5-2024

Acoustic-Perceptual Relations Between Fundamental Frequency and Expressiveness in Speakers With Hypokinetic Dysarthria

Alena Portnova
Utah State University, alena.portnova@usu.edu

ACOUSTIC-PERCEPTUAL RELATIONS BETWEEN FUNDAMENTAL FREQUENCY AND EXPRESSIVENESS IN SPEAKERS WITH HYPOKINETIC DYSARTHRIA

by

Alena Portnova

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Communication Sciences

Approved:

Annalise R. Fletcher, Ph.D. Major Professor	Stephanie A. Borrie, Ph.D. Committee Member
Alan A. Wisler, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2024

ABSTRACT

Acoustic-Perceptual Relations Between Fundamental Frequency and Expressiveness in Speakers with Hypokinetic Dysarthria

by

Alena Portnova, Master of Science

Utah State University, 2024

Major Professor: Dr. Annalise R. Fletcher
Department: Communicative Disorders and Deaf Education

This study answered two research questions. First, the study examined the degree to which tracking errors influence our ability to measure F0 mean and standard deviation across groups. Second, it compared how different statistical approaches to measuring F0 variability affect the relationship between acoustic measures and the perception of speech expressiveness. To evaluate the accuracy of F0 tracking, two Praat scripts were compared – the standard Praat extraction algorithm and a customized script that excluded tracking errors from the analysis. To measure the perception of speech expressiveness, a total of 11 listeners rated 90 sentences from nine speakers with Parkinson's disease and nine healthy control speakers answering the question "How expressive is this speaker?". To correlate the perception of expressiveness with F0 variability, six different statistics indexing F0 variation were calculated. Separate mixed-effects regression models were constructed to examine each research question. Results revealed that F0 summary statistics such as mean and SD are significantly affected by tracking errors in both groups of speakers. The results also demonstrated that F0 variability is a factor that contributes to the perception of speech expressiveness, and that some F0 variability statistics, such as F0 SD in semitones or F0 relative SD, are better metrics for capturing changes in F0 that are relevant to listeners' judgments.

(48 pages)

PUBLIC ABSTRACT

Acoustic-Perceptual Relations Between Fundamental Frequency and Expressiveness in Speakers with Hypokinetic Dysarthria

Alena Portnova

Hypokinetic dysarthria is a motor speech disorder that occurs in patients with Parkinson's disease (PD) and affects not only their speech intelligibility but also how their attitudes and emotions are perceived by listeners. People with PD have been judged as less happy, involved, friendly, and interested based only on their speech samples. A lack of speech expressiveness is one of the characteristics that is likely to be related to these negative listener judgments. Specifically, it has been suggested that a lack of fundamental frequency (F0) variation reduces speakers' ability to express various emotions. To investigate whether speech expressiveness is related to F0 variation, it is necessary to accurately measure how F0 changes across sentences. However, existing software produces a lot of tracking errors when measuring F0, so the accuracy of various statistical measures of F0 is questionable. The current study answered two research questions. First, we evaluated one of the most widely used software applications in the field of speech disorders to assess how tracking errors affected measurements of F0 mean and standard deviation in healthy speakers and people with PD. This was done by manually annotating all errors and calculating summary statistics that included and excluded these values. Second, we explored different statistical approaches to measuring F0 variability to determine which measurements had the strongest relationship with speech expressiveness. To address this question, listeners were presented with 90 sentences, read by both healthy speakers and speakers with PD, and were asked to rate "How expressive is this speaker?" using a visual analog scale. Statistical analysis evaluated the relationship between acoustic and perceptual variables and compared the results 1) before and after excluding tracking errors in F0 measurement; 2) when six different statistical approaches to measuring F0 variability were substituted into the model. Results demonstrated that tracking errors in the F0 contour significantly affect F0 summary statistics. This finding highlights the importance of manually screening and removing outliers when evaluating group differences in F0 statistics. Results also demonstrated that some (but not all) of the statistical approaches for measuring F0 variability show significant relationships with listeners' perceptions of speech expressiveness. A larger number of approaches showed a significant relationship with expressiveness scores when tracking errors were excluded from the analysis and the approaches that best accounted for listener ratings provided some normalization of sex-related differences in frequency values. These findings offer insight into the type of changes in F0 that are most relevant to listeners in their perception of speech expressiveness.

ACKNOWLEDGMENTS

I would like to thank the following people, without whom I would not have been able to complete this research, and without whom I would not have made it through my master's degree. The members of the Motor Speech Lab for the assistance with this project, and especially my research supervisor, Dr. Annalise Fletcher, for her mentorship, support, and patience during the research and writing process. The thesis committee, Dr. Alan Wisler and Dr. Stephanie Borrie, for their guidance and valuable feedback. Fulbright Student Program for the opportunity and financial support. My friends for always being there for me. And my special thanks to my parents for their continued love and support.

Alena Portnova

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Introduction

Dysarthria is a collective name for a group of neurologic speech disorders that reflect abnormalities in the strength, speed, range, steadiness, tone, or accuracy of movements required for speech production (Duffy, 2013). Parkinson's Disease (PD) is a neurodegenerative disease that causes dysarthria in up to 90% of patients (Müller et al., 2001). Dysarthria in PD typically manifests as hypokinetic dysarthria caused by breakdowns in the basal ganglia-thalamocortical circuits (Brabenec et al., 2017). The salient manifestations of hypokinetic dysarthria are perceptually monotonous speech in terms of pitch (voice highness or lowness) and loudness, reduced stress, deviating speech rate, underarticulation, as well as a harsh or breathy voice quality, all of which worsen with the progression of the disease (Duffy, 2013). These speech changes can have a significant impact on communication, contributing to reductions in intelligibility, as well as changes in speakers' participation in everyday situations (Borrie et al., 2022; McAuliffe et al., 2017; Spencer et al., 2020). Studies have highlighted that it is not only difficulty being understood that negatively affects people with PD, but also negative listener judgments of their cognitive skills and personalities (Miller et al., 2006; Yorkston et al., 2017). For example, a "flattening" of pitch in hypokinetic dysarthria may lead others to believe the speaker is depressed (Yorkston et al., 2017). Thus, it is important to consider how different features of hypokinetic dysarthria might influence not just intelligibility, but also emotional expression. Features of hypokinetic dysarthria, as well as speech in general, may be analyzed either in a more subjective way – through perceptual analysis to assess listeners' impressions of the speaker, or in a more objective way – by an acoustic analysis using speech analytics software. To best understand what aspects of the speech signal cause changes in listeners' perceptions of dysarthria, it is important to develop a strong understanding of the correlations between perceptual and acoustic measures.

There is a large body of literature focused on the acoustic characteristics of hypokinetic dysarthria and their effect on speech perception. Specifically, a lot of studies investigate intelligibility – a measure of how comprehensible the speech is to a listener. Several acoustic measurements have been suggested to index speakers' intelligibility. For example, a speaker's second formant (F2) slope within diphthongs has been found to be positively correlated with intelligibility scores in both Parkinson's disease and other neurological disorders (Kim et al., 2009, 2011; Tjaden et al., 2013; Weismer et al., 2001). However, the extent to which F2 slope is correlated with intelligibility varies across conditions, and therefore the measurement may be most relevant when intelligibility is more severely affected (e.g., when listening to speech in noisy environments or when deciphering uncommon, more complicated words) (Chiu et al., 2019). Intelligibility scores are also correlated with changes in the vowel space area in speakers with dysarthria. However, even though several studies show a clear relationship between these two measurements (e.g., Lansford & Liss, 2014; Tjaden & Wilding, 2004), the strength of the correlations vary, implying that some extraneous factors might be influencing the relationship. For example, Tjaden & Wilding (2004) showed that there is a difference between males and females in how vowel space area is related to the intelligibility ratings – with correlations found to be significant only in female speakers. Weismer et al. (2001) also found that the relationship may differ across dysarthria types, demonstrating that speakers with ALS had stronger relationship between vowel space area and intelligibility than speakers with PD.

