

Article

Not peer-reviewed version

Automated Detection of Dysprosody in Patients with Parkinson's Disease

[Fredrik Nylén](#) *

Posted Date: 18 December 2023

doi: 10.20944/preprints202312.1270.v1

Keywords: Automatic acoustic assessment; Dysprosody; Parkinson's disease



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Automated Detection of Dysprosody in Patients with Parkinson's Disease

Fredrik Nylén

Department of clinical Sciences, Umeå University, SE91333 Umeå, Sweden; fredrik.nylen@umu.se

Abstract: Dysprosody is a commonly described feature of speech deterioration due to Parkinson's disease. Descriptions of the tonal movements underlying dysprosody have been attempted but has not afforded automation. The current study assessed a fully automated acoustic analysis pipeline in terms of its ability to predict human raters' perception of dysprosody in patients with Parkinson's disease. Read speech samples of 68 speakers with PD (45 male and 23 female) aged 65.0 ± 9.8 years were assessed by four human clinical experts in terms of dysprosody severity. The recordings were also submitted to a speech processing pipeline in which portions with speech were identified and provided with a parametrization of intonation. Five models were trained and tuned based on 75% of utterances and their perceptual evaluations in a 10-fold cross-validation procedure and evaluated in terms of their ability to predict the perceptual assessments of recordings not included in the training. The performances of models were compared to human raters' agreement with the majority vote on a speech sample. The results showed that human raters' assessments of dysprosody can be approximated by the automated procedure. Variability in pitch does not adequately describe the level of dysprosody due to Parkinson's disease.

Keywords: automatic acoustic assessment; dysprosody; Parkinson's disease

1. Introduction

Dysprosody is a well-attested symptom of Parkinson's disease [1] 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" [2]. Dysprosody is an early onset symptom of the disease [1] that is a prominent factor behind reduced speech intelligibility of patients with Parkinsonian dysarthria [3–6]. While most often discussed in connection with the dysarthrias, and predominately in connection with Parkinson's disease, effects on expressive dysprosody have also been observed following lesions in the caudate nucleus, the globus pallidus, and the putamen [7], in case reports of left hemiparesis and right hemisphere tumors [8], and in approximately 2.7% of patients with epileptic seizures [9]. When occurring as a component of apraxia of speech [10], it has been observed that symptoms may be alleviated by neurobehavioral treatment [11].

While widely attested and often discussed in reports of speech effects of Parkinson's disease and other neurological diseases [12] or of dopaminergic treatments [13,14] there is currently no clinically validated objective measure of dysprosody by which the effect of treatment may be assessed [15]. One barrier to developing acoustic assessment methods for dysprosody originates in the complex nature of prosody itself [7], and the assessment is therefore often approximated by a singular focus on variability in fundamental frequency (f_0) in utterances [6,13,16–19]; in Parkinsonian dysarthria, smaller f_0 excursions have been observed when speaking.

Well-functional prosody is, however, a well-explored field of linguistics, with analysis frameworks that may offer a path toward analysis of dysprosody resulting from neurological diseases as well. The autosegmental-metrical framework analyses the realization and temporal alignment of language-specific intonational units (tones) and the strength of breaks (pauses) in speech. The aim of the analysis is to describe the functional properties of the speech signal that conveys prosodic information to the listener. The autosegmental-metrical framework has been applied to describe aspects of dysprosody in patients with Parkinson's disease. In these analyses, it

has been observed that the frequency of pitch accents and boundary tones in intonation, and pauses, and length of utterances on the rhythmic level of dysprosody, may be altered by the disease while the inventory of tones used remains unchanged [20–22]. The autosegmental-metrical analysis is a manual process, but Frota et al. recently proposed a quantification of the autosegmental-metrical analysis framework for Portuguese (P-ToBI) that incorporated expected and realized pitch accents and breaks into an intonational index of prosodic deviance [16] by which a difference in prosodic phrasing and use of nuclear tones in persons with Parkinson's disease was indicated [16]. The analysis requires manual annotation of where the utterance occurred and pitch accents and break indices within the utterance by a specifically trained annotator. Further, the analysis system has not been demonstrated to be generalizable to languages other than Portuguese, and the tone and break indices approach to prosodic analysis is well recognized to require that the analysis system is adapted to fit the individual language [23]. The index has not been successfully related to the perceived level of dysprosody in the patients. Thus, autosegmental-metrical approaches currently do not offer a path to automated and language-independent evaluation of neurogenic dysprosody.

Alternative approaches are acoustically defined composite measures applied to easily identified portions of the speech signal and therefore require less laborious manual annotation. The Stress Pattern Index [24], defined as $1 + \ln \left(\frac{f_{0max}}{f_{0min}} \right) \sum E$ where f_0 is the fundamental frequency and E the speech signal energy, applies to words that should have prominence in an utterance to signal appropriate meaning. Alternatively, the Syllabic Prosody Index [25], defined as $f_0 \sqrt{d} / \sqrt{E_{f < 1kHz}}$ in which d is the duration of the syllable and the speech signal energy E is computed after low-pass filtering, applies to prominent syllables. Both indices have been verified to quantify aspects of dysprosody due to Parkinson's disease. The Syllabic Prosody Index was, for instance, found to be reduced compared to control speakers only for male speakers [24,26]. The indices have not been evaluated in terms of classification performance. In terms of classification, the most reliable models of dysprosody severity due to dysarthria have shown an accuracy of 62.2-73.9%, depending on model type, when trained on a set of intonation (f_0) and rhythmic properties [27]. The application of the models investigated by Hernandez et al. [27], the Stress Pattern, and Syllabic Prominence indices requires language-specific manual annotation on one [24,26] or multiple [27] tiers before application and may be considered unwieldy in clinical application.

