal axis determines the patient ID while the vertical axis denotes each symptom. Right: Binary proportion of presence of symptoms for each PD patient.

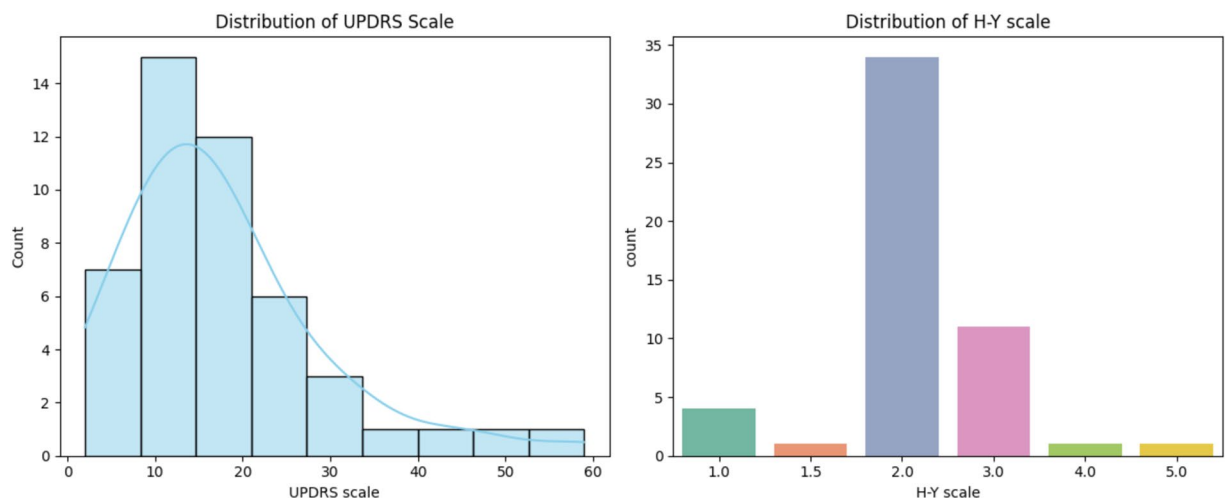


Fig. 8 Severity distribution of parkinsonian patients expressed in UPDRS and H-Y scales.

of PD including non-motor and motor experiences of daily living and motor complications. It characterizes the extent and burden of disease ranging from 0 and 260. On the other hand, the H-Y scale is used to characterize the progression of the disease, classifying the patient according to the severity of the symptoms ranging from 0 to 5. The date on which the patients were evaluated was annotated and anonymized to the first day of the month. Both clinical evaluations were carried out by four different physicians, all of them certified by the MDS for this purpose. Figure 8 illustrates a histogram of the results of these evaluations.

The findings of the direct examination of the vocal folds using fiber-laryngoscopic methods were also documented, revealing that all participants except one HC had normal vocal folds (this participant had a hypophonic right vocal corone). In any case, it is worth noting that a relevant proportion of participants declined this examination —28.30% of the PD group, and 38.11% of the HC group.

Metadata associated to the audio recordings. Each recording also has complementary information associated, such as the audio manual transcription, the GRBAS perceptual evaluations, or a set of voice quality characteristics. The metadata associated are dependent on the specific acoustic material in the recording, being different for vowels, monologues, or LR eloquences.

Audio transcriptions. To ensure accurate speech annotation, manual transcriptions were meticulously performed for both LR eloquences and monologues. This is a critical task, especially for the LR eloquences. The goal was to identify articulation mistakes made by the speakers or linguistic discrepancies between the utterances produced by the speakers and those expected. Trained transcribers carefully listened to each audio, recording the transcriptions word by word, and noting any deviations or mispronunciations.

Parameter	Abbreviation	Unit Measure
Perturbation Measures		
Absolute Jitter	Jitter	μ Seconds
Relative Jitter	rJitter	%
Relative Average Perturbation	RAP	%
Pitch Period Perturbation Quotient	rPPQ	%
Smoothed Pitch Period Perturbation Quotient	rSPPQ	%
Absolute Shimmer	ShimmerDb	dB
Relative Shimmer	rShimmer	%
Amplitude Perturbation Quotient	APQ	%
Smoothed Amplitude Perturbation Quotient	sAPQ	%
Cepstral Peak Prominence	CPP	dB
Noise Parameters		
Harmonics-to-Noise Ratio	HNR	dB
Cepstrum Harmonics-to-Noise Ratio	CHNR	dB
Glottal to Noise Excitation Ratio	GNE	Ratio
Normalised Noise Energy	NNE	dB
Tremor Parameters		
Frequency Tremor Intensity Index	FTRI	Arbitrary Units
Amplitude Tremor Intensity Index	ATRI	Arbitrary Units
Fundamental Frequency Tremor Frequency	FFTR	Hz
Amplitude Tremor Frequency	ATRF	Hz

Table 5. Voice measurements extracted from the recordings of the /a/ sustained vowels.