While these measurements of speech articulation have been of particular interest in the dysarthria literature, they cannot fully describe the symptoms of hypokinetic dysarthria, particularly in the early stages following diagnosis of PD. As highlighted by Ho et al., (1998), symptoms of voice impairment (e.g., reduced volume and monotonicity) almost always occur as the first symptoms of hypokinetic dysarthria, before articulatory impairments are apparent. Indeed, in a systematic review on hypokinetic dysarthria, Brabenec et al., (2017) divides a plethora of conventional acoustic measurements used in hypokinetic dysarthria analysis into five categories of speech production, with only one related directly to articulation: 1) features describing the impairment of vocal fold vibration (e.g., jitter and shimmer); 2) general voice quality deterioration (e.g., harmonic-to-noise ratio); 3) indicators of intonation impairment (e.g., standard deviation of fundamental frequency, range of fundamental frequency); 4) speech rate disturbances (e.g., total pause time, articulation rate); and 5) impairments of articulation and tongue movement (e.g., frequencies of the first three formants, vowel space area). Furthermore, Kovac et al. (2024) found that the most robust, language-dependent features for distinguishing speech patterns in PD included variability in fundamental frequency (F0) and number of pauses during a reading task, in addition to changes in the second formant frequencies (Kovac et al., 2024).

Examining these broader acoustic changes in hypokinetic dysarthria is important because the perceptual consequences of hypokinetic dysarthria are not limited to changes in articulation and intelligibility. Dysprosody – one of the most salient features of PD –

may be an additional barrier to effective communication, since prosodic variability carries a great communicative value of expressing attitudes and emotions (Rodero, 2011). The issue of emotional expression is considered in more detail in the following section.

Dysprosody and Emotions

Evidence suggests that people with hypokinetic dysarthria experience limitations in their ability to express emotions and emphasize important words within the speech signal. Studies investigating the ability of speakers with PD to express stress or certain emotions consistently show that speakers with PD are less expressive than healthy controls and, in particular, are less able to express negative emotions such as anger, disgust, or fear (Anzuino et al., 2023; Caekebeke et al., 1991; Möbes et al., 2008; Pell et al., 2006). For example, Pell et al., (2006) showed that happiness, disgust, and anger were often mislabeled as either sadness or neutral emotions because speakers lacked necessary variation in F0 and loudness required to express those feelings. Similiarly, Tykalova et al., (2014) demonstrated that people with PD have difficulties producing contrastive stress – emphasizing different words within a sentence – since that requires elevated F0 range and increased volume (Tykalova et al., 2014). Möbes et al., (2008) also found a statistically significant difference in the F0 and intensity ranges used to express neutral speech, happiness, and sadness in speakers with PD and healthy controls. Gnerre et al., (2023) confirmed these findings and showed that patients with PD had decreased changes in both prosodic features (related to speech rate, intensity and F0) and voice quality features (e.g. cepstral peak prominence) when expressing different emotions. However, they suggested that the voice quality findings were somewhat sex-specific – as females with hypokinetic dysarthria demonstrated greater changes in their expression of happiness, sadness and anger when compared to neurotypical speakers (Gnerre et al., 2023). In summary, there is considerable evidence that PD can affect speakers' abilities to convey different emotional states.

There is also evidence that acoustic changes associated with hypokinetic dysarthria can influence listeners' impressions of people with PD. For example, people with PD have been judged as less happy, involved, friendly, and interested based only on their speech samples (Jaywant & Pell, 2010; Pitcairn et al., 1990). These findings are particularly concerning, as changes in listener judgments may directly influence their likeliness to engage in communicative exchanges. It is hypothesized that changes in F0, perceptually recognized as pitch, are particularly relevant to these listener judgments. Pitch and pitch variability have been shown to be strongly correlated with liveliness (Traunmüller & Eriksson, 1995) and charisma (Strangert & Gustafson, 2008) in healthy speakers. However, despite monotonicity (i.e. lack of pitch variation) being considered one of the most prominent characteristics of hypokinetic dysarthria (Darley et al., 1969), and variability in F0 being one of the most robust acoustic features for PD detection (Kovac et al., 2024), studies of F0 variability have shown conflicting results across different groups of speakers with PD. While some studies have found that F0 variability is decreased in speakers with PD (Bowen et al., 2013; A. Goberman et al., 2005; Jiménez-Jiménez et al., 1997; Skodda et al., 2011; Tykalova et al., 2014), other works show significant decreases in F0 variability in females but not males (Holmes et al., 2000; Jaywant & Pell, 2010) or higher mean F0 values in patients with PD compared to healthy controls (A. M. Goberman & Blomgren, 2008). This inconsistency in results might be possibly explained by several factors –

different variability indices used, speech analytics software used, and the methods of dealing with F0 tracking errors. A better understanding of F0 variability is vital for interpreting how and why emotional expression is affected in PD. Thus, we need to consider how different methods of measuring F0 variability affect the values commonly reported in the motor speech literature.

Challenges Associated with Acoustic Measurement of F0

There is evidence that widely used acoustic speech analysis software tools make numerous errors when tracking F0. For example, Praat software is the most frequently used method for estimating F0 in linguistics, computer science, audiology, and speech-language pathology (Boersma & Weeninck, 2024; Strömbergsson, 2016). However, as discussed in Exner (2019), Praat's function "get pitch" is not robust to abnormal voice qualities such as breathy voices or vocal fry. Unfortunately, these voice characteristics are fairly common within dysarthric speech (Duffy, 2013). The following types of F0 tracking errors are frequently seen when Praat's "get pitch" function is used: 1) not tracking F0 at all during the voiced segments; 2) tracking F0 during unvoiced segments; 3) tracking subharmonic and/or overtone frequencies (frequencies that are lower or higher than the speaker's perceived pitch) (Exner, 2019). These errors result in outlying data points which could affect the summary statistics used for describing F0 and F0 variability. Therefore, before conducting a statistical analysis of F0, it may be necessary to manually review F0 data, especially when analyzing speech from people with dysarthria.

One option for dealing with F0 errors is to identify and discard these values from further statistical analysis. This strategy has been reported in numerous studies focused on both healthy and dysarthric speech. For instance, in their study of pitch in native and non-native Lombard speech, Marcoux & Ernestus (2019) deleted doubling, halving, and creaky voice prior to calculating minimum and median F0. Bowen et al. (2013) and Tykalova et al. (2014) reported removing all inconsistencies or incorrect detections of F0 after visual inspection of the pitch contour. Excluding the extreme, outlying values of minimum and maximum F0 can also be a strategy for removing errors, based on the location of those values within the utterance. For example, Van Der Bruggen et al. (2023) replaced extreme minimum and maximum values if they were isolated and were not associated with either phoneme boundaries or accented tones. The maximum values were replaced by the highest boundary or accentuation targets, and the minimum values were replaced by the lowest boundary or accentuation targets.

Another strategy for managing tracking errors is to make an attempt at directly correcting the errors in F0 estimates. For instance, Exner (2019) suggested the following procedure for correction of subharmonic and overtone frequencies: 1) record the total duration of the voiced segment; 2) measure F0 in any portion of the segment with normal F0 tracking and record the duration of those portions; 3) measure F0 in any subharmonic/overtone portion of the segment and multiply/divide it by 2, record the duration of those portions, and 4) repeat within the entire voiced segment and use the following formula to calculate the corrected mean F0:

$$\left(f0_1 * \left(\frac{d_1}{d_t}\right)\right) + \left((f0_2 * 2) * \left(\frac{d_2}{d_t}\right)\right) + \cdots + \left(f0_n * \left(\frac{d_n}{d_t}\right)\right)$$

where d_1 – total duration; d_n – duration of the analyzed portion of the segment; $f0_n$ - F0 value within the analyzed portion of the segment. This strategy has the advantage of minimizing the removal of F0 information from voiced segments. However, conducting manual corrections for every doubling/halving error is considerably more cumbersome than removing F0 values, and has the potential to introduce new errors if not performed accurately.

