frontiers | Frontiers in Human Neuroscience

TYPE Original Research
PUBLISHED 28 April 2025
DOI 10.3389/fnhum.2025.1566274

OPEN ACCESS

EDITED BY
Mirko Grimaldi,
University of Salento, Italy

REVIEWED BY
Melissa Annette Redford,
University of Oregon, United States
Neda Fatima,
Manav Rachna International Institute of Research and Studies (MRIIRS), India

*CORRESPONDENCE
Fredrik Nylén
✉ fredrik.nylen@umu.se

RECEIVED 24 January 2025
ACCEPTED 04 April 2025
PUBLISHED 28 April 2025

CITATION
Nylén F (2025) An acoustic model of speech dysprosody in patients with Parkinson's disease. *Front. Hum. Neurosci.* 19:1566274. doi: 10.3389/fnhum.2025.1566274

An acoustic model of speech dysprosody in patients with Parkinson's disease

Fredrik Nylén*

Department of Clinical Science, Faculty of Medicine, Umeå University, Umeå, Västerbotten, Sweden

Purpose: This study aimed to determine the acoustic properties most indicative of dysprosody severity in patients with Parkinson's disease using an automated acoustic assessment procedure.

Method: A total of 108 read speech recordings of 68 speakers with PD (45 male, 23 female, aged 65.0 ± 9.8 *years*) were made with active levodopa treatment. A total of 40 of the patients were additionally recorded without levodopa treatment to increase the range of dysprosody severity in the sample. Four human clinical experts independently assessed the patients' recordings in terms of dysprosody severity. Separately, a speech processing pipeline extracted the acoustic properties of prosodic relevance from automatically identified portions of speech used as utterance proxies. Five machine learning models were trained on 75% of speech portions and the perceptual evaluations of the speaker's dysprosody severity in a 10-fold cross-validation procedure. They were evaluated regarding their ability to predict the perceptual assessments of recordings excluded during training. The models' performances were assessed by their ability to accurately predict clinical experts' dysprosody severity assessments.

Results: The acoustic predictors of importance spanned several acoustic domains of prosodic relevance, with the variability in f_O change between intonational turning points and the average first Mel-frequency cepstral coefficient at these points being the two top predictors. While predominant in the literature, variability in utterance-wide f_O was *f*ound to be only the fifth strongest predictor.

Conclusion: Human expert raters' assessments of dysprosody can be approximated by the automated procedure, affording application in clinical settings where an experienced expert is unavailable. Variability in pitch does not adequately describe the level of dysprosody due to Parkinson's disease.

KEYWORDS

automatic acoustic assessment, dysprosody, Parkinson's disease, dysarthria, prosody

1 Introduction

Dysprosody is a well-attested symptom of Parkinson's disease (Schlenck et al., 1993) and is discussed in the literature as an "impaired melody of speech", speaking monotony in pitch or loudness ("monopitch" and "monoloudness", respectively), "hypophonia", or an "altered rate of speech" (Sidtis et al., 2006). Dysprosody is an early-onset symptom of the disease (Schlenck et al., 1993) and a prominent factor behind reduced speech intelligibility (Watson and Schlauch, 2008; Klopfenstein, 2009; Feenaughty et al., 2014; Martens et al., 2015) and communicative efficiency (Martens et al., 2011). Dysprosody is most often discussed in connection with dysarthrias and predominately

in connection with Parkinson's disease (PD) and Huntington's disease (Rusz et al., 2014). Effects on expressive dysprosody have, however, also been observed following lesions in the caudate nucleus, the globus pallidus, and the putamen (Sidtis and Sidtis, 2003), in case reports of left hemiparesis and right hemisphere tumors (Sidtis, 1984), and in ∼2.7% of patients with epileptic seizures (Peters et al., 2011). When occurring as a component of apraxia of speech (Ballard et al., 2016), symptoms have been observed to be alleviated by neurobehavioral treatment (Ballard et al., 2010). There is, therefore, a great need to further our understanding of dysprosody-causing symptoms of diseases and develop a robust assessment method to guide diagnosis and management of treatments across several neurological conditions. However, while widely attested and often discussed in reports of speech effects of PD and other neurological diseases, there is currently no objective measure of dysprosody by which the impact of treatment may be assessed (Steurer et al., 2022).

One barrier to developing acoustic assessment methods for dysprosody originates in the complex nature of prosody itself (Terken and Hermes, 2000; Sidtis and Sidtis, 2003; Ladd and Arvaniti, 2022). A recent review by Fumel et al. (2024) highlighted nine aspects (f_o, the variability in f_o, intensity, intensity variability, speech rate, articulation rate, pause duration, and proportion of pauses during speaking) that have been the focus of previous research on dysprosody. Utterance-wide variability in f_o is the most often used proxy measure for dysprosody, in which reduced f_o excursions have been a consistent finding in patients with Parkinson's disease compared to control speakers (Bocklet et al., 2011; MacPherson et al., 2011; Skodda, 2011; Feenaughty et al., 2014; Thies et al., 2019; Frota et al., 2021) across many languages (Fumel et al., 2024). As noted by Fumel et al. (2024), however, variability in f_o does not afford reliable interpretation in terms of dysprosody or severity since modulation of f_o is also linked with the perception of liveliness, emotional expressions, or speaker gender (Traunmüller and Eriksson, 1995; Avery and Liss, 1996; Martinho et al., 2024; Nylén et al., 2024). Dysprosody as a term should, however, be used only to describe specifically the effects that may reduce speech intelligibility, and other indexical properties should not be considered in the assessment (Fumel et al., 2024). Utterance-wide variation in speech intensity is the second most common proxy measure of dysprosody, for which Fumel et al. noted some evidence of a systematic reduction in patients with PD compared to control speakers in their review. Still, the effect was less consistent than the corresponding effect on f_o variability. Some dopaminergic treatments, e.g., levodopa administration or deep brain stimulation (DBS) in the subthalamic nucleus (STN), may alleviate the adverse effects of PD on f_o variability (Lundgren et al., 2011; Skodda, 2011; Karlsson et al., 2013; Thies et al., 2024). Possibly, good alleviation may require elevated treatment levels to have beneficial effects (Bobin et al., 2024). The beneficial effects are, however, not universally observed across dopaminergic treatments. DBS in the posterior subthalamic area may, in contrast, have no beneficial effect on PD speakers' ability to modulate f_o or intensity on the global level (Lundgren et al., 2011; Karlsson et al., 2013) or speech intelligibility (Johansson et al., 2014; Sandström et al., 2015).

Dysprosody can also manifest in speech rhythm through speech rate and articulation effects. According to the review by Fumel et al. (2024), these effects are more negligible and are less systematically observed across languages. As noted by Liss et al. (2009), speech rhythm deviation can be used to correctly classify speakers into dysarthria types with 80% prediction accuracy when quantified using manually annotated syllable nuclei and onset and coda component relationships (Liss et al., 2009). Retained control over articulation rate and consonant/vowel relationships in speech motor tasks can be used to correctly identify speakers with PD among controls (Karlsson and Hartelius, 2019) and to predict dysarthria severity (Karlsson et al., 2020), but increasing age of the speaker also affects these properties (Karlsson and Hartelius, 2021), which makes them challenging to use as markers of disease progression. A reduced speech rate has only been reliably observed in American English and Dutch (Fumel et al., 2024). The frequency and length of pauses in speech were a much more systematic observation separating PD speakers from healthy controls in the meta-analysis by Fumel et al. than speech or articulation rate effects. Dopaminergic treatments using levodopa or DBS in the STN have been observed to alleviate, but not entirely extinguish, the speech and articulation rate effects (Ho et al., 2008; Karlsson et al., 2011; Knowles et al., 2024) and can be further amplified by DBS stimulation that is adjusted in real-time in response to bioelectrical signals from the patient (Cernera et al., 2024).