Voice quality features. A set of voice quality features, detailed in Table 5, was calculated and included in the corpus. These features were extracted for the three repetitions of all sustained vowels. The AVCA-ByO open-source package⁵⁸ was used to extract three subsets of characteristics, namely: i) amplitude and frequency perturbation features; ii) noise parameters; and, iii) tremor parameters.

Perceptual evaluations. The GRBAS perceptual scale⁵³, is a globally recognized tool for vocal assessment^{59–61}, providing a structured method for clinicians and researchers to evaluate dysphonia⁶². It is both reliable and practical, posing no discomfort or inconvenience to either the patient or the evaluator⁶³. Each GRBAS element is scored with a unique value, and a cumulative score combining all is also obtained. The distribution of these evaluations for each group are detailed in Fig. 9.

A speech therapist with more than 15 years of professional experience evaluated each audio using the scale, scoring the five different elements of vocal quality. Grade (G) (the overall severity of dysphonia), Roughness (R), Breathiness (B), Asthenia (A) (vocal weakness), and Strain (S). The evaluation consists in assigning an integer value ranging from 0 to 3 to each element of the scale, where “0” represents normal, and “1”, “2”, and “3” indicate slight, moderate, and severe abnormalities, respectively.

For the sustained vowels, a GRBAS rating was assigned to every recording, along with additional comments from the speech therapist. For LR tasks and spontaneous monologues, the GRBAS evaluation was supplemented with an additional perceptual voice analysis, including an evaluation of the glottal attack (rated as normal, soft, hard, or somewhat hard), tone (normal, high-pitched, deep, or unstable), and phonation quality (normal, forced-strangled, breathy, wet, or blown). Intensity levels were categorized as normal, weak, or increased, while speech rate was evaluated as normal, slow, or accelerated. Resonance was described as normal, hypernasal, or hyponasal. Intelligibility was graded on a scale from 1 (normal) to 5 (severe deficiency), and speech prosody was assessed as normal, monotonous, variable intensity, or excessive accentuation. Lastly, articulation was evaluated as normal, slurred, or exaggerated.

Data Records

Data access is facilitated through the Zenodo repository⁶⁴. The structural representation of the data set can be seen in Fig. 10. The total size of the data set is 1.01 GB, comprising a collection of 2,977 audio files. On average, each patient contributes approximately with 26.88 ± 3.35 audio recordings.

The data is distributed into five distinct folders, namely, “audio features”, “audios”, “grbas”, “transcriptions” and “metadata”. In the subsequent sections, a detailed overview of the contents of each of these folders is provided.

Voice quality features. The “audio features” folder contains all precomputed voice quality features. These features are available exclusively for sustained vowels. These voice quality features are stored in a .csv file named `audio_features.csv`. The .csv file comprises 18 columns: the initial column indicates the path to the audio file, while the remaining 17 columns represent each of the precomputed voice quality features.

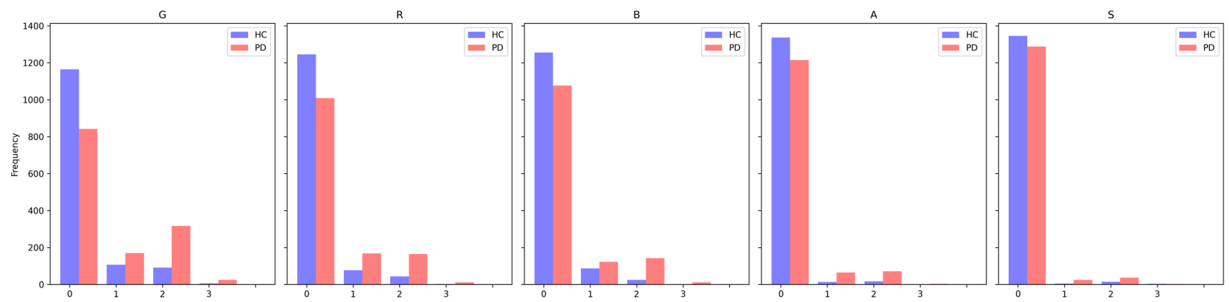


Fig. 9 Distribution of GRBAS individual evaluations for HC and PD patients.