Getting accurate values also depends on choosing appropriate settings when using speech analysis software. There are several studies that emphasize the importance of standardization and reporting of F0 settings and procedures when performing acoustic analyses of voice quality (Brockmann-Bauser & de Paula Soares, 2023; Vogel et al., 2009). When using Praat software, many studies in the field of speech disorders use the standard, or default, settings for identifying F0 since there is no other 'gold standard'. However, the standard floor/ceiling values might not always be appropriate for some clinical populations, especially for those whose speech is characterized by reduced F0 variability, lower F0 values in general, or glottal fry, and speaker-specific settings might be required for adequate F0 estimation.

Under ideal circumstances, F0 settings could be perfectly individualized based on the vocal characteristics of each speaker. However, as mentioned by Vogel et al., (2009) establishing speaker-appropriate pitch range settings requires determining an analysis window length, intensity cut-offs, and expert knowledge of software configurations. To successfully establish this for every individual speaker in a study would be extremely time and resource consuming. Manually altering settings for each person also introduces some degree of subjectivity (and opportunities for bias) in a previously objective acoustic measurement. A more helpful approach could be standardizing the selection of pitch range settings based on certain key speaker characteristics. For example, in their study, Vogel et al. (2009) found that even just using sex-specific pitch range settings improved the quality of F0 analysis. Indeed, these sex-specific settings demonstrated similar results to those obtained when speaker-specific, individualized floor/ceiling values were used.

Despite the improvements observed with sex-specific F0 settings, tracking errors remain a concerning issue, especially when it comes to analyses of disordered vocal qualities. Therefore, other techniques might be needed to determine appropriate floor/ceiling values and reduce the number of potential 'octave jumps' and other extreme F0 values. In the study by Looze et al. (2012), a several step standardized process was proposed for choosing appropriate F0 settings for each speaker: 1) floor/ceiling are set to extreme values of 60 and 600 Hz; 2) new floor and ceiling values are adapted based on the results obtained in the first step. To exclude the effects of outlying values, the authors suggest adapting floor and ceiling values based on the speakers' 15^{th} and 65^{th} F0 percentiles (e.g., q15*0.83 and q65*1.92 (where 'q' = percentile). This procedure was shown to exclude more F0 tracking errors than setting pitch parameters to 100-500 Hz range for female voices and 75-300 Hz range for male voices. Nevertheless, even though there are studies which demonstrate various techniques for dealing with F0 tracking errors and suggest other pitch settings to use for acoustic analysis, there is no agreed upon standard and much variability remains in analysis procedures.

It should be acknowledged that there are some alternative algorithms for F0 estimation, outside of Praat, which have been used to estimate pitch in dysarthric speech. For example, RAPT (Robust Algorithm for Pitch Tracking) and YIN have been used for pitch analysis in PD speech (Rodríguez-Pérez et al., 2019; Verkhodanova, 2021) and were stated to be the best performing algorithms for F0 estimation in single speakers. However, a systematic review by Strömbergsson (2016) comparing multiple F0 extraction algorithms, including the abovementioned, demonstrated that Praat is not only the most frequently used method in such research areas as Linguistics, Computer Science, Audiology, and Speech-Language Pathology, but also the most accurate. In this study, the accuracy was determined by comparing the reference F0 contours in the speech corpus to the F0 contours extracted by Praat, RAPT, and YIN algorithms. Four evaluation metrics were used – 1) gross pitch error (the proportion of frames where the relative pitch error is higher than 20%), 2) fine pitch error (the standard deviation of relative error values distribution from the frames that do not have gross pitch error), 3) voicing decision error (the proportion of frames for which an incorrect voiced/unvoiced decision is made), 4) F0 frame error (the proportion of frames for which an error is made). According to the results, Praat's overall performance was shown to be better than that of other two algorithms, mainly because of Praat's better voicing detection. Praat was also demonstrated to benefit the most from using gender-adapted pitch settings.

Thus, understanding how errors in Praat's F0 tracking affect our measurements of F0 variability is of particular importance when interpreting findings in speech pathology literature.

Statistics Used to Describe F0 Variability

For most studies of dysarthria which use at least some kind of F0 measurement, there are usually two summary statistics presented: mean F0 and standard deviation (SD) of F0 in Hz. While these statistics are frequently used, neither measure is robust to outliers or deviations from the normal distribution (Wilcox & Rousselet, 2018). Since a speakers' F0 values may not necessarily be normally distributed or free from outlying values, mean and standard deviation of F0 may not provide the most accurate representation of the voice quality. Another issue related to F0 measurement is inherent inter-speaker variability related to the size and shape of the vocal folds. For example, in males with larger vocal folds, we will tend to observe smaller absolute F0 changes in Hz, regardless of whether the speaker is perceived as monotone.

For these reasons, additional statistics might be helpful in the analysis of F0 to more accurately capture our perceptions of monotonicity. Besides mean and SD, the following F0 parameters have been used in the field of speech disorders to describe F0 variability: F0 variation range, relative variation range, interquartile range, coefficient of variation (relative standard deviation), and semitone standard deviation (Bowen et al., 2013; Brabenec et al., 2017; Verkhodanova, 2021). The calculation of these measures is described in more detail in Table 1. Some of these measures are more robust to outliers or deviations from the normal distribution (e.g., interquartile range), while others are more effective in normalizing sex differences in the F0 range (e.g., coefficient of variation, semitone standard deviation).

Table 1

Statistical Indices for F0 Variability

Statistical index	Calculation procedure
Variation range	max(F0) – min(F0)
Relative variation range	range(F0) / mean(F0) * 100
Interquartile range	q75 – q25
Relative standard deviation	sd(F0) / mean(F0) * 100

Statistical index	Calculation procedure
Semitone standard deviation	39.86 * log10((mean(F0) + sd(F0)) / mean(F0))

To summarize, there are several reasons why traditional statistical approaches for describing F0 variability, such as the standard deviation of F0, may fail to represent our true perceptions of a speaker's monotonicity. Therefore, it is important to consider a wider range of statistical approaches when analyzing F0 in dysarthria and compare their performance in capturing perceived changes in speakers' expressiveness.

Current study

The purpose of this experimental study is twofold: (1) to evaluate the accuracy of F0 tracking in speakers with hypokinetic dysarthria and healthy control speakers and examine the degree to which tracking errors influence our ability to measure differences in F0 mean and standard deviation across groups, and (2) to compare how different statistical approaches to measuring F0 variability affect the relationship between acoustic measures and the perception of speech expressiveness. We hypothesized that excluding F0 tracking errors from acoustic analysis would have significant effect on the measurements of F0. We also hypothesized that the exclusion of these errors would result in stronger relationships between acoustic measures and the perception of speech expressiveness. It was hypothesized that at least some of the statistics used to describe F0 variability would predict listeners' ratings of expressiveness in the PD population, but certain measurements (i.e.

ones that normalized sex differences or were more robust to outliers) would be more strongly correlated with listeners' perceptual ratings.

Methods

Speech stimuli

Speech stimuli used in this study included the recordings of 18 native speakers of American English – nine speakers with Parkinson's disease (4 women and 5 men) aged 57 to 77 years old (M = 69.00, SD = 6.73) who were assessed as having monotone voices, and nine age- and sex-matched neurotypical control speakers (4 women and 5 men) aged 58 to 86 years old (M = 67.44, SD = 8.28) who were assessed as having normal speech prosody. The speakers with PD were evaluated by two experienced speech-language pathologists to have mild to moderate hypokinetic dysarthria. Both groups of speakers were prompted to read The Caterpillar passage (Patel et al., 2012) consisting of 16 sentences.

Recordings were made using a cardioid lavalier microphone placed approximately 20 cm from the speaker's mouth. Sound was recorded to a laptop using custom software via a Shure X2U XLR-to-USB adapter, with a sampling rate of 48 kHz and 16 bits off quantization. All the recordings were scaled to have the average intensity of 70 dB for consistency in the analysis and perceptual experiment.