While observed with reasonable consistency, utterance-wide reductions in acoustic expressiveness due to PD, the effects are not large enough or observed with sufficient consistency for assessing dysprosody severity. Well-functional prosody is a well-explored field of linguistics, and analytical techniques have been used to provide a more detailed, time-aligned view of how PD affects speakers' prosodic ability. The autosegmental-metrical (AM) analysis, in which the realization and temporal alignment of language-specific intonational units (tones) and the strength of breaks (pauses) are categorized, has, in case reports, been used to observe a reduced concentration of prosodic tonal events due to PD. In contrast, the repertoire of tones used in communication has been observed to remain unaltered (Lowit et al., 2014). The AM framework has substantial descriptive value but does not provide direct insight that can be transferred to a measure of dysprosody in the speaker. Frota et al. (2021), however, recently proposed an extension to AM for Portuguese (P-ToBI) that deduces a prosodic index from the difference between the pitch accents and breaks that are produced by the PD patient compared to what would be expected in unimpaired speech. Thies et al. provided supporting evidence for this approach by showing that f_o peak in the syllable nuclei (the vowel) is lowered by PD (Thies et al., 2019). Manual multi-tier annotation of utterance, pitch accents, and break indices before analysis (Thies et al., 2019; Frota et al., 2021). The most reliable models of dysprosody severity due to dysarthria have shown an accuracy of 62.2–73.9%, depending on the model type, when trained on a set of intonation (f_o) and rhythmic properties extracted after the manual annotation of the utterance (Hernandez et al., 2020). Automatic segmental and syllable annotation procedures have been proposed. Still, they are challenged by syllable boundaries, to which tonal events are aligned, being more readily perceived as a cognitive construct with varying definitions (Vitale et al., 2024) than units that can unequivocally be segmented in recordings of fluent speech (Warren et al., 1996;

Reetzke et al., 2021). Therefore, the manual work required to perform the analysis offers a barrier to adoption beyond use in research for these analytical techniques. A less laborious work is required to find the stress pattern index (Tykalová et al., 2013), defined as $1 + ln(\frac{f_{o_{\max}}}{f_{o_{\min}}}) \sum E$) and the syllabic prosody index (Tavi and Werner, 2020) defined as $\frac{f_o\sqrt{d}}{\sqrt{E_{f<1kHz}}}$ (where f_o is the fundamental frequency, E the speech signal energy, and d is the duration) of words and syllables, respectively. These measures have, however, been evaluated in small samples of participants only (36 PD patients and controls) and only in terms of their ability to separate speaker groups. An evaluation of the affordance to accurately predict levels of dysprosody based on these metrics has not been attempted.

It should be observed that the properties discussed concerning dysprosody assessment in PD and other neurological diseases are only a subset of the features supporting appropriate prosody perception. Recent studies (Roessig et al., 2022; Arvaniti et al., 2024; Hu and Arvaniti, 2024) have highlighted additional cues that warrant renewed attention concerning the perception of prosodic entities. Vowels' spectral balance (spectral tilt) has long been attested to contribute to the perception of prominence in many languages (Sluijter and van Heuven, 1996; Heldner, 2003; Crosswhite, 2003) with a relative predictive strength rivaling the strongest cue (duration) (Sluijter and van Heuven, 1996; Heldner, 2003). Several proposals of how spectral tilt or spectral balance should be quantified have been proposed, and the Spectral Energy Ratio (SER) between a lower frequency band (0–1 kHz) and a higher frequency band (1–5 kHz) (Murphy et al., 2008) provides an intuitive base approximation of the tilt of the spectrum. However, the first Mel frequency cepstral coefficient (C1) and components of a first or sixth-order polynomial fitted to the logarithmic magnitude spectrum (SLF and SLF6D) have been shown to provide more robust quantifications (Kakouros et al., 2018). The level relation of the first and second, or first and third (Okobi, 2006), harmonic, both directly measured from the speech signal (L_2-L_1 and L_3-L_1) (Kakouros et al., 2018) and with correction (Iseli et al., 2007) for the effect of neighboring formants (corrected L_2-L_1 and L_3-L_1) (Hu and Arvaniti, 2024), have also been proposed to be potent cues.

A structure hinged on acoustic parameters is required to achieve an automatized framework for assessing dysprosody. Here, the outcome of observations from two different developments is fused to explore ways to circumvent the barrier to automatic assessment of dysprosody introduced by requirements for reliable syllable or segment isolation. First, it is observed that rhythmic structure can efficiently be extracted from the overall modulation of the speech signal (Leong et al., 2014) and that this information can be used to separate dysarthria types (LeGendre et al., 2009; Liss et al., 2010). While previous studies have explored envelope modulation to deduce rhythmic structure, it is proposed here that the timing of prosodically significant tonal events may provide an indirect cue to rhythmic properties by hinting at the approximate timing of the associated syllable.

Second, some previous studies have been directed toward stylizing a computed f_o curve into a more efficient and appropriate representation of the intonational structure, with microvariations removed. Taylor (1994) applied a *rise*, *fall*, and *connection element* classification scheme to a two-step median smoothed f_o and associated assigned elements with time and amplitude scaling factors (parameters) to deduce a representation of intonation that could replace the AM annotation in an automatic procedure. The degree of change in f_o was observed to provide a cue to the presence of a pitch accent. The subsequent Tilt model (Taylor, 2000) expanded the analysis by associating a rise amplitude, rise duration, fall amplitude, and fall duration with each element to derive a representation that could relatively faithfully reproduce manual annotations in analyzing a smoothed f_o curve in synthesis (Taylor, 2000). The reliance on a precomputed f_o with manual adjustments made to the speaker is a disadvantage to this approach when attempting automatic modeling.

In an alternative approach *Modeling melody* (Momel) algorithm (Hirst, 2005) extracts, the f_o contour in a two-step procedure, where the first quartile (q_1) of the distribution of f_o values obtained using very wide search space (typically 60–750 Hz) is used to derive the f_o curve forming the basis for subsequent computations within the search space of $0.75q_1$ Hz to 1.5 octaves above q_1. This two-step process is proposed to reduce the need for age and sex adjustment of parameters when deriving the f_o curve. The f_o curve is then separated into one quadratic spline function representation aimed at capturing the macro prosodic representation and a similar micro prosodic representation, which is not considered further here. The target points (Momel target points, MTPs) in the macro prosodic quadratic spline function are defined as (time, frequency) points that define significant tonal events in the utterance. The International Transcription System for Intonation (INTSINT) establishes a series of annotations of an intonational curve into (T)op, (H)igher (local maximum), (U)ppstepped, (S)ame, (M)id, (D)ownstepped, (L)ower (local minimum), and (B)ottom level. A parameter *key* is obtained by stepwise search originating from the mean f_o, and a *range* parameter is obtained in the 0.5–2.5 octave range. After defining the predicted f_o as an MTP by their INTSINT label as $T = key\sqrt{2^{span}}$, $M = key$, $B = \frac{key}{\sqrt{2^{span}}}$. The local maximum/minimum levels (H and L) model an f_o at the (log scaled) midpoint between the preceding MTPs and the T and B levels, respectively. Similarly, the up and downstepped levels (U and D) represent a point a quarter of the log-scaled distance between the previous MTPs and the T and B levels, respectively (Hirst, 2011). See Figure 1 for an illustration of the Momel and INTSINT automatic annotation output. The Momel and INTSINT annotation procedures have been given a canonical computer implementation (Hirst, 2007) and have been applied to describe intonation in relation to temporal events in several languages (Hirst and di Cristo, 1998; Véronis et al., 1998; Hirst et al., 2007; Chentir, 2009; Hirst, 2013; Celeste and Reis, 2021). The procedure has further been shown to reproduce human perception of tonal events with high accuracy (Hirst, 2011). If perceptually reliable, the MTPs could also serve as hints to prosodically relevant syllables and the rhythmic structure of speech. As the MTPs are defined in time and frequency, it is possible to associate intensity and spectral tilt measures with a time window surrounding the MPT to provide an augmented representation of prosody. Recent developments in vocal activity detection (Yin et al., 2018; Bredin et al., 2020; Bullock et al., 2020; Cristia et al., 2021) further suggest that units of speech approximating utterances could be automatically extracted from a speech recording prior to prosodic

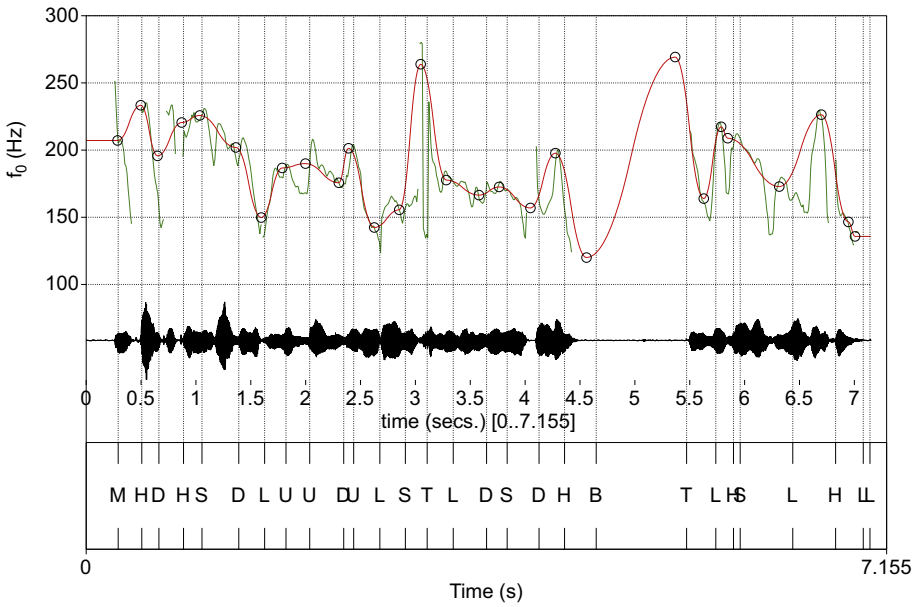

FIGURE 1
Example automatic INTSINT annotation (**bottom panel**) along with the audio waveform (**middle panel**) and computed f_o curve (**top panel, green**). The approximation of the macro-prosodic structure of the utterance estimated by the Momel algorithm is visualized is overlayed on the original f_o curve (**top panel, red**). Momel target points (MPTs) are marked with circles.

analysis in an unsupervised manner. When automatic assessments of all approximate utterances are combined, an assessment of dysprosody severity in the speaker may likely be deduced, and the combination of procedures into an analytical pipeline suggests a path toward a fully automated dysprosody assessment procedure.