A path to automated assessment of dysprosody severity may, however, be afforded by application of language-independent approaches to describe intonational contours. The International Transcription System for Intonation (INTSINT) re-encodes the surface intonational contour into language agnostic (T)op, (H)igher (local maximum), (U)ppstepped, (S)ame, (M)id, (D)ownstepped, (L)ower (local minimum), and (B)ottom levels defined relative to the f_0 key and f_0 range. Therefore, the original f_0 contour may be approximately reproduced given the INTSINT transcription labels and their locations in the utterance [28]. A companion f_0 contour simplification algorithm (Modeling melody, Momel) filters out micro-prosodic components of the contour by approximation using quadratic splines. With only the macro-prosodic intonational structure remaining the representation is refocused to include only intonational movements that are likely to serve a function in communication. See Hirst [29] for an overview. The macro-prosodic approximation of the intonational contour can then be used for the identification of INTSINT anchor (target) points in the utterance. Figure 1 exemplifies the result of a combined INTSINT and Momel analysis. The Momel and INTSINT annotation procedures have been automated, applied successfully to different languages [30–35], and given a canonical computer implementation [36]. An alternative f_0 stylization procedure is offered by ProsodyPro [37], which assumes a relatively constant segmental tier. This assumption was, however, considered not generally valid with regards to dysarthria [38] and this possible route to an automatic annotation was not pursued further. In contrast, the INTSINT / Momel analysis quantifies only the intonation contour and disregards the segmental information and pausing. The time marks identified by an INTSINT annotation provide reference points at which supplementary acoustic information could be computed directly from the sound. Whether an assessment procedure based on the result of an INTSINT / Momel analysis affords modeling of the level of dysprosody has, however, not been evaluated.

While the INTSINT and Momel procedures offer the opportunity to process intonation and other time-aligned acoustic information, the procedures presuppose that the utterance has already been identified and extracted manually. Recent developments in speaker diarization and vocal activity detection [39–44], however, offer the opportunity to preprocess speech recordings into vocal activities, approximating utterances, which can then be fed into an INTSINT / Momel analysis workflow in which the tonal movements are described in terms of timings and f_0 levels reached. The INTSINT / Momel analysis workflow provides a description of the micro and macro-prosodic intonation pattern of an utterance but not a summative quantification of prosody. The tonal levels provided after an INTSINT annotation can, however, be used as reference time points at which the Momel stylized f_0 and RMS amplitude of the speech signals can be extracted and given with a quantification that provides a prosodically relevant summation of the utterances.

The aim of the current study was to describe and evaluate an automated pipeline, with utterance identification and a novel quantification of intonation and speech intensity alterations, and to evaluate this pipeline in terms of its affordance for automatic assessment of dysprosody in patients with Parkinson's disease.

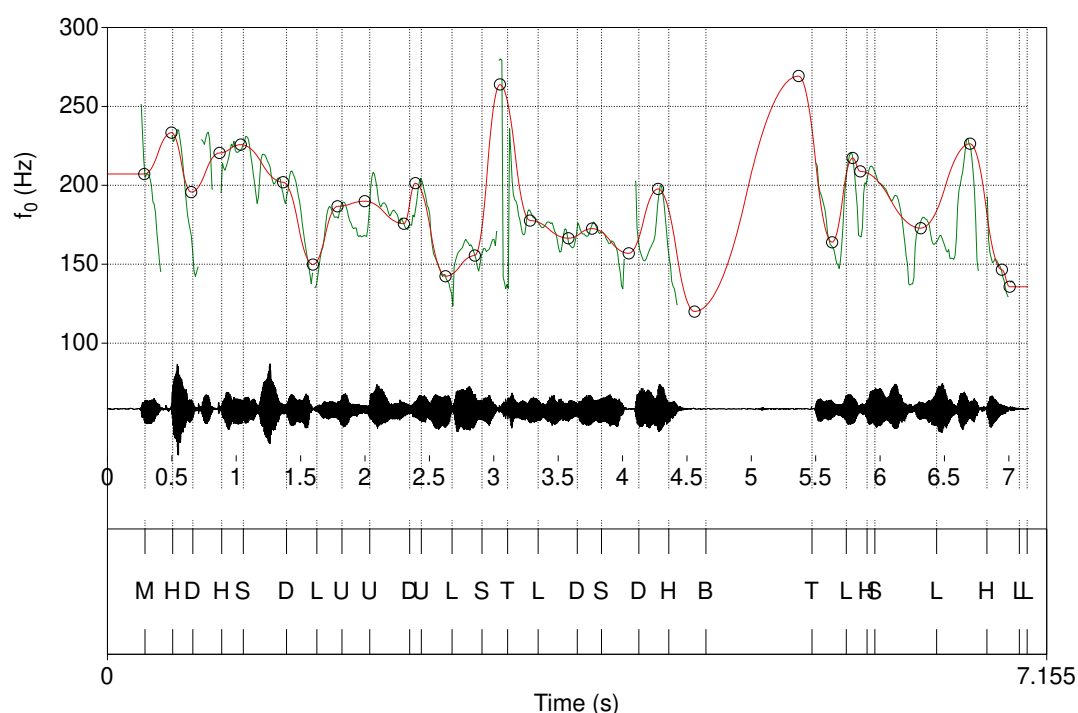


Figure 1. An example automatic INTSINT annotation (bottom panel) along with the audio waveform (middle panel) and computed f_0 curve (top panel, green). The approximation of the macro prosodic structure of the utterance estimated by the Momel algorithm is visualized is overlaid on the original f_0 curve (top panel, red).

2. Materials and Methods

2.1. Perceptual evaluation of dysprosody

Audio recordings of 68 speakers with PD (45 male and 23 female) aged 65.0 ± 9.8 yrs, 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, e.g., [45]) of 33 ± 12 were included in this study. The participants were asked to read a 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"]). The readings included were made either on or off L-dopa medication (1.5 h separation between sessions) to increase the range of speech impairment levels in the study. The recordings were made in a quiet

room in either 48 kHz (n=79) or 44.1 kHz (n=29) sampling rate using either a Sennheiser HSp 4 and a MZA 900 P phantom adapter or a Marantz PMD 660 digital audio flash recorder.

Four expert Speech-Language Pathologists (SLPs) with considerable (>20 years) experience in the assessment of dysarthria assessed all (108) readings of the standard text individually in four domains and focus of the current report is on the ratings of “Prosody”; the other domain assessments have been discussed in detail in a previous report [46]. All readings were assessed individually by the four raters, with individually randomized presentation order, with no time limitation, and the possibility to listen as many times as required. A rating scale of “No deviation,” “Mild,” “Moderate”, and “Severe” deviation was employed in the perceptual assessment. Similar to the procedure in [46] the moderate and severe categories were subsequently merged into a “Moderate to severe” category in order to support model training, due to too few ratings of severe deviation. An initial consensus training session in which four readings were rated and then discussed for consensus was performed prior to 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 using the pyannote framework [42,43]. The identified portions of speech acts were then submitted to an INTSINT / Momel processing pipeline in which the f_0 tracks, key, and f_0 range of the utterance were automatically identified. A 10ms analysis window was used when identifying f_0 . Subsequent steps used the computed f_0 track to derive INTSINT labels and compute their timing. Finally, Momel derived the macro-prosodic intonational structure associated with INTSINT annotations and identified their time points.