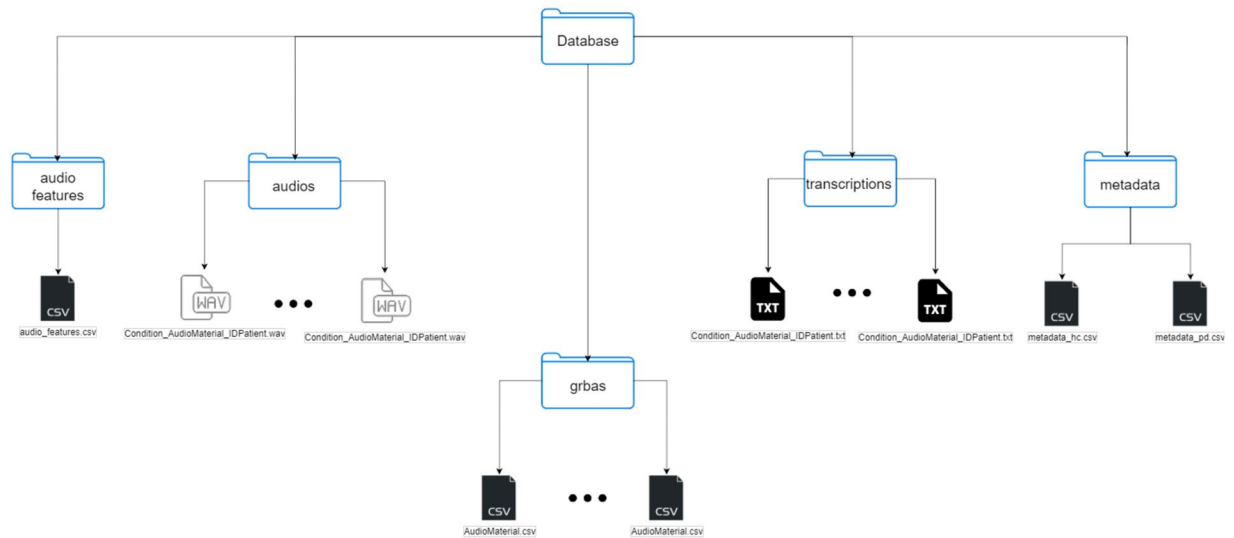


Fig. 10 Database tree structure: hierarchical representation of the data records and their organization within the database.

Audio recordings. The "audios" directory serves as the repository for all audio recordings, each stored in the .wav format. The naming convention of the audio files encapsulates critical information about the patient's condition, the specific audio material recorded, and the unique identifier of the patient, structured as follows:

"Condition_AudioMaterial#Repetition_IDPatient.wav".

This format is deciphered as follows:

- **Condition:** Identifies the patient's health status, categorized as either HC or PD, according to the health condition.
- **AudioMaterial:** Denotes the type of acoustic material, which may be "A", "E", "I", "O", or "U" for the sustained vowels; a keyword (as detailed in Table 3) for the text dependent utterances; "FREE" for free speech monologues; or "PATAKA" for the DDK test.
- **#Repetition:** Applicable only to the sustained vowels -performed three times-, indicating the sequence of the attempt (ranging from 1 to 3).
- **IDPatient:** A four-digit integer uniquely designating the patient.

For instance, the filename HC_A2_0034.wav represents the recording of the second attempt of the sustained vowel /a/ by patient 0034, who belongs to the HC cohort. Similarly, PD_GANGA_0066.wav corresponds to the recording of the phrase tagged as "GANGA" in Table 3 for patient 0066, belonging to the PD group.

GRBAS evaluations. The "grbas" folder contains all GRBAS evaluations performed by a speech therapist. These data are organized into 32 distinct .csv files, each named according to the AudioMaterial identifier.