Evaluating the accuracy of F0 tracking

To evaluate the accuracy of the F0 tracking in Praat, two scripts were compared. The first script utilized the standard Praat extraction algorithm to track F0 and extract the

F0 contour. The second script was designed to measure F0 without considering the segments that had been labeled as tracking errors. For the second customized script, pitch tracking errors were identified and labeled through manual examination of waveform and F0 contour. For this study, we labeled four main types of errors: 1) not tracking F0 during voiced segments; 2) tracking F0 within unvoiced segments/pauses; 3) tracking overtone frequencies; 4) tracking subharmonic frequencies. Mean F0 and standard deviation were then calculated for each sentence and each speaker using the results from both scripts. To extract F0 values, recommended Praat pitch settings were used – a range of 50-500 Hz for female speakers and a range of 50-300 Hz for male speakers. The floor value was set to 50 Hz for both groups of speakers to account for the vocal fry, according to pitch range recommendations (Boersma & Weeninck, 2024).

Measuring perception of speech expressiveness

Five sentences from each of the 18 speakers were selected to be included in the perceptual experiment based on two factors – all the sentences had to be declarative and of a similar length (M = 13.20, SD = 3.49). Speech stimuli consisted of 90 sentences in total.

Eleven speakers of American English (6 females and 5 males) aged 18 to 34 years old (M = 22.55, SD = 4.25) completed the listening task. Listeners were asked to answer the question of “How expressive is this speaker?” using a visual analog scale via a customized MATLAB program with extreme positions being “not expressive” and “very expressive”. Expressiveness was defined as “how dynamic and animated the speaker is". At the beginning of the perceptual experiment, listeners were presented with two example sentences to allow them to adjust the volume to the level they felt comfortable with and to familiarize themselves with the task interface. Then listeners were prompted to provide one rating following the presentation of each sentence. Sentences were the same for each listener, but the order of presentation was randomized.

To correlate the perception of expressiveness with F0 variability, several statistics were calculated – F0 range, F0 interquartile range, F0 relative range, F0 SD, F0 relative SD, and F0 semitone SD. These statistical indices were calculated according to the formulas in the Table 1.

Statistical analysis

All results were imported into R for statistical analysis. Separate mixed-effect regression models were constructed using the *lme4* package to examine each of our research questions. Firstly, we examined the effect of tracking errors on the F0 mean and F0 SD values for each sentence. F0 mean and F0 SD were input at dependent variables, with a fixed effect of script type (i.e. including or not including errors). Random intercepts were included for speaker and sentence, to account for repeated measures. Following this, we partitioned our data to examine the degree to which tracking errors influence our ability to measure differences in F0 mean and standard deviation across groups. One dataset included F0 tracking errors and the other did not. Two separate models were built to measure the fixed effects of sex and health status on F0 mean and F0 SD in each dataset. The same random intercepts for were included for speaker and sentence. The two-tailed significance level was set at $\alpha = .05$ for all models.

To answer our second research question, another set of models were used to examine listeners' ratings of expressiveness. In these models, the average rating of each sentence was the dependent variable, and there were fixed effects of health status (i.e. diagnosis of PD) and sex. Each model also included a fixed effect for one of the F0 variability measures. The effect of these variability measures was then compared across different models. For models in this set, random intercepts for speaker, sentence and listener were included to account for repeated measures. Again, the two-tailed significance level was set at **α** = .05 for all models.

Results

This study examined the accuracy of F0 tracking in speakers with hypokinetic dysarthria and healthy control speakers and the degree to which tracking errors influence our ability to measure differences in F0 mean and standard deviation across groups, and (2) to compare how different statistical approaches to measuring F0 variability affect the relationship between acoustic measures and the perception of speech expressiveness.

Accuracy of F0 tracking

Our first research question asked whether F0 tracking errors affected measures of mean F0 and SD F0. In the models which evaluated the influence of these tracking errors, F0 mean and F0 SD were dependent variables. They were regressed on sex, health status, and script type to assess differences due to F0 tracking errors (i.e. to compare the scripts that included and did not include the tracking errors).

Results demonstrated that script type (b = 2.36, SE = 1.08, $p < .05$) and sex (b = 53.27, SE = 13.94, $p < .01$) significantly influenced the measurements of F0 mean.

However, health status was shown to not significantly impact F0 mean (*b* = -8.03, SE = 13.85, $p > .05$). Overall, the model accounted for over 90% of the variance in F0 mean values ($R^2 = 0.906$). The model with F0 SD as the dependent variable showed that all three fixed effects of script type (*b* = 16.85, SE = 0.93, $p < .001$), sex (*b* = 13.70, SE = 1.96, $p < .001$), and health status (*b* = 4.71, SE = 1.95, $p < .05$) were significant predictors of F0 SD. However, the model of F0 SD did not account for as much variation in the data ($R^2 = 0.526$).

To understand how F0 tracking errors affect our ability to model group differences, our second analysis investigated the effects of speaker sex and the presence of dysarthria on F0 summary statistics. To accomplish this, we separated the data into two subsets, one that considered all F0 values, including errors, and one that excluded the tracking errors. We then built two separate models with either F0 mean or F0 SD as a dependent variable for each subset.

Table 2 reports the results for the models with F0 mean as a dependent variable. The results show that in both data subsets there was a significant effect of sex but no effect of dysarthria on the F0 mean measurement. The effect size of sex was higher when tracking errors were included (*b* = 59.03, SE = 13.01, $p < .001$) than when they were removed (*b* = 47.51, SE = 16.29, $p < .01$). However, the model that excluded tracking errors had a better overall fit ($R^2 = 0.97$, AIC = 2094.24) than the model including these errors ($R^2 = 0.95$, AIC = 2182.88), with lower AIC scores indicating that it would have improved predicative power when applied to new data. The results are presented in Fig. 1.

Table 2

Mixed-effects regression of F0 mean on sex and health status

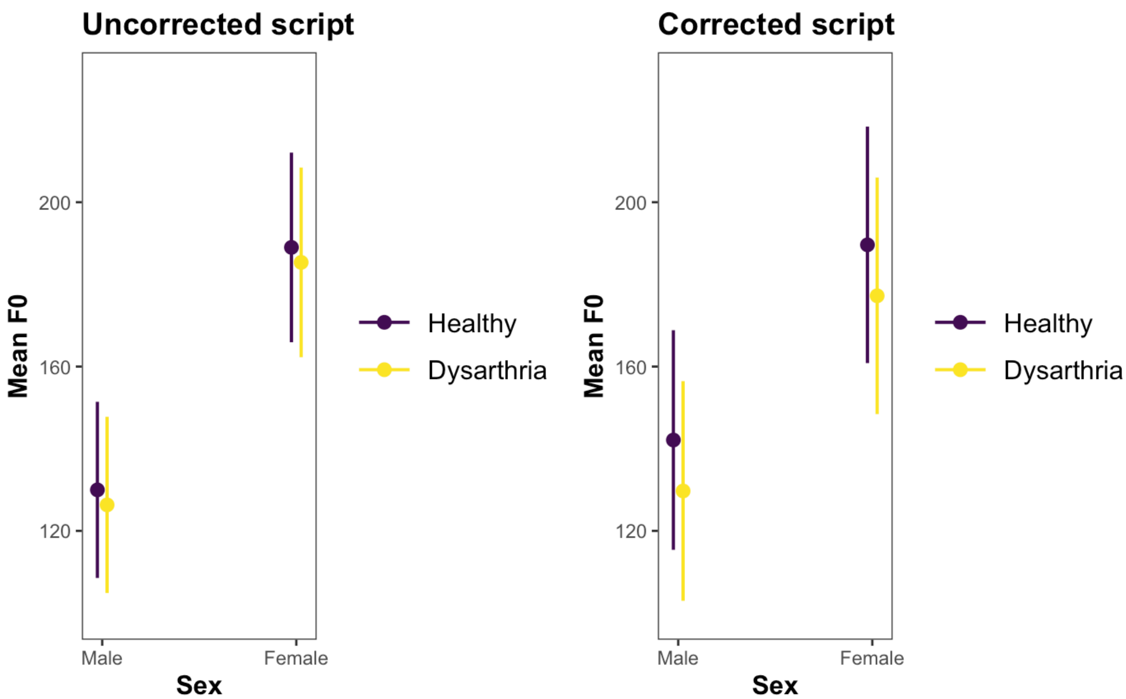
	Data containing tracking errors		Data excluding tracking errors	
Coefficient	Estimates	95% CI	Estimates	95% CI
Intercept	129.98 ***	108.53 – 151.42	142.11 ***	115.38 – 168.85
Sex (Female)	59.03 ***	33.41 – 84.64	47.51 **	15.44 – 79.58
Health status (Dysarthria)	-3.65	-29.11 – 21.80	-12.40	-44.28 – 19.47
Random Effects				
σ^2	80.69		57.19	
τ_{00}	747.39 $_{Speaker}$		1176.20 $_{Speaker}$	
	26.15 $_{Sentence}$		14.82 $_{Sentence}$	
ICC	0.91		0.95	
N	18 $_{Speaker}$		18 $_{Speaker}$	
	16 $_{Sentence}$		16 $_{Sentence}$	
Observations	287		287	
Marginal R^2 / Conditional R^2	0.504 / 0.953		0.324 / 0.969	

Note. CI = confidence interval.

