ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Contributions of Speech Timing and Articulatory Precision to Listener Perceptions of Intelligibility and Naturalness in Parkinson's Disease

Saul A. Frankford,[a] Alex Estrada,[a] and Cara E. Stepp[a]

[a] Boston University, MA

ARTICLE INFO

Article History:
Received December 22, 2023
Revision received April 9, 2024
Accepted June 6, 2024

Editor-in-Chief: Rita R. Patel
Editor: Jessica E. Huber

https://doi.org/10.1044/2024_JSLHR-23-00802

ABSTRACT

Purpose: Parkinson's disease (PD) results in hypokinetic dysarthria in as many as 90% of cases. Among the most distinctive features of hypokinetic dysarthria are atypical timing and articulatory imprecision in speech production. Here, we examined the contributions of perceived speech timing typicality and articulatory precision, both on their own and while controlling for the other, on intelligibility and naturalness in speakers with PD.
Method: Twenty speakers with PD and four healthy older adults read aloud the first paragraph of the Rainbow Passage. Twenty inexperienced listeners with typical hearing listened to these recordings and rated intelligibility, naturalness, timing typicality, and articulatory precision using separate visual analog scales. Ratings were averaged across listeners and entered into linear regression models with intelligibility and naturalness as dependent variables and timing typicality and articulatory precision as independent variables in each.
Results: Articulatory precision, but not timing typicality, was positively correlated with intelligibility on its own, but neither was associated with intelligibility after accounting for the other. Both timing typicality and articulatory precision were positively correlated with naturalness on their own as well as after controlling for the other variable.
Conclusion: These results contribute to the overall understanding of speech factors associated with intelligibility and naturalness in speakers with PD and indicate that considering the unique contributions of related perceptual constructs may provide more information than bivariate relationships alone.

Parkinson's disease (PD) is the second most common neurodegenerative disease (Tysnes & Storstein, 2017) and affects approximately 1 million individuals in the United States (Marras et al., 2018). Although principally characterized by its effects on gross motor skills such as gait, PD results in speech impairment (typically, hypokinetic dysarthria) in as many as 90% of individuals (Ho et al., 1999; Logemann et al., 1978), leading to reduced communicative participation (McAuliffe et al., 2017), negative mental health outcomes (Moya-Galé & Levy, 2019), and reduced quality of life (Miller et al., 2006; Moya-Galé & Levy, 2019). Despite this high prevalence and significant impact, current pharmacological and neurosurgical treatments do not adequately address speech symptoms (Fabbri et al., 2017; Ho et al., 2008; Moya-Galé & Levy, 2019). Popular evidence-based speech therapy protocols for people with PD (PwPD) are often a "one size fits all," typically focusing on addressing hypophonia (Pu et al., 2021). Given that hypokinetic dysarthria affects not only loudness and pitch variation but also articulatory precision and speaking rate, rhythm, and fluency (Darley et al., 1969), an exclusive focus on one speech characteristic may not be sufficient. Determining the extent to which features of hypokinetic speech contribute to clinically relevant outcomes may inform the development of more comprehensive therapies.

Correspondence to Saul A. Frankford: saulf@bu.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Two such outcomes that are commonly evaluated in individuals with dysarthria and used as targets in speech therapy are intelligibility and naturalness (Yorkston et al., 2010). *Intelligibility* is typically defined as the degree to which a speaker can be understood by a listener (Kent et al., 1989) and may be used as a measure of speech impairment and a target for improvement in speech therapy (Klopfenstein et al., 2020). *Naturalness*, on the other hand, is related to how speech sounds in comparison to the expectations of the listener, given a particular context. In this regard, naturalness may be more closely aligned with the prosodic components of a speech utterance. The main components of prosody, namely, changes in loudness, pitch, and timing (Cole, 2015), serve to provide additional context about what is being said, including the intended focus or intention of the utterance (e.g., statement vs. question), the speaker's emotions (e.g., excited vs. scared), and hidden or nonliteral meanings (e.g., sarcasm).