For recordings of sustained vowels, such as A1.csv for the first phonation of /a/ or E3.csv for the third phonation of /e/, the files feature 8 columns. These include: the filename, indicating the .wav file's name; five columns, holding an integer score with the evaluation corresponding to each of the five GRBAS features; a column for the overall GRBAS score; and a column to record the observations of the therapist about the recording.

Regarding other audio tasks, such as LR or DDK tasks, different .csv are found. These .csv files, for example Barbas.csv, contain 16 columns. The initial 8 mirror those for the vowel assessments, detailing

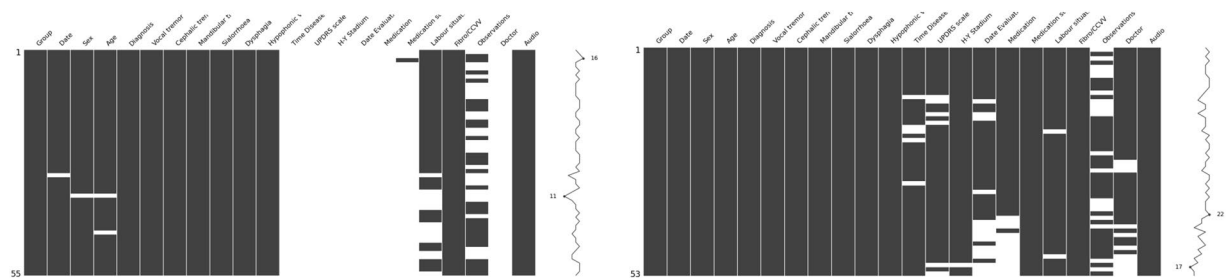


Fig. 11 Percentage of missing data per column variable. Blank spaces mean no available data. Left: HC group. Right: PD group.

filenames, GRBAS partial scores, total GRBAS score and comments. The subsequent eight columns provide the therapist's perceptual notes, including glottal attack (classified as normal, soft, hard, or somewhat hard), tone (normal, high-pitched, grave, or unstable), and quality of phonation (normal, forced-strangled, breathy, wet, or trembling). Intensity is labelled as normal, weak, or increased; speech pace as normal, slow, or fast; resonance as normal, hypernasal, or hyponasal. Intelligibility ranges from normal to severe deficiency. Prosody is assessed for normality, monotony, intensity variations, or excessive emphasis. And articulation is evaluated as normal, slurred, or overstated.

Transcriptions. The "transcriptions" directory serves as the repository for all audio manual transcriptions, each stored in a .txt file encoded in latin-1. The naming convention of the text files follows the same rules as the audio files, structured as follows:

"Condition_AudioMaterial_IDPatient.txt".

This format is deciphered as:

- **Condition:** Identifies the patient's health status, categorized as either HC or PD, reflecting the health condition.
- **AudioMaterial:** Denotes a keyword from specific phrases (as detailed in Table 3) or "FREE" for free speech monologues.
- **IDPatient:** A four-digit integer uniquely designating the patient.

For instance, the filename HC_ABLANDADA_0034.txt represents the text transcription of the first sentence in Table 3 by patient 0034, who falls under the HC category. Similarly, PD_GANGA_0066.txt corresponds to the transcription of the audio tagged as "GANGA" in Table 3 for patient 0066, identified with PD.

Other metadata. The "metadata" folder contains the clinical data and metadata recorded in addition to the audio files. It comprises two .csv files, one for each condition (i.e., HC and PD) named metadata_hc.csv and metadata_pd.csv.

Each .csv file is structured around 23 columns: the initial column represents the patient's ID; the second column specifies the cohort ("HC" or "PD"); and the subsequent 10 columns the detailed clinical data as outlined in the methodology. The last column contains the file path to the corresponding audio recording.

Technical Validation

The NeuroVoz corpus has been rigorously examined in several studies^{39,45–51}, establishing its usefulness to differentiate between parkinsonian and normophonic speech with a benchmark accuracy of 89 ± 7 reported in⁴⁹ using LR tasks.

Data integrity and completeness are visually represented in Fig. 11, providing a clear view of the missing data, and reinforcing the dataset's structured approach towards capturing distinct participant groups, while notably, PD-specific attributes are absent in the HC group.

Technical validation was further enhanced through a naïve classification procedure to categorize the samples into PD or HC. Initially, features extracted using the AVCA-ByO toolbox were divided into training and testing sets, ensuring an 80-20 split by patient's ID. After normalization, these features were used to train a simple Random Forest classifier and a Logistic Regression model, yielding balanced accuracies of 64% and 65%, respectively.