Following the speech segmentation and INTSINT / Momel preprocessing steps, quantifications designed to capture liveliness as well as the ability to provide linguistic function were applied to the INTSINT annotation time points and their associated f_0 levels, and an estimation of speech amplitude estimated as root mean squared (RMS) amplitude in the same time windows used when computing the f_0 contour. The measures were designed to capture micro-variability in f_0 and RMS amplitude of the speech signal in adjacent INTSINT tonal time points, as well as the macro-variability in the acoustic properties across an utterance. An overview of the measures computed as well as a description of them are presented in Table 1.

Table 1. Descriptions of quantifications applied to the INTSINT and Momel results.

| Domain | Measure name | Description |
|--------|-------------------|--|
| Time | Time diff MEAN | The mean time difference between consecutive INTSINT annotations (in ms) |
| | Time diff MEDIAN | The median time difference between consecutive INTSINT annotations (in ms) |
| | Time diff VAR | The variance in time differences between consecutive INTSINT annotations (in ms ²) |
| | Time diff MIN | The minimum time difference between consecutive INTSINT annotations (in ms) |
| | Time diff MAX | The maximum time difference between consecutive INTSINT annotations (in ms) |
| | INTSINT / s | The average INTSINT label concentration in the utterance |
| | Duration | The utterance duration |
| Pitch | f_0 diff MEAN | The mean difference in f_0 between consecutive INTSINT annotations (in Hz) |
| | f_0 diff MEDIAN | The median difference in f_0 between consecutive INTSINT annotations (in Hz) |

| | | |
|---------------------|--------------------|--|
| | f_0 diff VAR | The variance in f_0 differences between consecutive INTSINT annotations (in Hz) |
| | f_0 diff MIN | The minimum difference in f_0 between consecutive INTSINT annotations (in Hz) |
| | f_0 diff MAX | The maximum difference in f_0 between consecutive INTSINT annotations (in Hz) |
| | f_0 COV | The coefficient of variance in f_0 across the utterance |
| | f_0 diff VAR / s | The variance in f_0 differences between consecutive INTSINT annotations (in Hz), normalized by the duration (in seconds) of the utterance. |
| | f_0 key | The pitch key of the utterance |
| | f_0 range | The pitch range of the utterance |
| Amplitude | RMS COV | The coefficient of variance in RMS amplitude across the utterance |
| Intonational levels | Unique INTSINT | The number of distinct INTSINT labels encoded in the utterance |

2.3. Statistical modeling

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," "Moderate to Severe deviation") was evaluated in a cross-validation procedure. Five classification models were evaluated. The Naive Bayes is a supervised multiclass classifier based on the probability classification Bayesian theory [47] that has been shown to perform well in small sample speech classifications [48,49]. Decision trees are a supervised non-parametric learning method that derives a heretical set of simple decision rules that are applied to the predictors to derive a predicted class. Decision trees have received some use in the classification of speech or singing [50,51] but are predominately perceived as advantageous to explore as they can be clearly visualized and, therefore, easily accessible as input in clinical decision-making. The random forest procedure constructs multiple decision trees trained on random subsets of predictors and may provide enhanced prediction accuracy over simpler decision trees [52,53]. The polynomial support vector machine model maximizes the width of the distance between classes in a multidimensional space and has been used to identify both neurological diseases [53–55] and vocal expression of emotion [56,57]. Finally, elastic-net regularized ordinal regression optimizes the error as well as a tuned balance between penalty terms based on the summed squares and the norms of the coefficients and has been used previously in models of detailed motor deterioration of speech performance due to Parkinson's disease [46,58].

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' assessment of dysprosody in the patients' speech. The random division of the data into training (n=203, 75%) and evaluation (n=69, 25%) data sets was made in a stratified manner to ensure a similar distribution of levels of dysprosody in the two data sets. The model parameters were tuned in a 10-fold cross-validation procedure and the models were optimized based on their ability to predict the perceptual assessments of utterances in the 10th (holdout) fold of the training data. The mean logarithmic loss function was used as the measure for classification error in the model tuning procedure. The model tuning was performed using 1000 parameter candidates for each hyper-parameter (Table 2) that were spaced to maximize entropy in the distribution [59] with a variogram range of 2. The tuning procedure was repeated ten times, each time with a different holdout portion of the data, and the final models were then constructed by averaging all ten computed models of each type (Naive Bayes, Decision tree, Random forest, Support Vector Machine, and Penalized ordinal regression) to derive the final models.

Table 2. The hyper-parameters tuned for each model in the 10-fold cross-validation procedure.

| Model name | Hyperparameter |
|------------|----------------|
|------------|----------------|

| | |
|------------------------------|--|
| Naive Bayes | The relative smoothness of the class boundaries |
| Decision tree | The cost complexity Maximum tree depth |
| Random forest | The number of trees |
| Support Vector Machines | The cost of predicting a sample inside of or on the wrong side of the margin |
| Penalized Ordinal Regression | The total amount of regularization The proportion of L1 and L2 penalization |

The final models were assessed in terms of their accuracy in predicting human raters’ assessments of dysprosody in the 25% of utterances which was not included when training the models. The classification performances were assessed using the Balanced accuracy, F1-measure and the Mathew’s correlation coefficient (MCC) metrics, which was thought to provide an appropriate estimate of the performance despite the imbalanced nature of the classification task [60] and have multiclass extensions [61]. Similarly, human raters’ agreement with the most frequently occurring assessment of the reading (consensus rating) was computed. The consensus rating was selected over assessment based on the level of inter-rater agreement to allow for an increased robustness of dysprosody assessment by cooperating clinical colleagues. Both human raters and computational models were evaluated on recordings for which they had not been informed of the true outcome. The human raters were assessed using the same classification metrics as the trained statistical models.