Among the most distinctive features of hypokinetic dysarthria are abnormalities in the production of speech timing (Darley et al., 1969; Ho et al., 1999; Logemann et al., 1978), including fast or slow rate, intermittent rushes of speech, increased or decreased pause time, and disfluencies (Duffy, 2020), with increasing prevalence in those with greater speech impairment (Ho et al., 1999). Another highly characteristic feature of hypokinetic dysarthria in PD is reduced articulatory precision, including "slurring, inadequate sharpness, distortions, and lack of crispness" as well as "clumsiness in going from one sound to another" (Darley et al., 1969, p. 268; Duffy, 2020; Plowman-Prine et al., 2009, p. 144). This auditory-perceptual construct is often associated with terms such as speech *clarity* (Tjaden, 2000; Whitfield & Goberman, 2017), *distinctiveness* (Knowles et al., 2021), or *accuracy* (Chiu et al., 2021; Jiao et al., 2017), as well as with articulatory and acoustic measures such as range of articulator movement (*articulatory undershoot*; Duffy, 2020), degree of closure on stop consonants (Ackermann & Ziegler, 1991; Logemann & Fisher, 1981), and vowel space area (Fletcher et al., 2017; Rusz et al., 2013), each of which is atypical in PwPD (Ackermann & Ziegler, 1991; Kearney et al., 2017; Logemann & Fisher, 1981; Rusz et al., 2013; Skodda et al., 2011; Tjaden & Wilding, 2004; Walsh & Smith, 2012; Weismer, 1984; Whitfield & Goberman, 2014).

While representing distinct perceptual variables, articulatory precision and speech timing may be directly related to one another in speakers with PD. Fletcher et al. (2017) found that auditory-perceptual ratings of articulatory precision in speakers with various dysarthria subtypes were significantly related to acoustic rate and rhythm measures. Furthermore, it has been suggested that articulatory imprecision contributes to a perception of increased speaking rate in PwPD (Netsell et al., 1975; Tjaden, 2000), although there are currently no published data from PwPD to support this. Acoustically, habitual speaking rate is significantly related to articulatory distinctness in both healthy older adults and PwPD, such that people who speak more slowly tend to have a larger vowel space (Fletcher et al., 2015; Knowles et al., 2021) and more distinct voice onset times between voiced and voiceless stop consonants (Knowles et al., 2021). In studies in which rate was explicitly modulated, vowel space increased in slow compared to habitual conditions in typical speakers (Knowles et al., 2021; Tjaden & Wilding, 2004), but effects were mixed in PwPD (Knowles et al., 2021; McRae et al., 2002; Tjaden & Wilding, 2004). These relations complicate the task of assessing the relationships of speech timing and articulatory precision with intelligibility and naturalness for the purpose of improving therapy. The next sections highlight what is currently known about these relationships before turning to the present study, which seeks to disentangle these effects.

Speech Timing and Intelligibility in PD

The relationship between atypical timing characteristics of hypokinetic dysarthria and intelligibility is complex. Perceptually, high correlations between ratings of deviation in various timing characteristics (e.g., short rushes, variable rate, and fast rate) and reduced intelligibility have been found in PwPD (Darley et al., 1969). However, a later investigation did not corroborate these results (Plowman-Prine et al., 2009). Acoustically, one study found that indicators of disfluency, including repetitions, omissions, and pauses, could explain 54% of the variance in overall intelligibility in PwPD, even after accounting for dysarthria severity (Liu et al., 2022). Furthermore, when speaking rate was intentionally reduced in PwPD, some studies found improved intelligibility (Blanchet & Snyder, 2010; Martens et al., 2015; McAuliffe et al., 2014), whereas others found no effect (Tjaden et al., 2014; Tjaden & Wilding, 2004; van Brenk et al., 2022; Van Nuffelen et al., 2010) or impaired intelligibility (Van Nuffelen et al., 2009). These discrepancies may be explained by potential differences in speakers' habitual speech characteristics (Fletcher et al., 2017). Thus, there is currently mixed evidence for a relationship between speech timing characteristics and intelligibility.