* p<0.05 ** p<0.01 *** p<0.001

Figure 1

Effects of sex and health status on F0 mean for two script types. Uncorrected script = data including tracking errors. Corrected script = data excluding tracking errors.



In Table 3, the results for the models with F0 SD as a dependent variable are summarized. The results demonstrate that in both data subsets there was a significant effect of sex on F0 SD, but the effect was larger when tracking errors were included (*b* = 22.00, SE = 2.64, p < .001) than when they were removed (*b* = 5.40, SE = 1.93, p < .01). Interestingly, the effect of dysarthria on F0 SD values was only statistically significant when tracking errors were included in the dataset (*b* = 6.63, SE = 2.62, p < .05). When errors were removed, the effect of dysarthria was greatly reduced (*b* = 2.79, SE = 1.91, p > .05). However, again, the model that excluded tracking errors demonstrated a better fit with lower AIC values (R^2 = 0.51, AIC = 1806.88) than the model that contained uncorrected F0 data (R^2 = 0.51, AIC = 2311.79). These results are also presented in Fig. 2.

Table 3

Mixed-effects regression of F0 SD on sex and health status

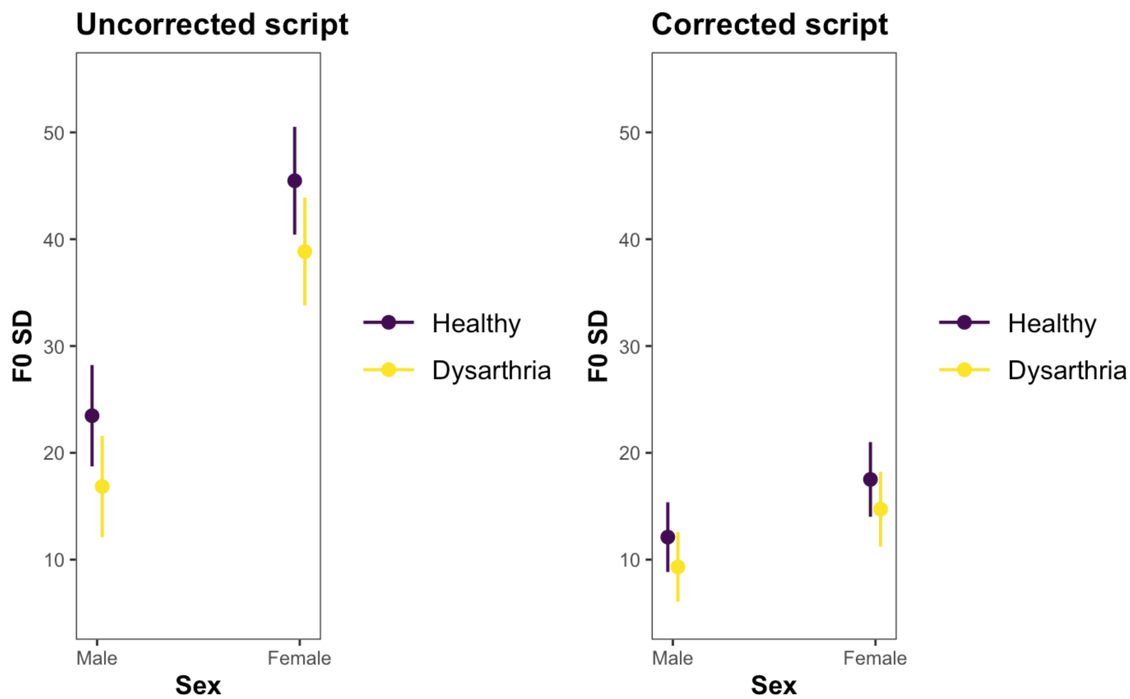
	Data containing tracking errors		Data excluding tracking errors	
Coefficient	Estimates	95% CI	Estimates	95% CI
Intercept	23.47 ***	18.72 – 28.22	12.11 ***	8.85 – 15.37
Sex (Female)	22.00 ***	16.81 – 27.19	5.40 **	1.61 – 9.19
Health status (Dysarthria)	-6.63 *	-11.79 – -1.47	-2.79	-6.56 – 0.98
Random Effects				
σ^2	163.74		25.82	
τ_{00}	20.65 $_{Speaker}$		14.87 $_{Speaker}$	
	16.00 $_{Sentence}$		2.92 $_{Sentence}$	
ICC	0.18		0.41	
N	18 $_{Speaker}$		18 $_{Speaker}$	
	16 $_{Sentence}$		16 $_{Sentence}$	
Observations	287		287	
Marginal R^2 / Conditional R^2	0.396 / 0.506		0.174 / 0.511	

Note. CI = confidence interval.

* p<0.05 ** p<0.01 *** p<0.001

Figure 2

Effects of sex and health status on F0 SD for two script types. Uncorrected script = data including tracking errors. Corrected script = data excluding tracking errors.



Effects of F0 variability on the perception of expressiveness

To answer the second research question of whether using different F0 variability statistics can improve our ability to predict perceptual ratings of expressiveness, we build 12 separate mixed effects linear models – six models for each of the following F0 variability measures: range, relative range, interquartile range, SD, relative SD, SD in semitones, from the two data subsets: including all F0 values with tracking errors vs. data excluding tracking errors. In these models, perceptual ratings were entered as a dependent variable; sex, dysarthria, and the F0 variability measure of interest were entered as fixed effects, while speaker, sentence, and listener were included as random intercepts.

Table 4 summarizes the results for the data subset that included tracking errors. The results show that for the F0 values which include Praat F0 tracking errors, only two out of six F0 variability statistics have a statistically significant effect on listener ratings of expressiveness – F0 relative SD (b = 0.019, SE = -0.01, p < .05) and F0 SD in semitones (b = 0.144, SE = -0.07, p < .05). As expected, across all models there is a significant effect of dysarthria on the perception of speech expressiveness.

Table 4

Fixed effects of six different F0 variability statistics on listener ratings of expressiveness. These models are based on the F0 values that contain tracking errors. Fixed effects of speaker sex and dysarthria are also included. Standard errors are included in brackets.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Perceptual ratings of expressiveness					
F0 range	-0.032					
	(-0.108)					
F0 IQ range		0.169				
		(-0.096)				
F0 relative range			-0.057			
			(-0.097)			
F0 SD				0.177		
				(-0.098)		
F0 relative SD					0.185*	
					(-0.092)	
F0 SD in semitones						0.193*
						(-0.094)
Dysarthria	-2.504***	-2.485***	-2.516***	-2.492***	-2.420***	-2.417***
	(-0.515)	(-0.53)	(-0.509)	(-0.552)	(-0.56)	(-0.562)
Sex	0.934	0.925	0.941	0.927	0.911	0.911
	(-0.518)	(-0.533)	(-0.512)	(-0.555)	(-0.562)	(-0.564)
AIC	4025.199	4022.425	4025.178	4022.398	4021.776	4021.599

*** = p < 0.001; ** = p < 0.01; * = p < 0.05

As presented in Table 5, for the F0 values extracted after excluding tracking errors, we found four significant fixed effects – F0 IQ range (*b* = 0.016, SE = -0.006, p < .01), F0 SD (*b* = 0.015, SE = -0.006, p < .01), F0 relative SD (*b* = 0.028, SE = -0.01, p < .01), and F0 SD in semitones (*b* = 0.213, SE = -0.07, p < .01). Again, as expected, there was also a significant effect of dysarthria on expressiveness ratings for all six models. AIC scores demonstrate that model fit is better for the models which were run using the corrected F0 values.