3. Results

The performance of statistical 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 the majority rating for the utterance. Table 4 summarizes the inter-rater agreement between pairs of raters. The lowest prediction accuracy was indicated for Naive bayes (0.54), which also showed a very low correlation between prediction and outcome ($MCC < 0$). The human raters showed an average balanced agreement (accuracy in prediction) with a consensus of 0.80 ± 0.06 (0.70-0.84), an average MCC of 0.56 ± 0.10 (0.41-0.65), and an average F1 score of 0.73 ± 0.09 (0.60-0.80). The Decision tree, Support Vector Machine and Penalized ordinal regression models formed a natural group in performances that were slightly worse than the human raters, with an average accuracy in predicting the assessments of humans of 0.65 ± 0.02 (0.63-0.67), an average MCC of 0.21 ± 0.05 (0.16-0.25), and an average F1 score of 0.53 ± 0.01 (0.51-0.54). The Decision tree model, however, showed misclassifications of utterances with No deviation as Moderate to severe deviation to a comparable degree as the Naive bayes and Rater 4, but which is not observed for Support Vector Machines, Penalized ordinal regression, and Raters 1-3. The Random forest model showed prediction performance of unseen human assessments (accuracy=0.74, MCC=0.42, F1=0.65) that were comparable with one of the human raters’ agreement with the consensus (accuracy=0.70, MCC=0.41, F1=0.60).

Table 5 presents the summed, normalized, Variable importances of acoustic predictors across all final models computed by the Feature Importance Ranking Measure approach [62].

Table 3. The agreement between individual rater’s assessment and the group consensus and the performances of trained models in predicting human ratings in unobserved data.

| Dysprosody severity (majority rating) | Individual assessment | Expert assessment | | | | Model assessment | | | | |
|--|---------------------------------|-------------------|------------|------------|------------|------------------|------------------|------------------|---------------------------|---------------------------------|
| | | Rater 1 | Rater 2 | Rater 3 | Rater 4 | Naive bayes | Decision tree | Random forest | Support Vector Machine | Penalized ordinal regression |
| No deviation | No deviation | 42 | 46 | 45 | 25 | 24 | 2 | 23 | 21 | 23 |
| | Mild deviation | 11 | 7 | 8 | 22 | 2 | 28 | 7 | 8 | 7 |
| | Moderate to severe deviation | 0 | 0 | 0 | 6 | 4 | 0 | 0 | 1 | 0 |
| Mild deviation | No deviation | 7 | 7 | 5 | 2 | 25 | 2 | 12 | 12 | 17 |
| | Mild deviation | 36 | 35 | 38 | 19 | 1 | 26 | 14 | 14 | 9 |
| | Moderate to severe deviation | 2 | 3 | 2 | 24 | 2 | 0 | 2 | 2 | 4 |
| Moderate to severe deviation | No deviation | 1 | 1 | 0 | 0 | 6 | 0 | 1 | 6 | 5 |
| | Mild deviation | 2 | 1 | 3 | 0 | 3 | 11 | 5 | 5 | 3 |
| | Moderate to severe deviation | 7 | 8 | 7 | 10 | 2 | 0 | 5 | 0 | 4 |
| Metrics | Balanced accuracy | 0.84 | 0.86 | 0.88 | 0.62 | 0.54 | 0.65 | 0.74 | 0.65 | 0.63 |
| | MCC | 0.63 | 0.69 | 0.71 | 0.30 | -0.02 | 0.25 | 0.42 | 0.21 | 0.16 |
| | F1 score | 0.79 | 0.82 | 0.83 | 0.50 | 0.39 | 0.54 | 0.65 | 0.53 | 0.51 |

Note: The majority vote in human experts’ assessments were considered the true outcome, against which individual assessments were compared. The Individual assessments were made either by an individual human expert rater or by an assessment model.

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

| Compared ratings | % Agreement | Cohen’s κ |
|-------------------|-------------|------------------|
| 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 |

Table 5. Total variable importance across all models.

| Measure name | Variable importance |
|-----------------------------|---------------------|
| f ₀ diff MIN | 9.3 |
| f ₀ diff VAR | 7.6 |
| f ₀ diff MAX | 7.2 |
| f ₀ key | 6.1 |
| f ₀ COV | 4.8 |
| f ₀ diff VAR / s | 2.9 |
| Time diff MEAN | 2.4 |
| Time diff MEDIAN | 1.8 |
| f ₀ range | 0.9 |
| f ₀ diff MEDIAN | 0.6 |
| Time diff MAX | 0.6 |
| f ₀ diff MEAN | 0.3 |
| Duration | 0.1 |
| INTSINT / s | 0.1 |
| RMS COV | 0.01 |

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 a 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 by means of a manual annotation procedure. The requirement of a laborious and time-consuming transcription task preceding assessment presents a clear barrier to clinical adoption of the assessment procedure. In this study, an automatic dysprosody assessment pipeline was constructed from tools that were already available for speech utterance identification and pitch contour stylization, and provide a novel quantification aimed at capturing aspects of variability in pitch and intensity of speech. The complete pipeline was then assessed in terms of its proficiency in assessing 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 made by four clinical raters 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 severity of dysprosody is not well described by single metrics, including the predominant proxy measure for dysprosody (variability in f₀), or by simpler statistical models (Naïve bayes or Decision tree classifiers). Simpler bases for classification tended to result in a strongly biased prediction that does not reflect human experts’ ratings well, and no acoustic predictor showed an influence on the classification that was strong enough to be able to serve as a proxy in determining dysprosody severity. The ensemble model Random forest and the strong ordinal classifier model type penalized ordinal regression showed stronger ability to learn how to identify utterances in the

evaluation set that human experts had determined to have Moderate to severe deviation in prosody. The Support Vector Machines models failed to reach similar levels of accuracy, particularly in identifying severe levels of dysprosody. Overall, the Random forest achieved a higher level of accuracy in predicting the severity of dysprosody in unseen utterances than one of the expert human raters.