Speech Timing and Naturalness in PD

Because timing is an important component of prosody and prosody is an important component of naturalness, there is an expected link between speech timing characteristics and naturalness. Furthermore, components of speech timing, such as rate, rhythm, and stress, are often explicitly included in operational definitions of naturalness (Yorkston et al., 2010), making timing an intrinsic component of naturalness judgments. However, there is little to

no direct evidence of an association of these percepts. Darley et al. (1969) found high correlations ($r > .5$) in PD between the perception of inappropriate silences, short rushes of speech, and the variable rate and perception of bizarreness, which could be considered the opposite of naturalness. Acoustically, faster articulation rate and shorter syllable durations were associated with reduced perceived naturalness in a within-speaker correlation analysis in PD (Klopfenstein, 2015). Relevant to speech therapy, rate reduction may have a negligible or detrimental effect on naturalness (Tjaden et al., 2014; Yorkston et al., 1990), potentially due to changes in prosodic contrasts. These studies suggest that naturalness is intrinsically related to speech timing, but evidence is either scant or indirect.

Articulatory Precision and Intelligibility in PD

Articulatory imprecision is highly correlated with intelligibility in PD (Darley et al., 1969; Plowman-Prine et al., 2009). Acoustic correlates and direct measures of articulator movement also show significant correlations with intelligibility in PwPD (Kearney et al., 2017; Tjaden & Wilding, 2004; Weismer et al., 2012). Finally, intelligibility is increased in speaking conditions in which articulatory distinctiveness is cued (i.e., clear speech; Stipancic et al., 2022; Tjaden et al., 2014; van Brenk et al., 2022). Overall, there is strong evidence for a relationship between articulatory precision and intelligibility in PD.

Articulatory Precision and Naturalness in PD

Interestingly, the relationship between articulatory precision and naturalness has been relatively underexplored. A single study has collected listener ratings of both naturalness and articulatory precision in PwPD, with the goal of predicting cognitive function (Brown & Spencer, 2023); although they did not examine the relationship between these variables, calculations based on their published values indicate a Pearson's r of .84. Additional indirect evidence comes from the acoustic measure of voice onset time, a cue for distinguishing voiced and voiceless stop consonants that was found to be a significant predictor of speech naturalness in individuals who stutter posttherapy (Metz et al., 1990). PwPD show a reduced contrast in this measure between voiced and voiceless consonants (Hammer, 2013; Knowles et al., 2021). Thus, articulatory precision is potentially related to naturalness in PwPD, but this relationship has not yet been explicitly examined.

Summary and Motivation for the Present Study

Together, the aforementioned studies provide substantial evidence of a relationship between speech timing and intelligibility and between articulatory precision and intelligibility, with fewer demonstrated relationships between these features and naturalness. However, given the potential relationship between articulatory precision and speech timing, it is difficult to determine whether correlations between these features and intelligibility and naturalness are independent, whether the relationship of one feature is driving correlations from another feature, or whether these findings are due to some underlying feature that explains variability in both speech features. Understanding these separate contributions to intelligibility and naturalness and their relative importance has clinical implications for selecting the most effective therapy for an individual given a limited amount of therapy time. For example, if improving speech naturalness is a therapy goal and articulatory precision does not affect naturalness beyond what is explained by speech timing, focusing only on modifying speech timing (and not articulatory precision) would be the most efficient way to achieve the goal. The overwhelming majority of studies examining these relationships in PwPD and both studies focusing on auditory-perceptual ratings (Darley et al., 1969; Plowman-Prine et al., 2009) evaluated each feature's effect separately using bivariate correlations (e.g., articulatory precision vs. intelligibility), which precludes understanding their separable contributions. A single study has used multiple regression to understand the dissociable effects of auditory-perceptual features on intelligibility in dysarthria (De Bodt et al., 2002). However, this analysis was carried out on speakers with a variety of dysarthria subtypes, and individual speaker characteristics were not included, making it difficult to translate specifically to PD.