Table 5

Fixed effects of six different F0 variability statistics on listener ratings of expressiveness. These models are based on the F0 values that do not contain tracking errors. Fixed effects of speaker sex and dysarthria are also included. Standard errors are included in brackets.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Perceptual ratings of expressiveness					
F0 range	0.159					
	(-0.105)					
F0 IQ range		0.253**				
		(-0.097)				
F0 relative range			0.114			
			(-0.094)			
F0 SD				0.270**		
				(-0.099)		
F0 relative SD					0.268**	
					(-0.091)	
F0 SD in semitones						0.281**
						(-0.093)
Dysarthria	-2.511***	-2.329***	-2.519***	-2.442***	-2.445***	-2.434***
	(-0.524)	(-0.51)	(-0.527)	(-0.516)	(-0.529)	(-0.53)
Sex	0.726	0.638	0.836	0.568	0.692	0.674
	(-0.544)	(-0.521)	(-0.536)	(-0.535)	(-0.538)	(-0.54)
AIC	4023.191	4018.682	4024.165	4018.154	4017.053	4016.542

*** = p < 0.001; ** = p < 0.01; * = p < 0.05

Discussion

The current study explored measures of F0 variability within two groups of speakers (speakers with hypokinetic dysarthria and neurotypical control speakers). The first aim was to examine to what degree F0 tracking errors would influence F0 summary statistics (such as F0 mean and F0 SD) and how F0 tracking errors might affect our ability to model group differences across speakers. The second aim was to investigate the relationship between F0 variability statistics and perceptual ratings of speech expressiveness. Overall, the study found that recordings from both speakers with dysarthria and age-matched, neurotypical adults were prone to F0 tracking errors when using Praat's standard pitch settings. There was a significant effect of tracking errors and sex on the F0 mean, and a significant effect of tracking errors, sex, and dysarthria on F0 standard deviation. Further investigation, which involved separate analysis of the F0 data that included tracking errors and the F0 data without these values, demonstrated a significant effect of sex on F0 mean and F0 SD in both data subsets. However, only the model that included tracking errors found a significant effect of dysarthria on F0 SD. The analysis of listeners' expressiveness ratings showed that F0 SD and F0 SD in semitones were significant predictors of speech expressiveness when tracking errors were included in the calculation of F0 statistics. However, when the tracking errors were excluded, a larger number of F0 metrics could be used to predict listener ratings. Specifically, F0 SD, F0 SD in semitones, F0 relative SD, and F0 IQ range were significant predictors of speech expressiveness when F0 tracking errors were excluded from the analysis. These results are discussed in more detail below.

The first part of this project focused on exploring how tracking errors affect the statistics commonly reported in the motor speech literature when describing dysarthria. The results established that a failure to control for F0 tracking errors will result in statistically significant differences in summary F0 statistics. This project specifically highlighted differences in mean F0 and F0 SD values, because these are the statistics most commonly examined in the speech pathology literature. These findings provide further evidence that data preprocessing, such as excluding outliers or correcting F0 tracking errors, might be needed to get more accurate results for F0 summary statistics. Failure to remove tracking errors resulted in higher measures of both F0 mean and F0 SD. While the difference appeared relatively minor for the F0 mean (the estimate of the effect was only 2.36 Hz), it was quite large for F0 SD (which had an estimated increase of 16.85 Hz when tracking errors were included). This finding is not unexpected, as tracking errors tend to result in extreme outlier values that occur in both directions (i.e. inaccurate F0 values that are too low and F0 values that are too high). When outliers occur in both directions, they may cancel each other out in terms of their impact on the mean. However, in case of SD, the presence of outliers in both directions increases the overall spread of the data, leading to a greater value of SD. In summary, these results suggest that it might be beneficial to control for the tracking errors especially when providing F0 summary statistics and comparing F0 SD values across speaker groups.

Running separate mixed-effects models for the dataset that included tracking errors and the dataset that removed them showed some expected and unexpected results. Firstly, our analysis showed a significant influence of sex on F0 mean and F0 SD – both measures were significantly higher for females than for males regardless of whether the data contained tracking errors. These results are not surprising since females overall have higher pitch and pitch range than males (Simpson, 2009). Unexpectedly, we found that the effect of dysarthria was significant only when tracking errors were included. This suggests that the tracking errors might erroneously inflate the group separation between neurotypical and dysarthric speech. Previous literature has suggested that hypokinetic dysarthria has significant effect on F0 SD and has showed higher F0 variability in control speakers (Bowen et al., 2013; Skodda et al., 2011). However, it has never been suggested that F0 tracking errors might be partially responsible for this outcome. These errors put analyses at high risk for false positive results and thus, if they are removed, the true group differences in F0 SD may not be as large or obvious as previously thought.

To compare model fit, we used Akaike Information Criterion (AIC). For both models with F0 mean and F0 SD as a dependent variable, the model fit improved when using the data without tracking errors. As discussed in the study by Cavanaugh & Neath (2019), improved AIC scores indicate that models for F0 mean and F0 SD prediction built using data without F0 tracking errors
will more accurately describe future data and will allow for better replicability of the outcome.

The second part of this thesis investigated the relationship between F0 variability and speech expressiveness. Findings revealed that, first, dysarthric speakers were in fact rated as less expressive than healthy speakers by at least 20%. This finding supports the previous literature on the acoustic-perceptual relationships in the dysarthric population which indicates that speakers with PD exhibit decreased emotional expression, even within relatively neutral (i.e. non-emotional) speaking contexts (Anzuino et al., 2023; Caekebeke et al., 1991; Möbes et al., 2008; Pell et al., 2006).

In examining F0 variability, we explored the relationship between six different F0 variability statistics (SD, relative SD, SD in semitones, range, interquartile range, relative range) and expressiveness scores. Nearly all the F0 variability statistics examined showed a positive correlation with expressiveness scores meaning that sentences with greater F0 range and greater SD were consistently rated as more expressive by the listeners, even though the effect was not statistically significant for all measures. The only exception was F0 range and relative range when tracking errors were included in the calculation of these values. In these cases, there was a negative correlation with the perceptual ratings. This finding is likely explained by the fact that F0 range will be especially sensitive to outlying values. If tracking errors resulting from subharmonic and overtone frequencies are the highest and lowest values present in a sentence, an analysis of F0 range (i.e. maximum F0 – minimum F0) would be based solely on these errors, and thus not include any real F0 values produced by the speaker. This interpretation is supported by the results from the data that excluded tracking errors, where both F0 range and relative range showed positive correlation with expressiveness scores, indicating that different minimum and maximum values were being selected.

Although the results of the analysis showed that there was a consistently positive effect of F0 variability on the perception of expressiveness, not every F0 variability statistic had a significant relationship with listener ratings. For the F0 statistics calculated from data that contained tracking errors, we found a significant effect of relative SD and SD in semitones. These two effects, together with the effect of interquartile range and F0 SD, were also found to be significant when tracking errors were removed. It appears that standard deviation, its variations, and interquartile range are appropriate statistics to use when trying to predict perceptual ratings of expressiveness on the data without F0 tracking errors. These effects were strongest when tracking errors were excluded from the analysis. The reason why some F0 variability measures performed better than others might be the nature of those measures. F0 relative SD and F0 SD in semitones, which were shown to be significant predictors in both sets of data (with/without errors), each help control for the differences in vocal range due to sex. SD in semitones was shown to have the greatest effect on perceptual ratings of expressiveness in our analysis, which might also be related to the fact that this measure better reflects our perception of changes in F0 frequencies. The human ear is more sensitive to changes in the lower end of the frequency spectrum, i.e. the just-noticeable difference between lower frequencies is smaller than the just-noticeable difference in higher frequencies. Measuring F0 SD in semitones captures this difference in perception since each semitone represents a consistent increase in pitch that is logarithmically, and not linearly, related to frequency. Thus, measuring standard deviation of F0 in semitones appears to correlate higher with the perception of monotonicity.