The result, therefore, demonstrates that dysprosody severity as perceived by human clinical experts in the assessment of speech in Parkinson's disease can be automatically deduced by a fully automatic speech processing pipeline in which utterances are automatically identified, and productive features for predicting dysprosody severity is identified using an established prosodic algorithm with a theoretical basis. It is therefore concluded that the developed pipeline constitutes a new development that can throw additional light onto what constitutes a symptom of perceived dysprosody due to Parkinson's disease. One may further observe in the results that the one acoustic feature that has to date predominately been investigated as a proxy to dysprosody, utterance-wide variability in f_0 , was not identified here as a robust predictor of the perceived level of dysprosody. Instead, the most productive predictors described local degrees of change in f_0 from the timing of one tonal level to the next; measures of utterance level variability in f_0 or RMS amplitude were less important. Thus, previous reports in which dysprosody has been evaluated solely based on the proxy measure of the standard deviation of f_0 are likely to have determined, in part, the level of liveliness [63] in speech. Liveness is essential and, in a communicative setting, likely contributes significantly to the experience of both parties. However, variability in f_0 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 of how intonation and intensity variation are used to distinguish portions of the speech signal of particular importance to the message from relatively less significant portions provided 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 [64–71] which may provide a partial explanation of the findings of significant predictors of clinically rated dysprosody. While an explanation for the converging observations in terms of neurofunctional correlates cannot be offered to date, the connection with the subcortical structures, the globus pallidus, and the putamen [7] 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 [72]. 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 [8]. The automatic processing pipeline does not presuppose a particular language or underlying disease causing the dysprosody, but the relative importance weights may, for instance, for reduced amplitude fluctuations, warrant an increase for other diseases. Such adjustments can, however, only be made with access to the appropriate speech recordings and the appropriate clinical expert raters in the language and are not considered further in this report. The components of the processing pipeline are, however, widely available and well documented [36,42,44,73,74], removing any barrier to replication, language or disease adjustments in weights, and replication efforts in later research.

5. Conclusions

Automatic assessment of dysprosody is feasible using the combination of overlap-aware segmentation of the speech signal to approximate utterances, the application of an automatic intonation stylization algorithm, and a novel quantification of the resulting representation of intonation, intensity, and inter-tonal timing. Using the proposed pipeline, a performance in dysprosody assessment that approximates the abilities of clinical expert raters was achieved. The variation in pitch over an utterance, which is the most often used quantification of dysprosody in

neurological diseases, does not contribute significantly to an understanding of the level of dysprosody in patients with Parkinson's disease.

Funding: The support of the Swedish Research Council (Grants numbers 2011-2294 and 421-2010-2131) for the projects in the framework of which the recordings were made is gratefully acknowledged.

Institutional Review Board Statement: The data collection procedure was approved by the regional Ethical review board (Ref. No.: 08- 093M; 2008-08-18).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study prior to recording.

Data Availability Statement: Under national law, audio recordings of speech are person-identifiable information and, therefore, cannot be shared outside the research group. Derived signal tracks and INTSINT / Momel annotations are available upon request.

Acknowledgments: The assistance of the clinical experts Lena Hartelius, Ellika Schalling, Katja Laakso, and Kerstin Johansson in making 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 (2018–2024, contract 2017-00626) and the 10 participating partner institutions, is gratefully acknowledged.

Conflicts of Interest: The author declare no conflict of interest.

References

- Schlenck, K.-J.; Bettrich, R.; Willmes, K. Aspects of Disturbed Prosody in Dysarthria. *Clinical Linguistics & Phonetics* **1993**, *7*, 119–128, doi:10.3109/02699209308985549.
- Sidtis, D.V.L.; Pachana, N.; Cummings, J.L.; Sidtis, J.J. Dysprosodic Speech Following Basal Ganglia Insult: Toward a Conceptual Framework for the Study of the Cerebral Representation of Prosody. *Brain & Language* **2006**, *97*, 135–153, doi:10.1016/j.bandl.2005.09.001.
- Watson, P.J.; Schlauch, R.S. The Effect of Fundamental Frequency on the Intelligibility of Speech with Flattened Intonation Contours. *American Journal of Speech-Language Pathology* **2008**, *17*, 348–355, doi:10.1044/1058-0360(2008/07-0048).
- Klopfenstein, M. Interaction between Prosody and Intelligibility. *Int J Speech-lang Pa* **2009**, *11*, 326–331, doi:10.1080/17549500903003094.
- Martens, H.; Nuffelen, G.V.; Dekens, T.; Huici, M.H.-D.; Hernández-Díaz, H.A.K.; Letter, M.D.; Bodt, M.S.D. The Effect of Intensive Speech Rate and Intonation Therapy on Intelligibility in Parkinson's Disease. *Journal of Communication Disorders* **2015**, *58*, 91–105, doi:10.1016/j.jcomdis.2015.10.004.
- Feenaughty, L.; Tjaden, K.; Sussman, J. Relationship between Acoustic Measures and Judgments of Intelligibility in Parkinson's Disease: A within-Speaker Approach. *Clinical Linguistics & Phonetics* **2014**, *28*, 857–878, doi:10.3109/02699206.2014.921839.
- Sidtis, J.J.; Sidtis, D.V.L. A Neurobehavioral Approach to Dysprosody. *Seminars in speech and language* **2003**, *24*, 93–105, doi:10.1055/s-2003-38901.
- Sidtis, J.J. Music, Pitch Perception, and the Mechanisms of Cortical Hearing. *Handbook of Cognitive Neuroscience*. **1984**, 91–114, doi:10.1007/978-1-4899-2177-2_5.
- Peters, A.S.; Rémi, J.; Vollmar, C.; Gonzalez-Victores, J.A.; Cunha, J.P.S.; Noachtar, S. Dysprosody during Epileptic Seizures Lateralizes to the Nondominant Hemisphere. *Neurology* **2011**, *77*, 1482–1486, doi:10.1212/wnl.0b013e318232abae.
- Ballard, K.J.; Azizi, L.; Duffy, J.R.; McNeil, M.R.; Halaki, M.; O'Dwyer, N.; Layfield, C.; Scholl, D.I.; Vogel, A.P.; Robin, D.A. A Predictive Model for Diagnosing Stroke-Related Apraxia of Speech. *Neuropsychologia* **2016**, *81*, 129–139, doi:10.1016/j.neuropsychologia.2015.12.010.
- Ballard, K.J.; Robin, D.A.; McCabe, P.; McDonald, J. A Treatment for Dysprosody in Childhood Apraxia of Speech. *J. Speech, Lang., Hear. Res.* **2010**, *53*, 1227–1245, doi:10.1044/1092-4388(2010/09-0130).
- Rusz, J.; Saft, C.; Schlegel, U.; Hoffman, R.; Skodda, S. Phonatory Dysfunction as a Preclinical Symptom of Huntington Disease. *PLoS ONE* **2014**, *9*, e113412, doi:10.1371/journal.pone.0113412.
- Skodda, S.; Grönheit, W.; Schlegel, U. Intonation and Speech Rate in Parkinson's Disease: General and Dynamic Aspects and Responsiveness to Levodopa Admission. *Journal of Voice* **2011**, *25*, e199-205, doi:10.1016/j.jvoice.2010.04.007.
- Karlsson, F.; Olofsson, K.; Blomstedt, P.; Linder, J.; Doorn, J. van Pitch Variability in Patients with Parkinson's Disease: Effects of Deep Brain Stimulation of Caudal Zona Incerta and Subthalamic Nucleus. *Journal of Speech, Language and Hearing Research* **2013**, *56*, 1–9, doi:10.1044/1092-4388(2012/11-0333).