The current study aimed to describe and evaluate a fully automated pipeline for the evaluation of dysprosody. Previous efforts to assess dysprosody have not had a solid base obtained from the perception of humans for evaluating the prediction accuracy of methods available to them, which is essential for a machine learning approach. Perceptual evaluations of dysprosody made by four speech and language pathologists with extensive experience in perceptual dysarthria assessment were used as ground truth for model training and evaluation. Four different machine learning model types with differing strengths and primary design aims were used to ensure that substantial contributions to our understanding of the perception of dysprosody could be represented regardless of the model's characteristics. Furthermore, the possibility of concentrating each model's strengths into an overall best-performing ensemble model was explored. The secondary aim was to describe the acoustic properties most indicative of increased dysprosody severity. A comparison of the predictive accuracy of the best-performing model with one with only one predictor, variability in f_o across an utterance, is also made.

2 Method

2.1 Perceptual evaluation of dysprosody

Audio recordings of 68 speakers with PD (45 male and 23 female) aged 65.0 ± 9.8 years, and with an average Hoehn & Yahr (H&Y) (Hoehn and Yahr, 1967) rating of 2.42 ± 0.57, and an average Unified Parkinson's Disease Rating Scale motor score [UPDRS part III (Goetz et al., 2008)] of 33 ± 12 were included in this study. The participants were asked to read an 89-word text, in which statements, interrogations, assertions, and instances of role changes are included and is the standard text used in dysarthria assessment in which dysprosody is an assessed component ["Ett svårt fall" ("A difficult case"); Eklund et al. (2014)]. The included recordings were made while on L-dopa medication. To increase the range of speech impairment levels in the study, recordings where the patient was off levodopa medication were also included when available ($n = 40$). Consequently, the total set of read speech recordings analyzed was 108. All recorded speeches were made in a quiet room at either a 48 kHz ($n = 79$) or 44.1 kHz ($n = 29$) sampling rate using either a Sennheiser HSp 4 and an MZA 900 P phantom adapter or a Marantz PMD 660 digital audio flash recorder.

Four expert speech-language pathologists (SLPs) with extensive (>20 years) experience in the assessment of dysarthria assessed all (108) readings of the standard text individually in four domains ("Articulation", "Voice", "Resonance", "Prosody", and "Overall impression"). A perceptual rating scale with four levels of deviant production ("No deviation", "Mild", "Moderate", or "Severe" deviation) was used in the perceptual assessment. As previously described (Karlsson et al., 2020), the moderate and severe categories were subsequently merged into a "Moderate to severe" category to support model training due to too few ratings of severe deviation. Spearman's correlations between the perceived level of deviation in prosody compared with other rated dimensions were strong for "Overall impression" ($r_s = 0.73$) and "Articulation" ($r_s = 0.65$), moderate for "Voice" ($r_s = 0.55$), and weak for "Resonance" ($r_s = 0.25$). An initial consensus training session in which four readings were rated and discussed for consensus was

TABLE 1 Descriptions of quantifications of each automatically extracted utterance.

Domain	Description of measure	Number of predictors per utterance
Time (s)	The time points relative to the start of the utterance associated with MTPs (in s)[y]	6
	The time distance from the previous MTPs (in s)	6
	The duration of the utterance (in s)	1
	The concentration of MTPs in an utterance (MTP/s)	1
	Duration of the utterance	1
f_o (Hz)	The Momel f_o associated with the MTPs[y]	6
	The change in Momel f_o from the previous MTPs[y]	6
	The f_o key of the utterance	1
	The range of the Momel f_o within the utterance	1
	The minimum of the Momel f_o within the utterance	1
	The maximum of the Momel f_o within the utterance	1
Intensity (dB)	The speech signal intensity associated with the MTPs[y]	6
	The change in speech signal intensity from the previous MTPs[y]	6
	The f_o key of the utterance	1
	The range of speech signal intensities within the utterance	1
	The minimum speech signal intensity within the utterance	1
	The maximum speech signal intensity within the utterance	1
Spectral tilt	The spectral energy ratio (in dB) between 0–1 kHz and 1–5 kHz at the MTPs	7
	The spectral energy ratio (in dB) between 0–1 kHz and 1–5 kHz from the previous MTPs[y]	7
	The L_2-L_1 (in dB) associated with the MTPs, both with and without correction[z] for the effect of nearby formants[y]	12
	The L_3-L_1 (in dB) associated with the MTPs, both with and without correction[z] for the effect of nearby formants[y]	12
	The change in L_2-L_1 (in dB) from the previous MTPs, both with and without correction[z] for the effect of nearby formants[y]	12
	The L_3-L_1 (in dB) associated with the MTPs[y], both with and without correction[z] for the effect of nearby formants	12
	The first Mel-frequency cepstral coefficient at the MTPs[y]	6
	The change in the first Mel-frequency cepstral coefficient from the previous MTPs[y]	6
	The slope of a first order polynomial to the short-term logarithmic magnitude spectrum at the MTPs[y]	6
	The change in slope of a first order polynomial to the short-term logarithmic magnitude spectrum from the previous MTPs[y]	6
	The coefficients of a sixth order polynomial to the short-term logarithmic magnitude spectrum at the MTPs[y]	36
	The change in coefficients of a sixth order polynomial to the short-term logarithmic magnitude spectrum from the previous MTPs[y]	36
Total number of predictors extracted per utterance		205

[y]Summary statistics extracted for each utterance: minimum, maximum, mean, standard deviation, coefficient of variability, and inter-quartal range.
[z]The correction was computed using the method presented by Iseli et al. (2006) and with formant bandwidths estimated using the method proposed by Hawks and Miller (1995).

performed before the perceptual assessment to strengthen inter-rater reliability. Laptops and Sennheiser HD 212Pro headsets were used in the perceptual evaluation.

2.2 Speech signal processing

The audio recordings were segmented into vocal activities approximating read speech sentences using overlap-aware speech detection (Bredin et al., 2020; Bullock et al., 2020). The identified portions of speech acts were then submitted to Momel & INTSINT processing, in which the f_o tracks (using a 10 ms analysis window), utterance f_o key, and f_o range were automatically identified, and MTPs were derived from the f_o track. INTSINT annotations were then assigned to each MPT on the macro-prosodic intonational structure. The entire utterance and each MPT were then provided with acoustic quantifications presented in Table 1 using the analysis procedure presented in Supplementary material A.

2.3 Machine learning

The ability of the quantifications of the prosody in the automatically extracted utterances to serve as predictors of human experts' ratings of dysprosody ("No deviation," "Mild," and "Moderate to Severe deviation") was evaluated in a cross-validation procedure. Five classification models with varying properties were selected for evaluation to explore their combined

use in support of the study's aims to (1) develop and evaluate a model and analysis pipeline that could facilitate an automatic assessment of dysprosody and (2) determine which acoustic properties provide the best support for classifying dysprosody severity. The polynomial support vector machine (SVM) model maximizes the distance between classes in a multidimensional space and has been used to identify both neurological diseases (Haq et al., 2019; Lahmiri and Shmuel, 2019; Arora and Tsanas, 2021) and other diseases based on voice samples (Vouzouneraki et al., 2024). While primarily considered for binary classification tasks, it has been extended to predict multiple classes and has been applied, for instance, to the prediction of vocal expression of emotion (Shahbakhi et al., 2014).