In the present study, we sought to expand upon previous research by assessing the extent to which inexperienced listeners' perceptions of speech timing and articulatory precision contribute to their perceptions of intelligibility and naturalness. We additionally assessed whether these relationships were significant after controlling for the effects of the other variable. To do so, we used a global measure of a listener's perception of speech rate and rhythm, termed *timing typicality*. This was motivated by a desire to quantify the overall severity of speech timing deviations in PD separate from other prosodic components while providing efficiency over the 10 auditory-perceptual features related to speech timing used in previous work (Dimensions 22–31 in the Appendix section of Darley et al., 1969). We hypothesized that, based on prior work associating speech timing percepts and imprecise consonants with both intelligibility and bizarreness (Darley et al., 1969; Plowman-Prine et al., 2009), both timing typicality and articulatory precision on their own would be significantly related to both intelligibility and naturalness as rated by inexperienced listeners. It is unclear, however, whether

these correlations will remain significant when the other variable is controlled for. On the basis of the high bivariate correlation values in Darley et al.'s (1969) study, we hypothesized that both perceived speech timing typicality and articulatory precision would be significantly associated with both intelligibility and naturalness, even after taking the effects of the other into account. Thus, both timing and articulation percepts would contribute unique information to explain variation in both intelligibility and naturalness.

Method

In this study, a cohort of young adults with typical speech, language, and hearing listened to speech samples from people with and without PD and made auditory-perceptual judgments about four aspects of their speech: intelligibility, naturalness, timing typicality, and articulatory precision. All participants (speakers and listeners) provided informed consent through the Boston University Institutional Review Board or the University of Washington Institutional Review Board, and all participants received compensation for their time.

Speakers

Speech recordings were selected from our laboratory's database of 165 individuals with idiopathic PD and 127 older adults with no speech, language, or neurological disorders. The first author selected a subset of 92 individuals with PD exhibiting a wide variety of timing and articulatory atypicalities (including some with very typical speech) using his auditory-perceptual judgment. A certified speech-language pathologist with 8 years of experience listening to speech in PwPD listened to speech samples from this subset and provided preliminary scores of intelligibility and naturalness for each using separate visual analog scales (VASs; see a description of the stimuli, VAS task, and instructions in subsequent sections). To maximize the ability to detect correlations of interest specific to intelligibility or naturalness in the final analysis, a preliminary procedure was carried out to reduce the (potential) correlation between intelligibility and naturalness. As in the work of Anand and Stepp (2015), speakers were grouped into four quadrants using these preliminary scores: (a) high intelligibility and high naturalness, (b) high intelligibility and low naturalness, (c) low intelligibility and high naturalness, and (d) low intelligibility and low naturalness. Preliminary data were then plotted, and select speakers were removed from each quadrant based on their preliminary scores to leave an approximately equal distribution across quadrants, a uniform continuum of speech naturalness and intelligibility, and a balance of male and female speakers. A preliminary Pearson correlation analysis was computed to confirm that naturalness and intelligibility were not significantly correlated ($r = .12$, $p = .57$).