15. Steurer, H.; Schalling, E.; Franzén, E.; Albrecht, F. Characterization of Mild and Moderate Dysarthria in Parkinson's Disease: Behavioral Measures and Neural Correlates. *Front Aging Neurosci* **2022**, *14*, 870998, doi:10.3389/fnagi.2022.870998.
16. Frota, S.; Cruz, M.; Cardoso, R.; Guimarães, I.; Ferreira, J.J.; Pinto, S.; Vigário, M. (Dys)Prosody in Parkinson's Disease: Effects of Medication and Disease Duration on Intonation and Prosodic Phrasing. *Brain Sci* **2021**, *11*, 1100, doi:10.3390/brainsci11081100.
17. Bocklet, T.; Nöth, E.; Stemmer, G.; Ruzickova, H.; Ruzs, J. *Detection of Persons with Parkinson's Disease by Acoustic, Vocal, and Prosodic Analysis; ... (ASRU)*; IEEE, 2011; ISBN 978-1-4673-0365-1.
18. MacPherson, M.K.; Huber, J.E.; Snow, D.P. The Intonation-Syntax Interface in the Speech of Individuals with Parkinson's Disease. *Journal of Speech, Language and Hearing Research* **2011**, *54*, 19–32, doi:10.1044/1092-4388(2010/09-0079).
19. Thies, T.; Mücke, D.; Lowit, A.; Kalbe, E.; Steffen, J.; Barbe, M.T. Prominence Marking in Parkinsonian Speech and Its Correlation with Motor Performance and Cognitive Abilities. *Neuropsychologia* **2019**, 107306, doi:10.1016/j.neuropsychologia.2019.107306.
20. Mennen, I.; Schaeffler, F.; Watt, N.; Miller, N. An Autosegmental-Metrical Investigation of Intonation in People with Parkinson's Disease. *Asia Pacific Journal of Speech, Language and Hearing* **2008**, *11*, 205–219.
21. Lowit, A.; Kuschmann, A. Characterizing Intonation Deficit in Motor Speech Disorders: An Autosegmental-Metrical Analysis of Spontaneous Speech in Hypokinetic Dysarthria, Ataxic Dysarthria, and Foreign Accent Syndrome. *Journal of Speech, Language and Hearing Research* **2012**, *55*, 1472–1484, doi:10.1044/1092-4388(2012/11-0263).
22. Lowit, A.; Kuschmann, A.; Kavanagh, K. Phonological Markers of Sentence Stress in Ataxic Dysarthria and Their Relationship to Perceptual Cues. *Journal of Communication Disorders* **2014**, *50*, 8–18, doi:10.1016/j.jcomdis.2014.03.002.
23. Pierrehumbert, J. Tonal Elements and Their Alignment. In *Prosody: Theory and Experiment; Studies Presented to Gösta Bruce*; Text, Speech and Language Technology; 2000; pp. 11–36 ISBN 9789048155620.
24. Tykalová, T.; Ruzs, J.; Cmejla, R.; Ruzickova, H.; Ruzicka, E. Acoustic Investigation of Stress Patterns in Parkinson's Disease. *Journal of Voice* **2013**, *28*, 129.e1–129.e8, doi:10.1016/j.jvoice.2013.07.001.
25. Tavi, L.; Werner, S. A Phonetic Case Study on Prosodic Variability in Suicidal Emergency Calls. *Int. J. Speech Lang. Law* **2020**, *0*, 59–74, doi:10.1558/ijssl.39667.
26. Tavi, L.; Penttilä, N. Functional Data Analysis of Prosodic Prominence in Parkinson's Disease: A Pilot Study. *Clin Linguist Phonet* **2023**, *ahead-of-print*, 1–18, doi:10.1080/02699206.2022.2158372.
27. Hernandez, A.; Kim, S.; Chung, M. Prosody-Based Measures for Automatic Severity Assessment of Dysarthric Speech. *Appl Sci* **2020**, *10*, 6999, doi:10.3390/app10196999.
28. Hirst, D.J. The Analysis by Synthesis of Speech Melody. *Journal of Speech Sciences* **2011**, *1*, 55–83.
29. Hirst, D.J. Form and Function in the Representation of Speech Prosody. *Speech Communication* **2005**, *46*, 334–347, doi:10.1016/j.specom.2005.02.020.
30. Hirst, D.; Cristo, A. di *Intonation Systems: A Survey of Twenty Languages*; Cambridge University Press, 1998; Vol. 76;.
31. Véronis, J.; Cristo, P.D.; Courtois, F.; Chaumette, C. A Stochastic Model of Intonation for Text-to-Speech Synthesis. *Speech Commun* **1998**, *26*, 233–244, doi:10.1016/s0167-6393(98)00063-6.
32. Hirst, D.; Cho, H.; Kim, S.; Yu, H. Evaluating Two Versions of the Momel Pitch Modelling Algorithm on a Corpus of Read Speech in Korean.; 2007; pp. 1649–1652.
33. Hirst, D. Melody Metrics for Prosodic Typology: Comparing English, French and Chinese. *Interspeech 2013* **2013**, 572–576, doi:10.21437/interspeech.2013-158.
34. Celeste, L.C.; Reis, C. Formal Intonative Analysis: Intsint Applied to Portuguese. *J Speech Sci* **2021**, *2*, 3–21, doi:10.20396/joss.v2i2.15026.
35. Chentir Extraction of Arabic Standard Micromelody. *J Comput Sci* **2009**, *5*, 86–89, doi:10.3844/jcssp.2009.86.89.
36. Hirst, D.J. A Praat Plugin for Momel and INTSINT with Improved Algorithms for Modelling and Coding Intonation.; 2007; pp. 1233–1236.
37. Xu, Y. ProsodyPro—A Tool for Large-Scale Systematic Prosody Analysis. In *Proceedings of the Tools and Resources for the Analysis of Speech Prosody*; Laboratoire Parole et Langage, France: Aix-en-Provence, France, 2013; pp. 7–10.
38. Liss, J.M.M.; White, L.; Mattys, S.L.; Lansford, K.; Lotto, A.J.; Spitzer, S.M.; Caviness, J.N. Quantifying Speech Rhythm Abnormalities in the Dysarthrias. *Journal of Speech, Language and Hearing Research* **2009**, *52*, 1334–1352, doi:10.1044/1092-4388(2009/08-0208).
39. Lavechin, M.; Métais, M.; Titeux, H.; Boissonnet, A.; Copet, J.; Rivière, M.; Bergelson, E.; Cristia, A.; Dupoux, E.; Bredin, H. Brouhaha: Multi-Task Training for Voice Activity Detection, Speech-to-Noise Ratio, and C50 Room Acoustics Estimation. *Arxiv* **2022**, doi:10.48550/arxiv.2210.13248.