The penalized ordinal regression optimizes the error with a tuned balance between penalty terms based on the summed squares and the norms of the coefficients, and it has been previously used in models of detailed motor deterioration of speech performance due to Parkinson's disease (Karlsson and Hartelius, 2019; Karlsson et al., 2020). Elastic-net regularization of the ordinal regression was chosen here as it has been shown to perform well in speech data with multicollinearity among predictors (Tomaschek et al., 2018). The Random Forest is an ensemble model-building procedure in which multiple decision trees are trained on random subsets of predictors and training data. Combined to make a single prediction, they are well-documented to provide good prediction accuracy (Noroozi et al., 2017; Arora and Tsanas, 2021; Vouzouneraki et al., 2024). The k nearest neighbors is a non-parametric model-building technique that consider*s* proximity between samples and has been shown to perform very well in classification tasks for the speech of individuals with Parkinson's disease (Amato et al., 2021) and specifically for dysprosody (Majda-Zdancewicz et al., 2024) in a small sample of speakers with PD.

The five models were trained on a training data set consisting of acoustic predictors extracted from 75% of the recorded readings, matched with all expert raters' assessments of dysprosody in the patients' speech. The data were randomly divided into training (75%) and evaluation (25%) data sets, and the data were stratified to ensure a similar distribution of dysprosody severity in the two data sets.

The model parameters were tuned in a 10-fold cross-validation procedure. The models were optimized based on their ability to predict the perceptual assessments of utterances in the training data's 10th (holdout) fold. The model-tuning procedure used the mean logarithmic loss function to measure classification error. Highly correlated predictors (Spearman's $r > 0.9$; 43 predictors) were substituted for the predictor with the highest correlation with the outcome variable (the rating of dysprosody severity) before model training to produce better conditions for model training. The model tuning was performed using 1,000 parameter candidates for each hyperparameter (Table 2) that were spaced to maximize entropy in the distribution (Dupuy et al., 2015) with a variogram range of 2. The tuning procedure was repeated 10 times, each time with a different holdout portion of the data, and the final models were then constructed by averaging all 10 computed models of each type (support vector machines, penalized ordinal regression, k nearest neighbors, and Random Forest) to derive the final models.

TABLE 2 Hyper-parameters tuned for each model in the 10-fold cross-validation procedure and the maximum and minimum hyper-parameter ranges in the tuning grid.

Model name	Hyperparameter	Min	Max
Polynomial support vector machines	The cost of predicting a sample inside of or on the wrong side of the margin	9.96×10^{-4}	31.6
Penalized ordinal regression	The total amount of regularization	0.0	9.98×10^{-1}
	The proportion of L1 and L2 penalization	6.7×10^{-4}	9.99×10^{-1}
Random forest	The number of trees	1	2,000
	The number of predictors sampled at each split	1	74
	The minimal number of data points at a node required for node split	2	40
k Nearest Neighbors	The number of neighbors considered	1	15
	The kernel function is used for weighing differences	Rectangular. Triangular. Epanechnikov. Bi-weight. Tri-weight. Cosine. Inverse. Gaussian. and Rank	
	The parameter used for calculating distance		

Furthermore, the models were combined into an ensemble model by model stacking and by weighing the predictions of each model relative to its strengths and weaknesses in prediction within the training data.

The importance of each variable in the most accurate model was computed using the feature importance ranking measure (FIRM) (Zien et al., 2009) procedure, which has the attractive property that it generalizes to the sum of the squared change in model output and, therefore, has a transparent interpretation independent of the model type investigated. The model training used a substantial set of acoustic predictors, 205 in total (Table 1). To reduce the risk of reporting a highly specialized ability of models to predict the data they were trained on and ensure generalizability, all model evaluations were performed on 25% of the data that were not part of model training but were set aside for model evaluation.

The final models were evaluated for their accuracy in predicting human raters' assessments of dysprosody in 25% of utterances not included during the training of the models. Similarly, the agreement among human raters on the most common assessment of the reading (consensus rating) was calculated. The consensus rating was chosen over assessment based on the level of inter-rater agreement to enhance the robustness of dysprosody assessment by collaborating clinical colleagues. Both human raters and computational models were tested on recordings for which they had not been informed of the actual outcome. The human raters were assessed using the same classification metrics as the trained machine learning models.

TABLE 3 Agreement between the assessment of an individual rater, human expert, or acoustic model, and the true assessment of unobserved (testing) data.

Truth	Prediction	Human raters					Acoustic models					
		Rater 1	Rater 2	Rater 3	Rater 4	Across all raters	Support vector machines	Penalized ordinal regression	*k* nearest neighbors	Random forest	Ensemble	Model using variability in f_0 as the only predictor
No deviation	No deviation	**26**	**25**	**31**	**20**	**102**	**74**	**77**	**83**	**80**	**23**	**14**
	Mild deviation	5	7	0	11	23	35	31	27	29	87	0
	Moderate to severe deviation	0	0	0	1	1	1	2	0	1	0	96
Mild deviation	No deviation	0	4	1	2	7	34	30	28	26	4	21
	Mild deviation	**19**	**18**	**24**	**11**	**72**	**54**	**59**	**63**	**63**	**94**	**0**
	Moderate to severe deviation	2	2	5	7	16	11	10	8	10	1	78
Moderate to severe deviation	No deviation	0	0	1	0	1	8	9	4	3	2	4
	Mild deviation	2	0	2	0	4	16	14	22	20	26	0
	Moderate to severe deviation	**4**	**6**	**3**	**3**	**16**	**9**	**10**	**7**	**10**	**5**	**28**
	Sensitivity	0.80	0.84	0.77	0.73	0.78	0.50	0.53	0.53	0.56	0.44	0.33
	Specificity	0.92	0.88	0.94	0.81	0.89	0.76	0.78	0.79	0.80	0.72	0.65
	Positive predictive value	0.80	0.78	0.75	0.56	0.71	0.53	0.56	0.58	0.59	0.69	–
	Negative predictive value	0.91	0.88	0.93	0.79	0.88	0.76	0.78	0.80	0.80	0.78	0.67
	Balanced accuracy	0.86	0.86	0.85	0.77	0.83	0.63	0.66	0.66	0.68	0.58	0.49
	F-score	0.80	0.80	0.75	0.56	0.73	0.51	0.54	0.54	0.57	0.40	0.20

The majority vote was considered the true outcome in human experts' assessments.

10.3389/fnhum.2025.1566274

3 Results

The automatic extraction identified ~926 utterances from the 108 passage readings. The stratified sampling procedure aimed to create a comparable distribution of dysprosody severity levels across the training and test sets of utterances. A total of 684 utterances were assigned to the training set, while 242 were designated for the test set, which was used only at the evaluation stage.

The performance of machine learning models in predicting the assessments of trained clinical professionals is presented in Table 3, along with the agreement of each of the four professionals' assessments with a majority rating for the utterance. Table 4 summarizes the inter-rater agreement between pairs of raters. Human raters showed an average consensus (balanced accuracy in prediction) of 0.83 ± 0.04 (0.77–0.86) and an average *F* score of 0.73 ± 0.11 (0.56–0.80). The Support Vector Machines, Penalized ordinal regression, k Nearest Neighbors, and Random Forest models showed an average balanced accuracy in predicting unseen data of 0.66 ± 0.02 (0.63–0.68), with an average *F* score of 0.54 ± 0.02 (0.51–0.57). The best-performing model overall was the Random Forest model, with a balanced accuracy of 0.68 and an *F* score of 0.57; the Ensemble model training failed to produce a model that generalized well into the test data and showed performances that were lower than most original models, except for a strengthened positive predictive value of 0.69. The receiver operating characteristics (ROC) curves for the best-performing model (Random Forest), the model using the least number of predictors (penalized ordinal regression), and the model stack of all directly trained models (model ensemble) presented in Figure 2 indicate that the superior performance of the Random Forest model is achieved primarily by the model's ability to accurately predict cases rated as having "No deviation".

Figure 3 presents the FIRM (Zien et al., 2009) variable importance of the top 30 acoustic predictors in the best-performing Random Forest model. As a point of reference, an ordinal regression model in which variability in f_o was the only predictor of dysprosody severity showed a reduced balanced accuracy (0.49) and *F* score (0.20) compared to other models.

TABLE 4 Inter-rater agreement in dysprosody severity assessments among human raters.