This speaker selection process yielded a sample of 20 PwPD (10 men, 10 women[1]) ranging in age from 46 to 81 years ($M = 69.7$, $SD = 7.7$) and with times since diagnosis from 1 to 20 years ($M = 9.0$, $SD = 5.8$). Recordings from all speakers were collected during the "ON" phase of their medication. To provide a rough reference for how PwPD were rated compared to speakers with no neurological disorder, four control speakers who reported typical speech, language, and hearing were selected, approximately matching the sex (two men, two women; Van Nuffelen et al., 2009) and age (range: 64–77 years, $M = 70.7$, $SD = 5.4$) of the group of PwPD. The same certified speech-language pathologist who provided preliminary scores also rated the overall severity of dysarthria on a scale from 0 (*no signs of dysarthria*) to 100 (*very severe dysarthria*). Dysarthria severity scores ranged from 10 to 65 ($M = 27.7$, $SD = 14.0$). See Table 1 for complete demographic information.

Speech Stimuli

Recordings consisted of each speaker reading the first six sentences of the Rainbow Passage (Fairbanks, 1960). For each construct in the auditory-perceptual listening experiment, listeners were presented with 24 recording samples (one from each speaker) and five (20%) repeated samples in order to calculate intrarater reliability. As timing and articulation are, to a large degree, dependent on the phonetic and contextual content of a speech sample, the same stimuli were used for all participants to maintain valid comparisons. Any stuttering behaviors were intentionally left in recordings to reflect the natural components of individuals' speech timing behaviors. All stimuli were amplitude-normalized such that the root-mean-square of the amplitude signal was the same across recordings. Multitalker babble was added to the samples during judgments of intelligibility at a signal-to-noise ratio of 0 dB to increase ecological validity and to avoid ceiling effects based on pilot testing (Tjaden et al., 2014). This multitalker babble comprised recordings of four male and four female older adults with typical speech not included in the data set (as in Abur et al., 2021). Multitalker babble was not added to the samples for ratings of the other constructs because it could have made it more difficult to perceive the important aspects of the actual speech signal, leading to less accurate estimates of those features. Although this may have reduced the ability to detect significant correlations between timing typicality and intelligibility and between articulatory precision and intelligibility, it

[1]Gender information was not available for this sample.

Table 1. Speaker demographics.

Speaker ID	Sex	Age	Years since diagnosis	Dysarthria severity[a]
PD01	F	59	18	25
PD02	F	70	7.5	60
PD03	F	68	17	25
PD04	F	66	11	25
PD05	F	76	10	15
PD06	F	77	16	35
PD07	F	68	3	20
PD08	F	77	12	25
PD09	F	68	6	30
PD10	F	68	5	10
PD11	M	73	20	25
PD12	M	69	2	25
PD13	M	46	10	30
PD14	M	71	9	65
PD15	M	66	15	10
PD16	M	81	2	40
PD17	M	75	1.5	20
PD18	M	79	6.5	18
PD19	M	71	2	25
PD20	M	66	7	25
C01	F	64		5
C02	F	77		5
C03	M	70		0
C04	M	72		0

Note. PD = speakers with Parkinson's disease; F = female; M = male; C = neurologically healthy control speakers.

[a]Higher severity ratings are more severe.

was a necessary trade-off that may need to be considered when interpreting the results.

Listeners

Recordings were presented to 20 inexperienced listeners (seven cisgender men, 10 cisgender women, two nonbinary people, and one agender person; nine men, 11 women) aged between 19 and 28 years (M = 21.8, SD = 2.4). All listeners' primary language was North American English, and they reported no history of speech, language, hearing, or neurological disorders. All participants passed a hearing screening using a 25–dB HL cutoff at all tested frequencies (125, 250, 500, 1000, 2000, 4000, and 8000 Hz; American Speech-Language-Hearing Association, 2005). To ensure that listeners' perceptions were not enhanced by previous exposure to speech in individuals with PD (Borrie et al., 2012), they were excluded if they were majors or had majored in speech, language, and hearing science or had any prior experience with auditory-perceptual ratings of motor speech disorders.