40. Räsänen, O.; Seshadri, S.; Lavechin, M.; Cristia, A.; Casillas, M. ALICE: An Open-Source Tool for Automatic Measurement of Phoneme, Syllable, and Word Counts from Child-Centered Daylong Recordings. *Behav Res Methods* **2021**, *53*, 818–835, doi:10.3758/s13428-020-01460-x.
41. Cristia, A.; Lavechin, M.; Scaff, C.; Soderstrom, M.; Rowland, C.; Räsänen, O.; Bunce, J.; Bergelson, E. A Thorough Evaluation of the Language Environment Analysis (LENA) System. *Behav Res Methods* **2021**, *53*, 467–486, doi:10.3758/s13428-020-01393-5.
42. Bullock, L.; Bredin, H.; Garcia-Perera, L.P. Overlap-Aware Diarization: Resegmentation Using Neural End-to-End Overlapped Speech Detection. *Icassp 2020 - 2020 IEEE Int Conf Acoust Speech Signal Process Icassp* **2020**, *00*, 7114–7118, doi:10.1109/icassp40776.2020.9053096.
43. Bredin, H.; Yin, R.; Coria, J.M.; Gelly, G.; Korshunov, P.; Lavechin, M.; Fustes, D.; Titeux, H.; Bouaziz, W.; Gill, M.-P. Pyannote.Audio: Neural Building Blocks for Speaker Diarization. *Icassp 2020 - 2020 IEEE Int Conf Acoust Speech Signal Process Icassp* **2020**, *00*, 7124–7128, doi:10.1109/icassp40776.2020.9052974.
44. Yin, R.; Bredin, H.; Barras, C. Neural Speech Turn Segmentation and Affinity Propagation for Speaker Diarization. *Interspeech 2018* **2018**, 1393–1397, doi:10.21437/interspeech.2018-1750.
45. Goetz, C.G.; Tilley, B.C.; Shaftman, S.R.; Stebbins, G.T.; Fahn, S.; Martin, P.M.; Poewe, W.H.; Sampaio, C.; Stern, M.B.; Dodel, R.; et al. Movement Disorder Society-sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale Presentation and Clinimetric Testing Results. *Movement Disorders* **2008**, *23*, 2129–2170, doi:10.1002/mds.22340.
46. Karlsson, F.; Schalling, E.; Laakso, K.; Johansson, K.M.; Hartelius, L. Assessment of Speech Impairment in Patients with Parkinson's Disease from Acoustic Quantifications of Oral Diadochokinetic Sequences. *The Journal of the Acoustical Society of America* **2020**, *147*, 839–851, doi:10.1121/10.0000581.
47. Jolad, B.; Khanai, R. An Art of Speech Recognition: A Review. *2019 2nd Int Conf Signal Process Commun Icspsc* **2019**, *00*, 31–35, doi:10.1109/icspsc46172.2019.8976733.
48. Bhakre, S.K.; Bang, A. Emotion Recognition on the Basis of Audio Signal Using Naive Bayes Classifier. *2016 Int Conf Adv Comput Commun Informatics Icacci* **2016**, 2363–2367, doi:10.1109/icacci.2016.7732408.
49. Sanchis, A.; Juan, A.; Vidal, E. A Word-Based Naïve Bayes Classifier for Confidence Estimation in Speech Recognition. *IEEE Transactions Audio Speech Lang Process* **2012**, *20*, 565–574, doi:10.1109/tasl.2011.2162403.
50. Liu, Z.-T.; Wu, M.; Cao, W.-H.; Mao, J.-W.; Xu, J.-P.; Tan, G.-Z. Speech Emotion Recognition Based on Feature Selection and Extreme Learning Machine Decision Tree. *Neurocomputing* **2018**, *273*, 271–280, doi:10.1016/j.neucom.2017.07.050.
51. Lavner, Y.; Ruinskiy, D. A Decision-Tree-Based Algorithm for Speech/Music Classification and Segmentation. *Eurasip J Audio Speech Music Process* **2009**, *2009*, 239892, doi:10.1155/2009/239892.
52. Noroozi, F.; Sapiński, T.; Kamińska, D.; Anbarjafari, G. Vocal-Based Emotion Recognition Using Random Forests and Decision Tree. *Int J Speech Technology* **2017**, *20*, 239–246, doi:10.1007/s10772-017-9396-2.
53. Arora, S.; Tsanas, A. Assessing Parkinson's Disease at Scale Using Telephone-Recorded Speech: Insights from the Parkinson's Voice Initiative. *Diagnostics* **2021**, *11*, 1892, doi:10.3390/diagnostics11101892.
54. Haq, A.U.; Li, J.P.; Memon, M.H.; Khan, J.; Malik, A.; Ahmad, T.; Ali, A.; Nazir, S.; Ahad, I.; Shahid, M. Feature Selection Based on L1-Norm Support Vector Machine and Effective Recognition System for Parkinson's Disease Using Voice Recordings. *IEEE Access* **2019**, *7*, 37718–37734, doi:10.1109/access.2019.2906350.
55. Lahmiri, S.; Shmuel, A. Detection of Parkinson's Disease Based on Voice Patterns Ranking and Optimized Support Vector Machine. *Biomed Signal Proces* **2019**, *49*, 427–433, doi:10.1016/j.bspc.2018.08.029.
56. Shahbakhi, M.; Far, D.T.; Tahami, E. Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine. *J Biomed Sci Eng* **2014**, *2014*, 147–156, doi:10.4236/jbise.2014.74019.
57. Novotný, M.; Ruz, J.; Cmejla, R.; Ruzicka, E. Automatic Evaluation of Articulatory Disorders in Parkinson's Disease. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **2014**, *22*, 1366–1378, doi:10.1109/taslp.2014.2329734.
58. Karlsson, F.; Hartelius, L. How Well Does Diadochokinetic Task Performance Predict Articulatory Imprecision? Differentiating Individuals with Parkinson's Disease from Control Subjects. *Folia Phoniatrica et Logopaedica* **2019**, 1–10, doi:10.1159/000498851.
59. Dupuy, D.; Helbert, C.; Franco, J. DiceDesign and DiceEval: Two R Packages for Design and Analysis of Computer Experiments. *J Stat Softw* **2015**, *65*, doi:10.18637/jss.v065.i11.
60. Chicco, D.; Jurman, G. The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. *Bmc Genomics* **2020**, *21*, 6, doi:10.1186/s12864-019-6413-7.
61. Kautz, T.; Eskofier, B.M.; Pasluosta, C.F. Generic Performance Measure for Multiclass-Classifiers. *Pattern Recogn* **2017**, *68*, 111–125, doi:10.1016/j.patcog.2017.03.008.
62. Zien, A.; Krämer, N.; Sonnenburg, S.; Rätsch, G. 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.* **2009**, 694–709, doi:10.1007/978-3-642-04174-7_45.