Compared ratings	% Agreement	Cohen's K
Rater 1–Rater 2	69	0.47
Rater 1–Rater 3	62	0.35
Rater 1–Rater 4	44	0.18
Rater 2–Rater 3	66	0.40
Rater 2–Rater 4	46	0.22
Rater 3–Rater 4	43	0.15

4 Discussion

Prosody is the language function that organizes the speech stream into manageable chunks for the listener to process, and failure to meet listeners' expectations is linked with reduced speech intelligibility. Prosody is inherently multidimensional in how it is signaled to the listener, and previous models aimed to detect neurogenic dysprosody severity have achieved 62.2–73.9% detection accuracy by incorporating information from intonation, rhythm, and pausing, information that was acquired through a manual annotation procedure. The requirement of a laborious

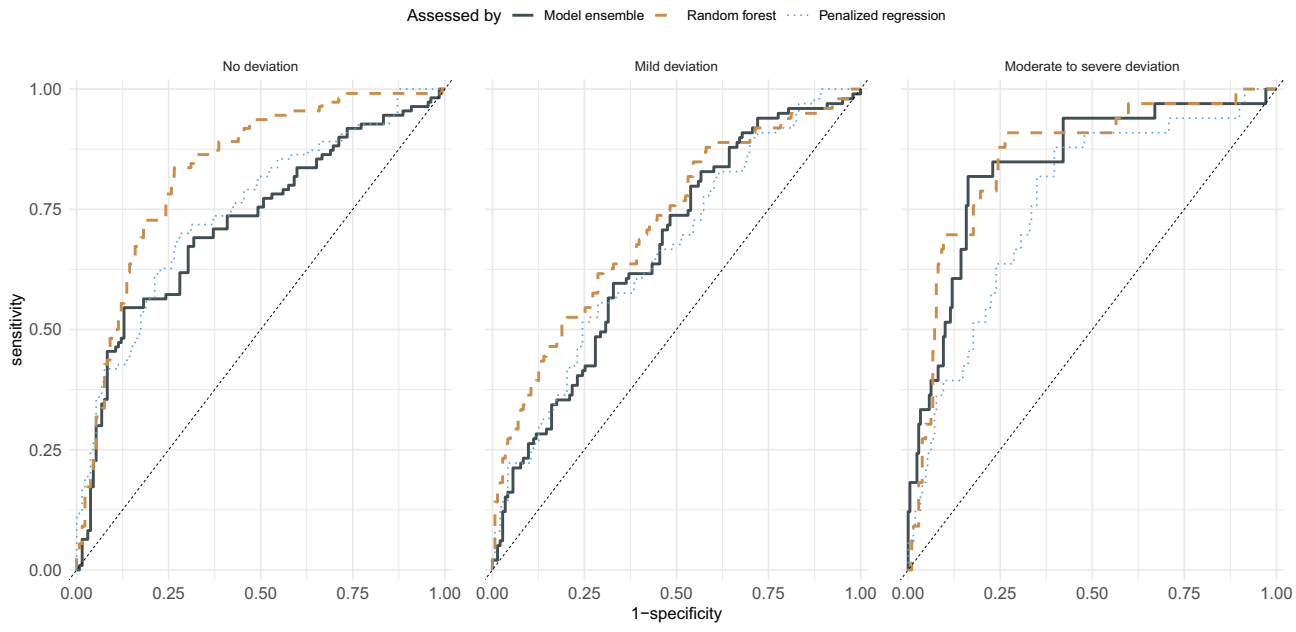

FIGURE 2
Receiver operating characteristic curves of random forest, the penalized ordinal regression, and ensemble models.

FIGURE 3
Variable importance of the 30 most important predictors extracted from MPT changes from one MPT to the next or across entire utterances. The importance of each variable was computed using the feature importance ranking measure (Zien et al., 2009) procedure. Notes: MTPs, Momel target points; LTAS, long-term average spectrum; C1, The first Mel-frequency cepstral coefficient; SLF, SLF6D, coefficients of a polynomial (with 1 and 6 components, respectively) fitted to the short-term logarithmic magnitude spectrum; MIN, MAX, MEAN, SD, and IQR indicate the summary statistic (minimum, maximum, mean, standard deviation, and interquartal range) applied to an utterance's quantifications to derive a single measure for each utterance. Measures described as "corrected" have spectral magnitude corrections (Iseli et al., 2007).

and time-consuming transcription task preceding assessment presents a clear barrier to the clinical adoption of the assessment procedure. In this study, an automatic dysprosody assessment pipeline for speech utterance identification and pitch contour preprocessing was constructed and provided with a comprehensive quantification aimed at capturing aspects established to be prosodically important from the speech signal of an utterance. The complete pipeline was then assessed regarding its proficiency in assessing the dysprosody severity of patients with Parkinson's disease based on a recording of speech patients' reading, with no prior pre-processing. Five models were trained on the individual assessments of levels of dysprosody severity made by four clinical raters with extensive experience in assessing dysarthria and evaluated in terms of their ability to predict the consensus

assessment of dysprosody severity among expert human raters on unobserved utterances.

The results suggest that the severity of dysprosody is not well described by single metrics, including the predominant proxy measure for dysprosody (utterance-wide variability in f_o). Simpler bases for classification tended to result in strongly biased predictions that do not reflect human experts' ratings well, and no acoustic predictor showed an influence on the classification that was strong enough to serve as a proxy in determining dysprosody severity. A model of dysprosody assessment based on the utterance-wide variability of f_o alone showed a strong preference for classifying most samples (84%) as having "Moderate to severe deviation". When using all predictors, the ensemble models Random Forest and penalized ordinal regression showed the best ability to identify utterances in the evaluation set that human experts had determined to have moderate to severe deviation in prosody. The support vector machine models failed to reach competitive levels of accuracy across all dysprosody severity levels. Overall, the Random Forest model achieved sensitivity, positive and negative predictive values, and an F score of predictions comparable to that of one rater (Rater 4) while not achieving similar sensitivity and balanced accuracy levels as the human rater. Overall, the acoustic models showed a lower sensitivity in their predictions than all human raters.

The result demonstrates that human clinical experts' assessment of dysprosody severity in Parkinson's disease can be partially modeled by a fully automatic speech processing pipeline in which utterances are automatically identified and that an intonation stylization can provide the scaffolding required for extracting acoustic cues. The developed models shed light on what constitutes a symptom of perceived dysprosody due to Parkinson's disease. While utterance-wide variability in f_o was not identified as a robust indication of the perceived level of dysprosody, the degree of variability, as well as the minimum of how much f_o changed from one MPT to the next, were identified as strong predictors. The modeling further highlighted that disregarding the first Mel spectrum coefficient and the level differences between the first and second, as well as the first and third harmonics, severely reduces the ability of an acoustic model to approximate human perception of dysprosody. The expert raters studied were not specifically experts in assessing dysprosody but were well-established experts in assessing dysarthria in a clinical setting, and the findings can, therefore, be transferred to a clinical setting.

Thus, previous reports in which dysprosody has been evaluated solely based on the proxy measure of the standard deviation of f_o are likely to have determined, in part, the level of liveliness (Traunmüller and Eriksson, 1995) in speech. Liveness is essential to our speech and likely contributes significantly to the experience of both parties in a communicative setting. However, utterance-wide variability in f_o alone does not ensure a retained linguistically functional intonation that adequately supports the transfer of information from the speaker to the listener. Instead, estimates of more local alterations in intonation, spectral balance, and intensity are used to distinguish portions of the speech signal of particular importance for the message from relatively less significant portions, providing a better model of clinical judgments of reduced prosodic functioning in patients with Parkinson's disease. Patients with Parkinson's disease have previously been observed to be reduced in their rapid regulation of phonation (Goberman et al., 2002; Goberman and Blomgren, 2008; Karlsson et al., 2012; Eklund et al., 2014; Tsuboi et al., 2014; Tanaka et al., 2015; Whitfield and Goberman, 2015), which may provide a partial explanation of the finding of less rapid local changes in f_o being significant predictors of clinically rated dysprosody specifically for patients with Parkinson's disease. While an explanation for the observations in terms of neurofunctional correlates cannot be offered to date, the connection with the subcortical structures, the globus pallidus, and the putamen (Sidtis and Sidtis, 2003) is congruent with an interpretation that failure to achieve tonal targets by persons with Parkinson's disease may be related to a failure to initiate an alteration of state in the phonatory musculature rather than an effect of muscular inability or fatigue or conflicting signaling in the direct, indirect, or hyper-direct pathways from the striatum to the cortex (Utter and Basso, 2008). This interpretation is, however, tentative and requires experimental support before being accepted.

Dysprosody is discussed here and in other parts of the literature as a single symptom. While discussed under a single term, dysprosody of a rated severity due to Parkinson's disease may differ from dysprosody caused by other neurological conditions (Sidtis, 1984). The automatic processing pipeline developed here does not presuppose a particular language or underlying disease causing dysprosody, but the relative importance of weights may likely be different for other diseases. Adjustments can, however, only be made with access to clinical raters with sufficient levels of experience and expertise. The procedure used in extracting acoustic parameters is made publicly available (Supplementary material A), and the procedures used for utterance segmentation and intonation modeling are widely available and well-documented (Hirst, 2007; Hirst et al., 2007; Origlia et al., 2013; Jadoul et al., 2018; Yin et al., 2018; Bredin et al., 2019; Bullock et al., 2020), which, when combined, removes any barrier to replication, language or disease estimates, adjustments in weights, and replication efforts in later research.