Additionally, audio files underwent preprocessing and were classified using a pre-trained ResNet-18 Convolutional Neural Network (CNN) from the Pytorch[®] timm library. After segmenting audios into 400 ms frames with 50% of overlap and computing mel-spectrograms, these normalized spectrograms were used to fine-tune the ResNet-18 over 10 epochs, classifying the samples into PD or HC with a balanced accuracy of 69%.

These validation efforts, employing both audio features and direct audio analysis, do not aim to improve the benchmark results in the state of the art, but demonstrate the utility of the data set for the screening of PD. Thus, the noted discrepancy between these results and those reported in previous works underscores the need for more advanced methodologies and further experiments.

Usage Notes

The use of the dataset is restricted to a Data Usage Agreement (DUA) available in the repository⁶⁴. Final users should copy the entire DUA into the 'request message' box at the Zenodo repository, also filling in the personal information requested. All applications that agree to the terms and conditions will be considered to access the corpus.

The NeuroVoz dataset is complemented with a set of Python scripts for data management and analysis. These scripts were designed to provide a practical introduction to working with the dataset and to facilitate initial experiments for the screening of PD using this corpus.

The script "simple_predictors_af.py" offers a straightforward demonstration on processing and analyzing voice quality features extracted with the AVCA-ByO toolbox, followed by a simple application using Random Forest and Logistic Regression models to classify the speech samples as either HC or PD. This script is intended to inspire further analytical explorations.

For those interested in working directly with the audio files, the "simple_predictors_audio.py" script is included. This script outlines the main steps for audio preprocessing, mel-frequency spectrogram calculations, and subsequent classification with a pre-trained ResNet-18 CNN using the Pytorch timm library. It exemplifies how to leverage DL for PD detection from the speech, potentially paving the way for novel research directions.

The notebook "extract_audio_features.ipynb" provides code to extract 6373 features using OpenSmile Compare 2016. This comprehensive feature extraction process is essential for detailed audio analysis and facilitates subsequent ML applications to classify speech data.

Code availability

The codes used for data preprocessing and cleaning can be accessed in a specific GitHub® repository (https://github.com/BYO-UPM/NeuroVoz_Dababase). Additionally, the referenced scripts used for classification purposes are included in the same repository. The AVCA-ByO toolbox (<https://github.com/BYO-UPM/AVCA-ByO>) used to extract the voice quality features is also available as an open-source package⁵⁸.

Received: 19 July 2024; Accepted: 28 November 2024;