Experimental Procedure

Listeners were seated in a sound-attenuating booth in front of a computer screen displaying a custom MATLAB graphical user interface. They were then asked to make judgments about speech recordings for each of the four constructs: intelligibility, naturalness, timing typicality, and articulatory precision. Speech samples were presented through over-the-ear headphones (Sennheiser HD 280 Pro headphones) in four experimental runs (one for each construct) of 29 trials (one sample from each speaker plus five repeated samples). During the session, participants were first familiarized with descriptions of intelligibility and naturalness. *Intelligibility* was defined as "the degree to which a speaker's message can be recovered by a listener" (Kent et al., 1989; p. 483). *Naturalness* was defined as how much speech "conforms to the listener's standards of rate, rhythm, intonation, stress patterning, and if it conforms to the syntactic structure of the utterance being produced" (Yorkston et al., 2010; p. 288). Listeners then completed a practice run comprising four speech samples of control speakers not included in the experiment in order to get used to the task and provide feedback on loudness so the experimenter could adjust the intensity in their headphones to a comfortable level. These samples underwent the same intensity normalization and mixture of multitalker babble as the experimental intelligibility samples. Following this, listeners completed the intelligibility and naturalness runs (first rating all speakers on intelligibility and then rating all speakers on naturalness). The intelligibility run was always completed before the naturalness run to reduce familiarity effects, especially since the stimuli for rating intelligibility were presented with multitalker babble, whereas the other ratings were completed on samples with nothing added. Participants were then familiarized with definitions of timing typicality and articulatory precision. *Timing typicality* was operationally defined as "the extent to which a speaker's speech cadence is as expected in a typical speaker. Atypical timing can be halting, rushed, or too evenly paced for the context. This does not take into account factors like pitch or loudness." *Articulatory precision* was defined as "the clarity of a speaker's speech sounds. Imprecise articulation shows slurring, inadequate sharpness, distortions, lack of crispness and clumsiness in going from one sound to another" (adapted from Plowman-Prine et al., 2009, and Darley et al., 1969). The order of the timing typicality and articulatory precision runs was counterbalanced across participants to minimize potential order effects.

Within each run, samples were presented in a random order that was different for each listener. Furthermore, the five repeated samples were randomly selected for each listener and were always presented (in a random

order) at the end of the run. On a given trial, listeners were presented with a horizontal sliding bar and were instructed to rate the sample on a VAS from 0 to 100 according to the construct definition. The end points of the sliding bar were labeled "0" and "100," and the instructions directly above the bar stated, "Please rate the speech sample from 0 (Completely A) to 100 (Completely B)," where A and B represent both ends of the continuum (i.e., completely unintelligible and completely intelligible, respectively). At the beginning of each trial, the slider was positioned in the center of the sliding bar (equivalent to 50). The definition of the construct they were rating remained on screen for the duration of the experimental run. The speech sample could be played up to two times before submitting their rating (as in Abur et al., 2021), but listeners had to listen to the entire sample at least once to ensure comparable stimuli were rated by all listeners. Breaks could be taken at any time.

Despite the use of speech samples from our laboratory's database, none of the outcome data from this article have been reported previously. All auditory-perceptual tasks were completed explicitly for this study.

Statistical Analysis

All statistical analyses were carried out using either RStudio (Version 2013.06.1) or custom MATLAB scripts. Intra- and interrater reliability were assessed using intraclass correlation coefficients (ICCs) with ratings from both PwPD and control speakers. Intrarater reliability was determined using the formula ICC(A, 1) because the repeated ratings did not systematically differ and single-rater values were of most interest. Interrater reliability was determined using the formula ICC(C, *k*) because the repeated ratings systematically differed by listener and the reliability of the group average was of most importance for the present study (Koo & Li, 2016; McGraw & Wong, 1996). Note that because ICC(C, *k*) evaluates reliability based on the average of several ratings, it is inherently more reliable than the intrarater reliability metric. This measure is consistent with prior auditory-perceptual literature of intelligibility (e.g., Knowles et al., 2021; Tjaden et al., 2014; van Brenk et al., 2022). ICC values were classified as poor (less than .5), moderate (.5–.75), good (.75–.9), and excellent (greater than .9) reliability according to the guidelines of Koo and Li (2016). Because the spread of values across repeated trials was large for some participants and small for others, we also calculated absolute mean difference scores between the first and second repeated samples as an additional measure of intrarater reliability.