5 Conclusion

The perception of dysprosody can be approximated using an intonation stylization algorithm and an associated comprehensive acoustic assessment with no manually added temporal or tonal information. A performance in dysprosody assessment that approximates the abilities of clinical expert raters was achieved, which affords the transfer of a clinical assessment to remote situations where an experienced clinical expert is unavailable. The variation in pitch across an utterance, which is the most often used quantification of dysprosody in neurological disease, is not a reliable predictor of the level of dysprosody in patients with Parkinson's disease.

Data availability statement

All derived data supporting the conclusions of this article will be made available by the authors, without undue reservation. The speech recordings are sensitive personal identifiable information under national law and cannot be shared.

Ethics statement

The studies involving humans were approved by Regional Ethical Review Boards of Umeå (Case number 2012-368-31M) and Gothenburg (Case number 044-11). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

FN: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author declares that no financial support was received for the research and/or publication of this article.

Acknowledgments

The assistance of the clinical experts Ellika Schalling, Katja Laakso, Kerstin Johansson, and Lena Hartelius for performing the perceptual evaluations and the technical support of the Visible Speech (VISP) platform developed as part of the Swedish national research infrastructure Språkbanken and Swe-Clarin, funded jointly by the Swedish Research Council (2018–2028, contracts 2017-00626 and 2023-00161-16) and the 10 participating partner institutions, is gratefully acknowledged.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author declares that no Gen AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2025.1566274/full#supplementary-material

References

Amato, F., Borzì, L., Olmo, G., and Orozco-Arroyave, J. R. (2021). An algorithm for Parkinson's disease speech classification based on isolated words analysis. *Heal. Inf. Sci. Syst.* 9, 32. doi: 10.1007/s13755-021-00162-8

Arora, S., and Tsanas, A. (2021). Assessing Parkinson's disease at scale using telephone-recorded speech: insights from the Parkinson's voice initiative. *Diagnostics* 11, 1892. doi: 10.3390/diagnostics11101892

Arvaniti, A., Katsika, A., and Hu, N. (2024). Variability, overlap, and cue trading in intonation. *Language* 100, 265–307. doi: 10.1353/lan.2024.a929737

Avery, J. D., and Liss, J. M. (1996). Acoustic characteristics of less-masculine-sounding male speech. *J. Acoust. Soc. Am.* 99, 3738–3748. doi: 10.1121/1.414970

Ballard, K. J., Azizi, L., Duffy, J. R., McNeil, M. R., Halaki, M., O'Dwyer, N., et al. (2016). A predictive model for diagnosing stroke-related apraxia of speech. *Neuropsychologia* 81, 129–139. doi: 10.1016/j.neuropsychologia.2015.12.010

Ballard, K. J., Robin, D. A., McCabe, P., and McDonald, J. (2010). A treatment for dysprosody in childhood apraxia of speech. *J. Speech, Lang., Hear. Res.* 53, 1227–1245. doi: 10.1044/1092-4388(2010/09-0130)

Bobin, M., Sulzer, N., Bründler, G., Staib, M., Imbach, L. L., Stieglitz, L. H., et al. (2024). Direct subthalamic nucleus stimulation influences speech and voice quality in Parkinson's disease patients. *Brain Stimul.* 17, 112–124. doi: 10.1016/j.brs.2024.01.006

Bocklet, T., Nöth, E., Stemmer, G., Ruzickova, H., and Rusz, J. (2011). *Detection of Persons with Parkinson's Disease by Acoustic, Vocal, and Prosodic Analysis*. Piscataway: IEEE.

Bredin, H., Yin, R., Coria, J. M., Gelly, G., Korshunov, P., Lavechin, M., et al. (2019). *pyannote.audio: neural building blocks for speaker diarization*. Arxiv. doi: 10.1109/ICASSP40776.2020.9052974

Bredin, H., Yin, R., Coria, J. M., Gelly, G., Korshunov, P., Lavechin, M., et al. (2020). "Pyannote.Audio: neural building blocks for speaker diarization," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* Piscataway: IEEE), 7124–7128.

Bullock, L., Bredin, H., and Garcia-Perera, L. P. (2020). "Overlap-aware diarization: resegmentation using neural end-to-end overlapped speech detection," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Piscataway: IEEE), 7114–7118.

Celeste, L. C., and Reis, C. (2021). Formal intonative analysis: intsint applied to Portuguese. *J Speech Sci.* 2, 3–21. doi: 10.20396/joss.v2i2.15026

Cernera, S., Long, S., Kelberman, M., Hegland, K. W., Hicks, J., Smith-Hublou, M., et al. (2024). Responsive versus continuous deep brain stimulation for speech in essential tremor: a pilot study. *Mov. Disord.* 39, 1619–1623. doi: 10.1002/mds.29865

Chentir (2009). Extraction of Arabic standard micromelody. *J Comput Sci* 5, 86–89. doi: 10.3844/jcssp.2009.86.89

Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., et al. (2021). A thorough evaluation of the Language Environment Analysis (LENA) system. *Behav. Res. Methods* 53, 467–486. doi: 10.3758/s13428-020-01393-5

Crosswhite, K. (2003). "Spectral tilt as a cue to word stress in Polish, Macedonian, and Bulgarian," in *Proceedings of the 15th International Congress of Phonetic Sciences* (Barcelona: Causal Productions), 767–770.

Dupuy, D., Helbert, C., and Franco, J. (2015). DiceDesign and DiceEval : two R packages for design and analysis of computer experiments. *J. Stat. Softw.* 65, 1–38. doi: 10.18637/jss.v065.i11

Eklund, E., Qvist, J., Sandström, L., Viklund, F., van Doorn, J., and Karlsson, F. (2014). Perceived articulatory precision in patients with Parkinson's disease after

deep brain stimulation of subthalamic nucleus and caudal zona incerta. *Clin. Linguist. Phonet.* 29, 150–166. doi: 10.3109/02699206.2014.971192

Feenaughty, L., Tjaden, K., and Sussman, J. (2014). Relationship between acoustic measures and judgments of intelligibility in Parkinson's disease: a within-speaker approach. *Clin. Linguist. Phonet.* 28, 857–878. doi: 10.3109/02699206.2014.921839

Frota, S., Cruz, M., Cardoso, R., Guimarães, I., Ferreira, J. J., Pinto, S., et al. (2021). (Dys)Prosody in Parkinson's disease: effects of medication and disease duration on intonation and prosodic phrasing. *Brain Sci.* 11, 1100. doi: 10.3390/brainsci11081100

Fumel, J., Bahuaud, D., Weed, E., Fusaroli, R., and Basirat, A. (2024). A systematic review and bayesian meta-analysis of acoustic measures of prosody in Parkinson's disease. *J. Speech Lang. Hear. Res.* 67, 2548–2564. doi: 10.1044/2024_JSLHR-23-00588

Goberman, A. M., and Blomgren, M. (2008). Fundamental frequency change during offset and onset of voicing in individuals with parkinson disease. *J. Voice* 22, 178–191. doi: 10.1016/j.jvoice.2006.07.006

Goberman, A. M., Coelho, C., and Robb, M. (2002). Phonatory characteristics of Parkinsonian speech before and after morning medication: the ON and OFF states. *J. Commun. Disord.* 35, 217–239. doi: 10.1016/S0021-9924(01)00072-7

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martin, P. M., et al. (2008). Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement Disord.* 23, 2129–2170. doi: 10.1002/mds.22340

Haq, A. U., Li, J. P., Memon, M. H., khan, J., Malik, A., Ahmad, T., et al. (2019). Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings. *IEEE Access* 7, 37718–37734. doi: 10.1109/ACCESS.2019.2906350

Hawks, J. W., and Miller, J. D. (1995). A formant bandwidth estimation procedure for vowel synthesis. *J. Acoust. Soc. Am.* 97, 1343–1344. doi: 10.1121/1.412986

Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of *fo*cal accents in Swedish. *J. Phonet.* 31, 39–62. doi: 10.1016/S0095-4470(02)00071-2

Hernandez, A., Kim, S., and Chung, M. (2020). Prosody-based measures for automatic severity assessment of dysarthric speech. *Appl. Sci.* 10, 6999. doi: 10.3390/app10196999

Hirst, D. (2013). Melody metrics for prosodic typology: comparing English, French and Chinese. *Interspeech* 2013, 572–576. doi: 10.21437/Interspeech.2013-158

Hirst, D., Cho, H., Kim, S., and Yu, H. (2007). "Evaluating two versions of the momel pitch modelling algorithm on a corpus of read speech in Korean," in *INTERSPEECH*, 1649–1652.

Hirst, D., and di Cristo, A. (1998). *Intonation Systems: A Survey of Twenty Languages.* Cambridge University Press.