63. Traunmüller, H.; Eriksson, A. The Perceptual Evaluation of F0 Excursions in Speech as Evidenced in Liveliness Estimations. *The Journal of the Acoustical Society of America* **1995**, *97*, 1905–1915, doi:10.1121/1.412942.
64. Tanaka, Y.; Tsuboi, T.; Watanabe, H.; Kajita, Y.; Fujimoto, Y.; Ohdake, R.; Yoneyama, N.; Masuda, M.; Hara, K.; Senda, J.; et al. Voice Features of Parkinson's Disease Patients with Subthalamic Nucleus Deep Brain Stimulation. *Journal of Neurology* **2015**, *262*, 1–9, doi:10.1007/s00415-015-7681-z.
65. Tsuboi, T.; Watanabe, H.; Tanaka, Y.; Ohdake, R.; Yoneyama, N.; Hara, K.; Nakamura, R.; Watanabe, H.; Senda, J.; Atsuta, N.; et al. Distinct Phenotypes of Speech and Voice Disorders in Parkinson's Disease after Subthalamic Nucleus Deep Brain Stimulation. *Journal of Neurology, Neurosurgery & Psychiatry* **2014**, *86*, jnnp-2014-308043, doi:10.1136/jnnp-2014-308043.
66. Tsuboi, T.; Watanabe, H.; Tanaka, Y.; Ohdake, R.; Yoneyama, N.; Hara, K.; Ito, M.; Hirayama, M.; Yamamoto, M.; Fujimoto, Y.; et al. Characteristic Laryngoscopic Findings in Parkinson's Disease Patients after Subthalamic Nucleus Deep Brain Stimulation and Its Correlation with Voice Disorder. *Journal of Neural Transmission* **2015**, *122*, 1663–1672, doi:10.1007/s00702-015-1436-y.
67. Karlsson, F.; Blomstedt, P.; Olofsson, K.; Linder, J.; Nordh, E.; Doorn, J. van Control of Phonatory Onset and Offset in Parkinson Patients Following Deep Brain Stimulation of the Subthalamic Nucleus and Caudal Zona Incerta. *Parkinsonism & Related Disorders* **2012**, *18*, 824–827, doi:10.1016/j.parkreldis.2012.03.025.
68. Eklund, E.; Qvist, J.; Sandström, L.; Viklund, F.; Doorn, J. van; Karlsson, F. Perceived Articulatory Precision in Patients with Parkinson's Disease after Deep Brain Stimulation of Subthalamic Nucleus and Caudal Zona Incerta. *Clinical Linguistics & Phonetics* **2014**, *29*, 150–166, doi:10.3109/02699206.2014.971192.
69. Goberman, A.M.; Blomgren, M. Fundamental Frequency Change During Offset and Onset of Voicing in Individuals with Parkinson Disease. *Journal of Voice* **2008**, *22*, 178–191, doi:10.1016/j.jvoice.2006.07.006.
70. Whitfield, J.A.; Goberman, A.M. The Effect of Parkinson Disease on Voice Onset Time: Temporal Differences in Voicing Contrast. *The Journal of the Acoustical Society of America* **2015**, *137*, 2432–2432, doi:10.1121/1.4920874.
71. Goberman, A.M.; Coelho, C.; Robb, M. Phonatory Characteristics of Parkinsonian Speech before and after Morning Medication: The ON and OFF States. *Journal of Communication Disorders* **2002**, *35*, 217–239, doi:10.1016/s0021-9924(01)00072-7.
72. Utter, A.A.; Basso, M.A. The Basal Ganglia: An Overview of Circuits and Function. *Neurosci. Biobehav. Rev.* **2008**, *32*, 333–342, doi:10.1016/j.neubiorev.2006.11.003.
73. Bredin, H.; Yin, R.; Coria, J.M.; Gelly, G.; Korshunov, P.; Lavechin, M.; Fustes, D.; Titeux, H.; Bouaziz, W.; Gill, M.-P. Pyannote.Audio: Neural Building Blocks for Speaker Diarization. *Arxiv* **2019**.
74. Origlia, A.; Abete, G.; Cutugno, F. A Dynamic Tonal Perception Model for Optimal Pitch Stylization. *Computer Speech & Language* **2013**, *27*, 190–208, doi:10.1016/j.csl.2012.04.003.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.