Published online: 18 December 2024

References

- Bloem, B. R., Okun, M. S. & Klein, C. Parkinson's disease. *The Lancet* **397**, 2284–2303 (2021).
- Aarsland, D. *et al.* Cognitive decline in parkinson disease. *Nature Reviews Neurology* **13**, 217–231 (2017).
- Friedman, J. H. *et al.* Fatigue in parkinson's disease: report from a multidisciplinary symposium. *NPJ Parkinson's Disease* **2**, 1–6 (2016).
- Pfeiffer, R. F. Non-motor symptoms in parkinson's disease. *Parkinsonism & Related Disorders* **22**, S119–S122 (2016).
- Koga, S. *et al.* When dlb, pd, and psp masquerade as msa: an autopsy study of 134 patients. *Neurology* **85**, 404–412 (2015).
- Rossi, M., Perez-Lloret, S. & Merello, M. How much time is needed in clinical practice to reach a diagnosis of clinically established parkinson's disease? *Parkinsonism & Related Disorders* **92**, 53–58 (2021).
- Tolosa, E., Garrido, A., Scholz, S. W. & Poewe, W. Challenges in the diagnosis of parkinson's disease. *The Lancet Neurology* **20**, 385–397 (2021).
- Pujols, J. *et al.* Small molecule inhibits α -synuclein aggregation, disrupts amyloid fibrils, and prevents degeneration of dopaminergic neurons. *Proceedings of the National Academy of Sciences* **115**, 10481–10486 (2018).
- Ramig, L. O., Fox, C. & Sapir, S. Speech treatment for parkinson's disease. *Expert Review of Neurotherapeutics* **8**, 297–309 (2008).
- Robbins, J. A., Logemann, J. A. & Kirshner, H. S. Swallowing and speech production in parkinson's disease. *Annals of Neurology* **19**, 283–287 (1986).
- Weismer, G. *et al.* Articulatory characteristics of parkinsonian dysarthria: Segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination. *The Dysarthrias: Physiology, Acoustics, Perception, Management* 101–130 (1984).
- Skodda, S., Visser, W. & Schlegel, U. Vowel articulation in parkinson's disease. *Journal of Voice* **25**, 467–472 (2011).
- Rusz, J., Cmejla, R., Ruzickova, H. & Ruzicka, E. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease. *The Journal of the Acoustical Society of America* **129**, 350–367 (2011).
- Ackermann, H. & Ziegler, W. Articulatory deficits in parkinsonian dysarthria: an acoustic analysis. *Journal of Neurology, Neurosurgery & Psychiatry* **54**, 1093–1098 (1991).
- Kegl, J., Cohen, H. & Poizner, H. Articulatory consequences of parkinson's disease: perspectives from two modalities. *Brain and Cognition* **40**, 355–386 (1999).
- Duffy, J. R., Strand, E. A. & Josepshs, K. A. Motor speech disorders associated with primary progressive aphasia. *Aphasiology* **28**, 1004–1017 (2014).
- Moro-Velazquez, L., Gomez-Garcia, J. A., Arias-Londoño, J. D., Dehak, N. & Godino-Llorente, J. I. Advances in parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects. *Biomedical Signal Processing and Control* **66**, 102418 (2021).
- Moro-Velazquez, L. & Dehak, N. A review of the use of prosodic aspects of speech for the automatic detection and assessment of parkinson's disease. In *Automatic Assessment of Parkinsonian Speech Workshop, Cambridge, Massachusetts, USA*, 42–59 (Springer, 2020).
- Walsh, B. & Smith, A. Basic parameters of articulatory movements and acoustics in individuals with parkinson's disease. *Movement Disorders* **27** (2012).
- Svensson, P., Henningson, C. & Karlsson, S. Speech motor control in parkinson's disease: a comparison between a clinical assessment protocol and a quantitative analysis of mandibular movements. *Folia Phoniatrica* **45**(4), 157–64 (1993).
- Chenausky, K. V., MacAuslan, J. & Goldhor, R. S. Acoustic analysis of pd speech. *Parkinson's Disease* **2011** (2011).
- Godino-Llorente, J. I., Shattuck-Hufnagel, S., Choi, J.-Y., Moro-Velázquez, L. & Gómez-García, J. A. Towards the identification of idiopathic parkinson's disease from the speech. new articulatory kinetic biomarkers. *PLoS ONE* **12** (2017).
- Antolík, T. K. & Fougerson, C. Consonant distortions in dysarthria due to parkinson's disease, amyotrophic lateral sclerosis and cerebellar ataxia. In *Interspeech* (2013).
- Weismer, G. & Wildermuth, J. Formant trajectory characteristics in persons with parkinson, cerebellar, and upper motor neuron disease. *Journal of the Acoustical Society of America* **103**, 2892–2892 (1998).
- Rusz, J. *et al.* Imprecise vowel articulation as a potential early marker of parkinson's disease: effect of speaking task. *Journal of the Acoustical Society of America* **134** 3, 2171–81 (2013).