Hirst, D. J. (2005). Form and function in the representation of speech prosody. *Speech Commun.* 46, 334–347. doi: 10.1016/j.specom.2005.02.020

Hirst, D. J. (2007). "A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation," in *16th International Congress of Phonetic Sciences ICPhS XVI*, 1233–1236.

Hirst, D. J. (2011). The analysis by synthesis of speech melody. *J. Speech Sci.* 1, 55–83. doi: 10.20396/joss.v1i1.15011

Ho, A. K., Bradshaw, J. L., and Iansek, R. (2008). For better or worse: the effect of levodopa on speech in Parkinson's disease. *Mov. Disord.* 23, 574–580. doi: 10.1002/mds.21899

Hoehn, M. M., and Yahr, M. D. (1967). Parkinsonism onset, progression, and mortality. *Neurology* 17, 427–427. doi: 10.1212/WNL.17.5.427

Hu, N., and Arvaniti, A. (2024). Individual variability in the use of tonal and non-tonal cues in intonationa. *JASA Express Lett.* 4, 095203. doi: 10.1121/10.0028613

Iseli, M., Shue, Y-. L., and Alwan, A. (2006). "Age- and gender-dependent analysis of voice source characteristics," in *2006 IEEE International Conference on Acoustics, Speech and Signal Processing* 1, I-389–I-392. doi: 10.1109/ICASSP.2006.1660039

Iseli, M., Shue, Y-. L., and Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *J. Acoust. Soc. Am.* 121, 2283–2295. doi: 10.1121/1.2697522

Jadoul, Y., Thompson, B., and de Boer, B. (2018). Introducing Par*s*elmouth: a Python interface to Praat. *J. Phonetics*. 71, 1–15. doi: 10.1016/j.wocn.2018.07.001

Johansson, L., Möller, S., Olofsson, K., Linder, J., Nordh, E., Blomstedt, P., et al. (2014). Word-level intelligibility after caudal zona incerta stimulation for Parkinson's disease. *Acta Neurologica Scandinavica* 130, 27–33. doi: 10.1111/ane.12210

Kakouros, S., Räsänen, O., and Alku, P. (2018). Comparison of spectral tilt measures for sentence prominence in speech—Effects of dimensionality and adv*e*rse noise conditions. *Speech Commun.* 103, 11–26. doi: 10.1016/j.specom.2018.08.002

Karlsson, F., Blomstedt, P., Olofsson, K., Linder, J., Nordh, E., van Doorn, J., et al. (2012). Control of phonatory onset and offset in Parkinson patients following deep brain stimulation of the subthalamic nucleus and caudal zona incerta. *Parkinsonism Related Disord.* 18, 824–827. doi: 10.1016/j.parkreldis.2012.03.025

Karlsson, F., and Hartelius, L. (2019). How well does diadochokinetic task performance predict articulatory imprecision? Differentiating Individuals with Parkinson's Disease from Control Subjects. *Folia Phoniatrica et Logopaedica* 71, 251–260. doi: 10.1159/000498851

Karlsson, F., and Hartelius, L. (2021). On the Primary Influences of Age on Articulation and Phonation in Maximum Performance Tasks. *Languages* 6, 174. doi: 10.3390/languages6040174

Karlsson, F., Olofsson, K., Blomstedt, P., Linder, J., and van Doorn, J. (2013). Pitch variability in patients with Parkinson's disease: effects of deep brain stimulation of caudal zona incerta and subthalamic nucleus. *J. Speech Lang. Hear. Res.* 56, 1–9. doi: 10.1044/1092-4388(2012/11-0333)

Karlsson, F., Schalling, E., Laakso, K., Johansson, K., and Hartelius, L. (2020). Assessment of speech impairment in patients with Parkinson's disease from acoustic quantifications of oral diadochokinetic sequences. *J. Acoust. Soc. Am.* 147, 839–851. doi: 10.1121/10.0000581

Karlsson, F., Unger, E., Wahlgren, S., Blomstedt, P., Linder, J., Zafar, H., et al. (2011). Deep brain stimulation of caudal zona incerta and subthalamic nucleus in patients with Parkinson's disease: Effects on diadochokinetic rate. *Parkinson's Dis.* 2011, 1–10. doi: 10.4061/2011/605607

Klopfenstein, M. (2009). Interaction between prosody and intelligibility. *Int. J. Speech-Lang. Pa.* 11, 326–331. doi: 10.1080/17549500903003094

Knowles, T., Adams, S. G., and Jog, M. (2024). Effects of speech rate modifications on phonatory acoustic outcomes in Parkinson's disease. *Front. Hum. Neurosci.* 18, 1331816. doi: 10.3389/fnhum.2024.1331816

Ladd, D. R., and Arvaniti, A. (2022). Prosodic prominence across languages. *Annu. Rev. Linguistics* 9, 171–193. doi: 10.1146/annurev-linguistics-031120-101954

Lahmiri, S., and Shmuel, A. (2019). Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. *Biomed Signal Proces.* 49, 427–433. doi: 10.1016/j.bspc.2018.08.029

LeGendre, S. J., Liss, J. M., and Lotto, A. J. (2009). Discriminating dysarthria type and predicting intelligibility from amplitude modulation spectra. *J. Acoust. Soc. Am.* 125, 2530–2530. doi: 10.1121/1.4783544

Leong, V., Stone, M. A., Turner, R. E., and Goswami, U. (2014). A role for amplitude modulation phase relationships in speech rhythm perception. *J. Acoust. Soc. Am.* 136, 366–381. doi: 10.1121/1.4883366

Liss, J. M., LeGendre, S., and Lotto, A. J. (2010). Discriminating dysarthria type from envelope modulation spectra. *J. Speech, Lang., Hear. Res.* 53, 1246–1255. doi: 10.1044/1092-4388(2010/09-0121)

Liss, J. M. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., et al. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *J. Speech Lang. Hear. Res.* 52, 1334–1352. doi: 10.1044/1092-4388(2009/08-0208)

Lowit, A., Kuschmann, A., and Kavanagh, K. (2014). Phonological marke*r*s of sentence stress in ataxic dysarthria and their relationship to perceptual cues. *J. Commun. Disord.* 50, 8–18. doi: 10.1016/j.jcomdis.2014.03.002

Lundgren, S., Saeys, T., Karlsson, F., Olofsson, K., Blomstedt, P., Linder, J., et al. (2011). Deep brain stimulation of caudal zona incerta and subthalamic nucleus in patients with Parkinson´s disease: effects on voice intensity. *Parks. Dis.* 2011, 658956. doi: 10.4061/2011/658956

MacPherson, M. K., Huber, J. E., and Snow, D. P. (2011). The intonation-syntax interface in the speech of individuals with Parkinson's disease. *J. Speech Lang. Hear. Res.* 54, 19–32. doi: 10.1044/1092-4388(2010/09-0079)

Majda-Zdancewicz, E., Potulska-Chromik, A., Nojszewska, M., and Kostera-Pruszczyk, A. (2024). Speech signal analysis in patients with Parkinson's disease, taking into account phonation, articulation, and prosody of speech. *Appl. Sci.* 14, 11085. doi: 10.3390/app142311085

Martens, H., Nuffelen, G. V., Cras, P., Pickut, B., Letter, M. D., Bodt, M. S. D., et al. (2011). Assessment of prosodic communicative efficiency in Parkinson's disease as judged by professional listeners. *Parkinson's Dis.* 2011, 129310–10. doi: 10.4061/2011/129310

Martens, H., Nuffelen, G. V., Dekens, T., Huici, M. H-. D., Hernández-Díaz, H. A. K., Letter, M. D., et al. (2015). The effect of intensive speech rate and intonation therapy on intelligibility in Parkinson's disease. *J. Commun. Disord.* 58, 91–105. doi: 10.1016/j.jcomdis.2015.10.004

Martinho, D. H. d. a. C., Lopes, L. W., Dornelas, R., and Constantini, A. C. (2024). Can acoustic measurements predict gender perception in the voice? *PLoS ONE* 19, e0310794. doi: 10.1371/journal.pone.0310794

Murphy, P. J., McGuigan, K. G., Walsh, M., and Colreavy, M. (2008). Investigation of a glottal related harmonics-to-noise ratio and spectral tilt as indicato*r*s of glottal noise in synthesized and human voice signals. *J. Acoust. Soc. Am.* 123, 1642–1652. doi: 10.1121/1.2832651

Noroozi, F., Sapiński, T., Kamińska, D., and Anbarjafari, G. (2017). Vocal-based emotion recognition using random forests and decision tree. *Int. J. Speech Technol.* 20, 239–246. doi: 10.1007/s10772-017-9396-2

Nylén, F., Holmberg, J., and Södersten, M. (2024). Acoustic cues to femininity and masculinity in spontaneous speech. *J. Acoust. Soc. Am.* 155, 3090–3100. doi: 10.1121/10.0025932

Okobi, A. O. (2006). *Acoustic Correlates of Word Stress in American English*. Cambridge: Massachusetts Institute of Technology.