These observations once again highlight the importance of, on the one hand, excluding F0 tracking errors from the analysis when calculating summary and variability statistics and, on the other hand, choosing appropriate statistical measures to describe the data, especially if working with data which were not examined for the presence of outliers.

Limitations and Future Directions

The study included nine speakers with PD and nine neurotypical healthy controls. The exploration of a bigger sample might be beneficial for further understanding of certain effects such as, for instance, the effect of dysarthria on the SD which was significant only when tracking errors were included in the calculation F0 SD. For future research, it will also be helpful to add another level of labeling when identifying the type of F0 tracking errors, so we can better understand if there are differences in the type of errors made when analyzing neurotypical and dysarthric speech. This type of labeling will allow to explore further which errors are more frequent in which group of speakers and have better understanding of why we can better distinguish between dysarthric and healthy speakers when tracking errors are included in the data. Additionally, the speakers with PD selected for this study were assessed to have either mild or moderate dysarthria. Future studies might also include speakers with severe dysarthria in the analysis to examine whether there will be relationships between F0 variability and perceived expressiveness for this group as well.

One final issue in the current study was the combined analysis the modal register together with creaky voice. Low F0 values in creaky voice are not a result of tracking errors. However, the presence of creaky voice is likely to affect the distribution of F0 values across sentences—potentially increasing the range and variability in F0 values. For measures of F0 variability to be meaningful, we must be clear on what type of F0 variation we are attempting to index. In the speech disorder literature, the perception of monotonicity is typically related to a flattening of the F0 contour in the *modal* voice register, while the presence of *vocal fry or creaky voice* is considered a separate feature of voice quality (Duffy, 2013). However, if F0 variation is measured without removing creaky voice, a flattening of F0 in the modal register might be difficult to detect. For example, if a speaker has a flat F0 contour in their modal range, but high levels of vocal fry, they may appear to have large variability in F0 values. In fact, there is evidence that the distribution of the F0 values may be bimodal in both healthy and speakers with dysarthria since both populations use modal register and creaky voice. Dorreen (2017), which focused on finding the most efficient F0 parameters for forensic speaker comparison, demonstrated that all bilingual speakers had a bimodal distribution of F0 in at least one of the languages they spoke. They concluded that it is important to take this distribution type into consideration since bimodality with significant amounts of creak phonation will shift the F0 mean – one of the most widely used statistics for describing F0 – downwards. To more accurately explain F0 values, creak phonation and modal phonation were analyzed separately with the first antimode (point that has the least frequent F0) used as a splitting point in this study. In the dysarthria literature, it would be important to more closely explore F0 distributions before conducting similar analyses to make sure that all the speakers have the same distribution of F0 values and establish whether bimodal distributions are occurring. If some speakers have a bimodal distribution of F0 values and others do not, it might be useful to look only at the modal register, excluding creaky voice and lower frequencies from the analysis.

Conclusions

As previously demonstrated in the literature, listeners regard speakers with PD as being less expressive, which may reduce their likelihood of engaging in conversation with them. This study demonstrated that measures of F0 variation can be used to predict these reductions in speech expressiveness. However, we also demonstrated that it can be challenging to accurately measure F0 variation, since common methods of F0 tracking are prone to errors which significantly affect F0 statistics. The results of this study offer insight into how we can best to capture changes in F0 that are relevant to listeners' perceptual judgements. By focusing on these acoustic measurements, we may be able to more objectively assess improvements in speech expressiveness following speech treatment.

References

Anzuino, I., Baglio, F., Pelizzari, L., Cabinio, M., Biassoni, F., Gnerre, M., Blasi, V., Silveri, M. C., & Di Tella, S. (2023). Production of emotions conveyed by voice in Parkinson's disease: Association between variability of fundamental frequency and gray matter volumes of regions involved in emotional prosody. *Neuropsychology*, *37*(8), 883–894. https://doi.org/10.1037/neu0000912

Boersma, P., & Weeninck, D. (2024). *Praat: Doing phonetics by computer (Computer program).* https://www.praat.org/

Bowen, L. K., Hands, G. L., Pradhan, S., & Stepp, C. E. (2013). Effects of Parkinson's Disease on Fundamental Frequency Variability in Running Speech. *Journal of Medical Speech-Language Pathology*, *21*(3), 235–244.

Brabenec, L., Mekyska, J., Galaz, Z., & Rektorova, I. (2017). Speech disorders in Parkinson's disease: Early diagnostics and effects of medication and brain stimulation. *Journal of Neural Transmission*, *124*(3), 303–334. https://doi.org/10.1007/s00702-017-1676-0

Brockmann-Bauser, M., & de Paula Soares, M. F. (2023). Do We Get What We Need from Clinical Acoustic Voice Measurements? *Applied Sciences*, *13*(2), Article 2. https://doi.org/10.3390/app13020941

Caekebeke, J. F., Jennekens-Schinkel, A., Linden, M. E. van der, Buruma, O. J., & Roos, R. A. (1991). The interpretation of dysprosody in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *54*(2), 145–148. https://doi.org/10.1136/jnnp.54.2.145

Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics*, *11*(3), e1460. https://doi.org/10.1002/wics.1460

Chenausky, K., MacAuslan, J., & Goldhor, R. (2011). Acoustic Analysis of PD Speech. *Parkinson's Disease*, *2011*, e435232. https://doi.org/10.4061/2011/435232

Chiu, Y.-F., Forrest, K., & Loux, T. (2019). Relationship Between F2 Slope and Intelligibility in Parkinson's Disease: Lexical Effects and Listening Environment.

American Journal of Speech-Language Pathology, *28*(2S), 887–894. https://doi.org/10.1044/2018_AJSLP-MSC18-18-0098

Dorreen, K. (2017). *Fundamental Frequency Distributions of Bilingual Speakers in Forensic Speaker Comparison*. http://hdl.handle.net/10092/13549

Duffy, J. R. (2013). *Motor Speech Disorders: Substrates, Differential Disgnosis, and Management* (3rd ed.).

Exner, A. H. (2019). *The Effects of Speech Tasks on the Prosody of People with Parkinson Disease* [Thesis, Purdue University Graduate School]. https://doi.org/10.25394/PGS.9936275.v1

Gnerre, M., Malaspina, E., Di Tella, S., Anzuino, I., Baglio, F., Silveri, M. C., & Biassoni, F. (2023). Vocal Emotional Expression in Parkinson's Disease: Roles of Sex and Emotions. *Societies*, *13*(7), Article 7. https://doi.org/10.3390/soc13070157

Goberman, A., Coelho, C., & Robb, M. (2005). Prosodic characteristics of Parkinsonian speech: The effect of levodopa-based medication. *Journal of Medical Speech-Language Pathology*. https://www.semanticscholar.org/paper/Prosodic-characteristics-of-Parkinsonian-speech%3A-of-Goberman-Coelho/05fbed6f5bba737a8aa2e260a61696e935554fb7

Goberman, A. M., & Blomgren, M. (2008). Fundamental Frequency Change During Offset and Onset of Voicing in Individuals with Parkinson Disease. *Journal of Voice*, *22*(2), 178–191. https://doi.org/10.1016/j.jvoice.2006.07.006

Holmes, R. J., Oates, J. M., Phyland, D. J., & Hughes, A. J. (2000). Voice characteristics in the progression of Parkinson's disease. *International Journal of Language & Communication Disorders*, *35*(3), 407–418. https://doi.org/10.1080/136828200410654

Jaywant, A., & Pell, M. D. (2010). Listener impressions of speakers with Parkinson's disease. *Journal of the International Neuropsychological Society*, *16*(1), 49–57. https://doi.org/10.1017/S1355617709990919