26. McRae, P.-A., Tjaden, K. & Schoonings, B. Acoustic and perceptual consequences of articulatory rate change in parkinson disease. *Journal of Speech, Language, and Hearing Research* **45** 1, 35–50 (2002).
27. Logemann, J. A., Fisher, H. B., Boshes, B. & Blonsky, E. R. Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients. *Journal of Speech and Hearing Disorders* **43**, 47–57 (1978).
28. Midi, I. *et al.* Voice abnormalities and their relation with motor dysfunction in parkinson's disease. *Acta Neurologica Scandinavica* **117**, 26–34 (2008).
29. Perez, K. S., Ramig, L. O., Smith, M. E. & Dromey, C. The parkinson larynx: tremor and videostroboscopic findings. *Journal of Voice* **10**, 354–361 (1996).
30. Ramig, L. A., Titze, I. R., Scherer, R. C. & Ringel, S. P. Acoustic analysis of voices of patients with neurologic disease: rationale and preliminary data. *Annals of Otology, Rhinology & Laryngology* **97**, 164–172 (1988).
31. Rahn III, D. A., Chou, M., Jiang, J. J. & Zhang, Y. Phonatory impairment in parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis. *Journal of Voice* **21**, 64–71 (2007).
32. Tanaka, Y., Nishio, M. & Niimi, S. Vocal acoustic characteristics of patients with parkinson's disease. *Folia Phoniatrica et logopaedica* **63**, 223–230 (2011).
33. Rusz, J. *et al.* Evaluation of speech impairment in early stages of parkinson's disease: a prospective study with the role of pharmacotherapy. *Journal of Neural Transmission* **120**, 319–329 (2013).
34. J. Holmes, R., M. Oates, J., J. Phylard, D. & J. Hughes, A. Voice characteristics in the progression of parkinson's disease. *International Journal of Language & Communication Disorders* **35**, 407–418 (2000).
35. Ngo, Q. C. *et al.* Computerized analysis of speech and voice for parkinson's disease: A systematic review. *Computer Methods and Programs in Biomedicine* 107133 (2022).
36. Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J. & Ramig, L. O. Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease. *IEEE Transactions on Biomedical Engineering* **59**, 1264–1271 (2012).
37. Hawi, S. *et al.* Automatic parkinson's disease detection based on the combination of long-term acoustic features and mel frequency cepstral coefficients (mfcc). *Biomedical Signal Processing and Control* **78**, 104013 (2022).
38. Vásquez-Correa, J. C. *et al.* Multimodal assessment of parkinson's disease: a deep learning approach. *IEEE Journal of Biomedical and Health Informatics* **23**, 1618–1630 (2018).
39. Arias-Londoño, J. D. & Gómez-García, J. A. Predicting updrs scores in parkinson's disease using voice signals: A deep learning/transfer-learning-based approach. In *Automatic Assessment of Parkinsonian Speech Workshop, Cambridge, Massachusetts, USA*, 100–123 (Springer, 2020).
40. Fujita, T., Luo, Z., Quan, C., Mori, K. & Cao, S. Performance evaluation of rnn with hyperbolic secant in gate structure through application of parkinson's disease detection. *Applied Sciences* **11**, 4361 (2021).
41. Dimauro, G., Di Nicola, V., Bevilacqua, V., Caivano, D. & Girardi, F. Assessment of speech intelligibility in parkinson's disease using a speech-to-text system. *IEEE Access* **5**, 22199–22208 (2017).
42. Orozco-Arroyave, J. R., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Gonzalez-Rátiva, M. C. & Nöth, E. New spanish speech corpus database for the analysis of people suffering from parkinson's disease. In *LREC*, 342–347 (2014).
43. Bot, B. M. *et al.* The mpower study, parkinson disease mobile data collected using researchkit. *Scientific data* **3**, 1–9 (2016).
44. Jaeger, H., Trivedi, D. & Stadtschnitzer, M. Mobile device voice recordings at king's college london (mdvr-kcl) from both early and advanced parkinson's disease patients and healthy controls. *Zenodo* (2019).
45. Moro-Velázquez, L. *et al.* Use of acoustic landmarks and gmm-ubm blend in the automatic detection of parkinson's disease. In *Proceedings of the Models and Analysis of Vocal Emissions for Biomedical Applications, Firenze, Italy*, 13–15 (2017).
46. Moro-Velázquez, L. *et al.* Study of the automatic detection of parkinson's disease based on speaker recognition technologies and allophonic distillation. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, 1404–1407 (IEEE, 2018).
47. Moro-Velázquez, L., Gómez-García, J. A., Dehak, N. & Godino-Llorente, J. I. Analysis of phonatory features for the automatic detection of parkinson's disease in two different corpora. *Proceedings of the Models and Analysis of Vocal Emissions for Biomedical Applications, Firenze, Italy* 33 (2019).
48. Moro-Velázquez, L. *et al.* A forced gaussians based methodology for the differential evaluation of parkinson's disease by means of speech processing. *Biomedical Signal Processing and Control* **48**, 205–220 (2019).
49. Moro-Velázquez, L. *et al.* Phonetic relevance and phonemic grouping of speech in the automatic detection of parkinson's disease. *Scientific Reports* **9**, 19066 (2019).
50. Godino-Llorente, J. *et al.* Approaches to evaluate parkinsonian speech using artificial models. In *Automatic Assessment of Parkinsonian Speech Workshop, Cambridge, Massachusetts, USA*, 77–99 (Springer, 2020).
51. Ibarra, E. J., Arias-Londoño, J. D., Zañartu, M. & Godino-Llorente, J. I. Towards a corpus (and language)-independent screening of parkinson's disease from voice and speech through domain adaptation. *Bioengineering* **10**, 1316 (2023).
52. Guerrero-López, A., Arias-Londoño, J. D., Shattuck-Hufnagel, S. & Godino-Llorente, J. I. MARTA: a model for the automatic phonemic grouping of the parkinsonian speech. *Authorea Preprints* (2024).
53. Hirano, M. & McCormick, K. R. Clinical examination of voice by minoru hirano (1986).
54. on Rating Scales for Parkinson's Disease”, M. T. F. The unified parkinson's disease rating scale (updrs): status and recommendations. *Movement Disorders* **18**, 738–750 (2003).
55. Hoehn, M. M. & Yahr, M. D. Parkinsonism: onset, progression, and mortality. *Neurology* **17**, 427–427 (1967).
56. Hoodin, R. B. & Gilbert, H. R. Parkinsonian dysarthria: An aerodynamic and perceptual description of velopharyngeal closure for speech. *Folia Phoniatrica et Logopaedica* **41**, 249–258 (1989).
57. Godino-Llorente, J. I., Saenz-Lechon, N., Osma-Ruiz, V., Aguilera-Navarro, S. & Gómez-Vilda, P. An integrated tool for the diagnosis of voice disorders. *Medical Engineering & Physics* **28**, 276–289 (2006).
58. Gómez-García, J., Moro-Velázquez, L., Arias-Londoño, J. D. & Godino-Llorente, J. I. On the design of automatic voice condition analysis systems. part iii: Review of acoustic modelling strategies. *Biomedical Signal Processing and Control* **66**, 102049 (2021).
59. Bhuta, T., Patrick, L. & Garnett, J. D. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *Journal of Voice* **18**, 299–304 (2004).
60. Karnell, M. P. *et al.* Reliability of clinician-based (grbas and cape-v) and patient-based (v-rqol and ipvi) documentation of voice disorders. *Journal of Voice* **21**, 576–590 (2007).
61. Kojima, T. *et al.* Objective assessment of pathological voice using artificial intelligence based on the grbas scale. *Journal of Voice* (2021).
62. Isshiki, N., Okamura, H., Tanabe, M. & Morimoto, M. Differential diagnosis of hoarseness. *Folia Phoniatrica et Logopaedica* **21**, 9–19 (1969).
63. Nemr, K. *et al.* Grbas and cape-v scales: high reliability and consensus when applied at different times. *Journal of Voice* **26**, 812–e17 (2012).
64. Mendes-Laureano, J. *et al.* NeuroVoz: a Castilian Spanish corpus of parkinsonian speech (v.3.0.0), <https://doi.org/10.5281/zenodo.10777656> (2024).