Origlia, A., Abete, G., and Cutugno, F. (2013). A dynamic tonal perception model for optimal pitch stylization. *Comput. Speech Lang.* 27, 190–208. doi: 10.1016/j.csl.2012.04.003

Peters, A. S., Rémi, J., Vollmar, C., Gonzalez-Victores, J. A., Cunha, J. P. S., Noachtar, S., et al. (2011). Dysprosody during epileptic seizures lateralizes to the nondominant hemisphere. *Neurology* 77, 1482–1486. doi: 10.1212/WNL.0b013e318232abae

Reetzke, R., Gnanateja, G. N., and Chandrasekaran, B. (2021). Neural tracking of the speech envelope is differentially modulated by attention and language experience. *Brain Lang.* 213, 104891. doi: 10.1016/j.bandl.2020.104891

Roessig, S., Winter, B., and Mücke, D. (2022). Tracing the phonetic space of prosodic focus marking. *Front. Artif. Intell.* 5, 842546. doi: 10.3389/frai.2022.842546

Rusz, J., Klempír, J., Tykalová, T., Baborová, E., Cmejla, R., RuŽička, E., et al. (2014). Characteristics and occurrence of speech impairment in Huntington's disease: possible influence of antipsychotic medication. *J. Neural Trans.* 121, 1529–1539.

Sandström, L., Hägglund, P., Johansson, L., Blomstedt, P., and Karlsson, F. (2015). Speech intelligibility in Parkinson's disease patients with zona incerta deep brain stimulation. *Brain Behav.* 5, e00394. doi: 10.1002/brb3.394

Schlenck, K-. J., Bettrich, R., and Willmes, K. (1993). Aspects of disturbed prosody in dysarthria. *Clin. Linguist. Phonet.* 7, 119–128. doi: 10.3109/02699209308985549

Shahbakhi, M., Far, D. T., and Tahami, E. (2014). Speech analysis for diagnosis of Parkinson's disease using genetic algorithm and support vector machine. *J. Biomed. Sci. Eng.* 2014, 147–156. doi: 10.4236/jbise.2014.74019

Sidtis, D. V. L., Pachana, N., Cummings, J. L., and Sidtis, J. J. (2006). Dysprosodic speech following basal ganglia insult: toward a conceptual framework for the study of the cerebral representation of prosody. *Brain Lang.* 97, 135–153. doi: 10.1016/j.bandl.2005.09.001

Sidtis, J. J. (1984). "Music, pitch perception, and the mechanisms of cortical hearing," in *Handbook of Cognitive Neuroscience* (Boston, MA: Springer), 91–114.

Sidtis, J. J., and Sidtis, D. V. L. (2003). A neurobehavioral approach to dysprosody. *Sem. Speech Lang.* 24, 93–105. doi: 10.1055/s-2003-38901

Skodda, S. (2011). Aspects of speech rate and regularity in Parkinson's disease. *J. Neurol. Sci.* 310, 231–236. doi: 10.1016/j.jns.2011.07.020

Sluijter, A. M. C., and van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *J. Acoust. Soc. Am.* 100, 2471–2485. doi: 10.1121/1.417955

Steurer, H., Schalling, E., Franzén, E., and Albrecht, F. (2022). Characterization of mild and moderate dysarthria in Parkinson's disease: behavioral measures and neural correlates. *Front. Aging Neurosci.* 14, 870998. doi: 10.3389/fnagi.2022.870998

Tanaka, Y., Tsuboi, T., Watanabe, H., Kajita, Y., Fujimoto, Y., Ohdake, R., et al. (2015). Voice features of Parkinson's disease patients with subthalamic nucleus deep brain stimulation. *J. Neurol.* 262, 1–9. doi: 10.1007/s00415-015-7681-z

Tavi, L., and Werner, S. (2020). A phonetic case study on prosodic variability in suicidal emergency calls. *Int. J. Speech Lang. Law* 27, 59–74. doi: 10.1558/ijsll.39667

Taylor, P. (1994). The rise/fall/connection model of intonation. *Speech Commun.* 15, 169–186. doi: 10.1016/0167-6393(94)90050-7

Taylor, P. (2000). Analysis and synthesis of intonation using the Tilt model. *J. Acoust. Soc. Am.* 107, 1697–1714. doi: 10.1121/1.428453

Terken, J., and Hermes, D. (2000). Prosody: theory and experiment, studies presented to gösta bruce. *Text, Speech Lang. Technol.* 14, 89–127. doi: 10.1007/978-94-015-9413-4_5

Thies, T., Barbe, M. T., and Mücke, D. (2024). Prosody matters: Preserved prominence marking strategies in people with Parkinson's disease independent of motor status. *PLoS ONE* 19, e0308655. doi: 10.1371/journal.pone.0308655

Thies, T., Mücke, D., Lowit, A., Kalbe, E., Steffen, J., Barbe, M. T., et al. (2019). Prominence marking in parkinsonian speech and its correlation with motor performance and cognitive abilities. *Neuropsychologia* 137, 107306. doi: 10.1016/j.neuropsychologia.2019.107306

Tomaschek, F., Hendrix, P., and Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *J. Phon.* 71, 249–267. doi: 10.1016/j.wocn.2018.09.004

Traunmüller, H., and Eriksson, A. (1995). The perceptual evaluation of F0 excursions in speech as evidenced in liveliness estimations. *J. Acoust. Soc. Am.* 97, 1905–1915. doi: 10.1121/1.412942

Tsuboi, T., Watanabe, H., Tanaka, Y., Ohdake, R., Yoneyama, N., Hara, K., et al. (2014). Distinct phenotypes of speech and voice disorders in Parkinson's disease after subthalamic nucleus deep brain stimulation. *J. Neurol. Neurosurg. Psychiat.* 86, jnnp-2014-308043. doi: 10.1136/jnnp-2014-308043

Tykalová, T., Rusz, J., Cmejla, R., Ruzickova, H., and Ruzicka, E. (2013). Acoustic Investigation of Stress Patterns in Parkinson's Disease. *J. Voice* 28, 129.e1–129.e8. doi: 10.1016/j.jvoice.2013.07.001

Utter, A. A., and Basso, M. A. (2008). The basal ganglia: an overview of circuits and function. *Neurosci. Biobehav. Rev.* 32, 333–342. doi: 10.1016/j.neubiorev.2006.11.003

Véronis, J., Cristo, P. D., Courtois, F., and Chaumette, C. (1998). A stochastic model of intonation for text-to-speech synthesis. *Speech Commun.* 26, 233–244. doi: 10.1016/S0167-6393(98)00063-6

Vitale, V. N., Cutugno, F., Origlia, A., and Coro, G. (2024). Exploring emergent syllables in end-to-end automatic speech recognizers through model explainability technique. *Neural Comput. Appl.* 36, 6875–6901. doi: 10.1007/s00521-024-09435-1

Vouzouneraki, K., Nylén, F., Holmberg, J., Olsson, T., Berinder, K., Höybye, C., et al. (2024). Digital voice analysis as a biomarker of acromegaly. *J. Clin. Endocrinol. Metab.* 110, 983–990. doi: 10.1530/endoabs.99.OC5.6

Warren, R. M., Healy, E. W., and Chalikia, M. H. (1996). The vowel-sequence illusion: intrasubject stability and intersubject agreement of syllabic forms. *J. Acoust. Soc. Am.* 100, 2452–2461. doi: 10.1121/1.417953

Watson, P. J., and Schlauch, R. S. (2008). The effect of fundamental frequency on the intelligibility of speech with flattened intonation contours. *Am. J. Speech Lang. Pathol.* 17, 348–355. doi: 10.1044/1058-0360(2008/07-0048)

Whitfield, J. A., and Goberman, A. M. (2015). The effect of Parkinson disease on voice onset time: temporal differences in voicing contrast. *J. Acoust. Soc. Am.* 137, 2432–2432. doi: 10.1121/1.4920874

Yin, R., Bredin, H., and Barras, C. (2018). Neural speech turn segmentation and affinity propagation for speaker diarization. *Interspeech* 2018, 1393–1397. doi: 10.21437/Interspeech.2018-1750

Zien, A., Krämer, N., Sonnenburg, S., and Rätsch, G. (2009). *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2009, Bled, Slovenia, September 7-11, 2009, Proceedings, Part II. Lect. Notes Comput. Sci.* (Berlin: Springer) 694–709.