Jiménez-Jiménez, F. J., Gamboa, J., Nieto, A., Guerrero, J., Orti-Pareja, M., Molina, J. A., García-Albea, E., & Cobeta, I. (1997). Acoustic voice analysis in untreated patients with Parkinson’s disease. *Parkinsonism & Related Disorders*, *3*(2), 111–116. https://doi.org/10.1016/s1353-8020(97)00007-2

Kim, Y., Kent, R. D., & Weismer, G. (2011). An Acoustic Study of the Relationships Among Neurologic Disease, Dysarthria Type, and Severity of Dysarthria. *Journal of Speech, Language, and Hearing Research*, *54*(2), 417–429. https://doi.org/10.1044/1092-4388(2010/10-0020)

Kim, Y., Weismer, G., Kent, R. D., & Duffy, J. R. (2009). Statistical Models of F2 Slope in Relation to Severity of Dysarthria. *Folia Phoniatrica et Logopaedica*, *61*(6), 329–335. https://doi.org/10.1159/000252849

Kovac, D., Mekyska, J., Aharonson, V., Harar, P., Galaz, Z., Rapcsak, S., Orozco-Arroyave, J. R., Brabenec, L., & Rektorova, I. (2024). Exploring digital speech biomarkers of hypokinetic dysarthria in a multilingual cohort. *Biomedical Signal Processing and Control*, *88*, 105667. https://doi.org/10.1016/j.bspc.2023.105667

Lansford, K. L., & Liss, J. M. (2014). Vowel Acoustics in Dysarthria: Mapping to Perception. *Journal of Speech, Language, and Hearing Research*, *57*(1), 68–80. https://doi.org/10.1044/1092-4388(2013/12-0263)

Looze, C. D., Ghio, A., Scherer, S., Pouchoulin, G., & Viallet, F. (2012). *Automatic analysis of the prosodic variations in Parkinsonian read and semi-spontaneous speech*. 71–74. https://doi.org/10.21437/SpeechProsody.2012-21

Marcoux, K., & Ernestus, M. (2019). Pitch in native and non-native Lombard speech. *19th International Congress of Phonetic Sciences (ICPhS 2019)*, 2605–2609.

Miller, N., Noble, E., Jones, D., & Burn, D. (2006). Life with communication changes in Parkinson’s disease. *Age and Ageing*, *35*(3), 235–239. https://doi.org/10.1093/ageing/afj053

Möbes, J., Joppich, G., Stiebritz, F., Dengler, R., & Schröder, C. (2008). Emotional speech in Parkinson’s disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *23*(6), 824–829. https://doi.org/10.1002/mds.21940

Müller, J., Wenning, G. K., Verny, M., McKee, A., Chaudhuri, K. R., Jellinger, K., Poewe, W., & Litvan, I. (2001). Progression of Dysarthria and Dysphagia in Postmortem-Confirmed Parkinsonian Disorders. *Archives of Neurology*, *58*(2), 259. https://doi.org/10.1001/archneur.58.2.259

Patel, R., Connaghan, K., Franco, D., Edsall, E., Forgit, D., Olsen, L., Ramage, L., Tyler, E., & Russell, S. (2012). “The Caterpillar”: A Novel Reading Passage for Assessment of Motor Speech Disorders. *American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association*, *22*. https://doi.org/10.1044/1058-0360(2012/11-0134)

Pell, M. D., Cheang, H. S., & Leonard, C. L. (2006). The impact of Parkinson’s disease on vocal-prosodic communication from the perspective of listeners. *Brain and Language*, *97*(2), 123–134. https://doi.org/10.1016/j.bandl.2005.08.010

Pitcairn, T. K., Clemie, S., Gray, J. M., & Pentland, B. (1990). Impressions of parkinsonian patients from their recorded voices. *British Journal of Disorders of Communication*, *25*(1), 85–92. https://doi.org/10.3109/13682829009011965

Rodero, E. (2011). Intonation and Emotion: Influence of Pitch Levels and Contour Type on Creating Emotions. *Journal of Voice*, *25*(1), e25–e34. https://doi.org/10.1016/j.jvoice.2010.02.002

Rodríguez-Pérez, P., Fraile, R., García-Escrig, M., Sáenz-Lechón, N., Gutiérrez-Arriola, J. M., & Osma-Ruiz, V. (2019). A transversal study of fundamental frequency contours in parkinsonian voices. *Biomedical Signal Processing and Control*, *51*, 374–381. https://doi.org/10.1016/j.bspc.2019.02.021

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621–640. https://doi.org/10.1111/j.1749-818X.2009.00125.x

Skodda, S., Grönheit, W., & Schlegel, U. (2011). Intonation and Speech Rate in Parkinson’s Disease: General and Dynamic Aspects and Responsiveness to Levodopa Admission. *Journal of Voice*, *25*(4), e199–e205. https://doi.org/10.1016/j.jvoice.2010.04.007

Strangert, E., & Gustafson, J. (2008). What makes a good speaker? Subject ratings, acoustic measurements and perceptual evaluations. *Interspeech 2008*, 1688–1691. https://doi.org/10.21437/Interspeech.2008-368

Strömbergsson, S. (2016). *Today’s Most Frequently Used F0 Estimation Methods, and Their Accuracy in Estimating Male and Female Pitch in Clean Speech* (p. 529). https://doi.org/10.21437/Interspeech.2016-240

Tjaden, K., Richards, E., Kuo, C., Wilding, G., & Sussman, J. (2013). Acoustic and perceptual consequences of clear and loud speech. *Folia Phoniatrica et Logopaedica: Official Organ of the International Association of Logopedics and Phoniatrics (IALP)*, *65*(4), 214–220. https://doi.org/10.1159/000355867

Tjaden, K., & Wilding, G. E. (2004). Rate and Loudness Manipulations in Dysarthria. *Journal of Speech, Language, and Hearing Research*, *47*(4), 766–783. https://doi.org/10.1044/1092-4388(2004/058)

Traunmüller, H., & Eriksson, A. (1995). The perceptual evaluation of F0 excursions in speech as evidenced in liveliness estimations. *The Journal of the Acoustical Society of America*, *97*(3), 1905–1915. https://doi.org/10.1121/1.412942

Tykalova, T., Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2014). Acoustic Investigation of Stress Patterns in Parkinson's Disease. *Journal of Voice*, *28*(1), 129.e1-129.e8. https://doi.org/10.1016/j.jvoice.2013.07.001

Van Der Bruggen, S., De Letter, M., & Rietveld, T. (2023). Effects of near-monotonous speech of persons with Parkinson's disease on listening effort and intelligibility. *Clinical Linguistics & Phonetics*, *0*(0), 1–14. https://doi.org/10.1080/02699206.2023.2272032

Verkhodanova, V. (2021). *Acoustic change over time in speech of one bilingual individual with Parkinson's disease*. https://doi.org/10.17605/OSF.IO/9BSQY

Vogel, A. P., Maruff, P., Snyder, P. J., & Mundt, J. C. (2009). Standardization of pitch range settings in voice acoustic analysis. *Behavior Research Methods*, *41*(2), 318–324. https://doi.org/10.3758/BRM.41.2.318

Weismer, G., Jeng, J.-Y., Laures, J. S., Kent, R. D., & Kent, J. F. (2001). Acoustic and Intelligibility Characteristics of Sentence Production in Neurogenic Speech Disorders. *Folia Phoniatrica et Logopaedica*, *53*(1), 1–18. https://doi.org/10.1159/000052649

Wilcox, R. R., & Rousselet, G. A. (2018). A Guide to Robust Statistical Methods in Neuroscience. *Current Protocols in Neuroscience*, *82*, 8.42.1-8.42.30. https://doi.org/10.1002/cpns.41

Yorkston, K., Baylor, C., & Britton, D. (2017). Speech Versus Speaking: The Experiences of People With Parkinson's Disease and Implications for

Intervention. *American Journal of Speech-Language Pathology*, *26*(2 Suppl), 561–568. https://doi.org/10.1044/2017_AJSLP-16-0087