Acknowledgements

This work was funded by the Ministry of Economy and Competitiveness of Spain (grants TEC-2012-38630-C04-01, DPI2017-83405-R1, PID2021-128469OB-I00 and TED2021-131688B-I00), Ministry of Health (FIS PI21/00771), and by Comunidad de Madrid, Spain. UPM also supports Julián D. Arias-Londoño through a María Zambrano UP2021-035 grant funded by European Union-NextGenerationEU. The authors thank the Neurology and Otorhinolaryngology Services of HGUGM, Madrid, and HUF, Madrid, for the facilities provided, and UPM for providing computing resources on the Magerit Supercomputer. Finally, the authors also thank the Madrid ELLIS unit (European Laboratory for Learning & Intelligent Systems) for its indirect support, and to all patients who selflessly participated in the study.

Author contributions

J.G.G. and J.I.G. designed the experiment. J.G.G., J.D.A. and J.I.G. designed the methodology. J.M.L. and E.L.B. validated the experiment. J.M.L. and E.L.B. collected the data. J.M.L., E.L.B. and F.G.P. provided the resources. J.G.G., A.G.L. and J.I.G. wrote the initial draft version. A.G.L., J.D.A. and J.I.G. reviewed and edited the manuscript. A.G.L. provided the software to analyse the data. J.D.A. and J.I.G. supervised. J.I.G. and F.G.P. administrated the project. J.I.G. acquired the funding. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.I.G.-L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024