Linear regression models were used to determine the association between independent (timing typicality and articulatory precision) and dependent (intelligibility and naturalness) variables of interest. All models only included ratings of PwPD so that the results specifically pertained to this population. We first assessed whether our dependent variables of interest (intelligibility and naturalness) were correlated using Pearson product–moment correlations to determine the degree to which results may have been affected by dependent variables that were related to one another. We then constructed four regression models in which the bivariate relationships between articulatory precision and timing typicality as independent variables and intelligibility and naturalness as dependent variables were assessed, including the associated coefficients of determination (R^2). Two multiple regression models were then constructed to assess the unique associations between the independent and dependent variables and the amount of variance (R^2) explained by each. The first included articulatory precision and timing typicality as predictor variables, with intelligibility as the outcome variable. The second had the same predictor variables, but the outcome variable was naturalness. Pearson product–moment correlations between our independent variables as well as variance inflation factors (VIFs) were calculated to determine whether multicollinearity would impact separable estimates of each variable. All statistical tests were carried out with a significance level of $\alpha = .05$.

The approximate distributions of shared and unique variance among the variables of each model were visualized with Euler plots (Ip, 2001) using the *eulerr* package in R. The area of each segment is proportional to the sum of squares unique to or shared by the variables as derived using the *aov* function in R. As a result, the ratio between the overlapping and nonoverlapping parts of each variable is approximately equal to the calculated R^2.

Results

Reliability

Intrarater reliability as measured by ICC(A, 1) was moderate on average and highly variable across listeners for intelligibility ($M_{ICC} = 0.59$, $SD = 0.33$), naturalness ($M_{ICC} = 0.61$, $SD = 0.39$), timing typicality ($M_{ICC} = 0.73$, $SD = 0.22$), and articulatory precision ($M_{ICC} = 0.71$, $SD = 0.25$). Because repeated ratings had a wider spread of values for some listeners than others, which could negatively bias ICC scores, mean absolute difference scores of VAS ratings (out of 100) were also calculated for intelligibility ($M = 14.8$, $SD = 7.4$), naturalness ($M = 16.2$, $SD = 6.6$), timing typicality ($M = 12.3$, $SD = 5.9$), and articulatory precision ($M = 13.9$, $SD = 6.8$). Interrater reliability as measured by ICC(C, *k*) was excellent for intelligibility (.98), naturalness (.96), timing typicality (.95), and articulatory precision (.97).

Association Between Outcome Variables

The relationship between intelligibility and naturalness was not significant in PwPD (r = .28, p = .24), as expected given the way the sample was selected (see Figure 1). All four control speakers were found to have higher naturalness scores than any of the PwPD and higher intelligibility scores than 70% of the PwPD.

Association Between Independent Variables

There was high correlation between timing typicality and articulatory precision (r = .75, p < .001, VIF = 2.31 between the two variables in the following models; see Figure 2).

Predictors of Intelligibility

Linear regressions with each independent variable evaluated separately found that timing typicality was not significantly associated with intelligibility, β = .49, $t(18)$ = 1.69, p = .11, R^2 = .14, but articulatory precision was, β = .68, $t(18)$ = 2.44, p = .025, R^2 = .25 (see Figure 3). Using a linear regression model with both timing typicality and articulatory precision as the independent variables and intelligibility as the dependent variable, neither timing typicality, β = −.02, $t(17)$ = −0.04, p = .97, nor articulatory precision, β = .70, $t(17)$ = 1.59, p = .13, was significant after controlling for the other, and the model did not explain a significant amount of variance in intelligibility, R^2 = .25, $F(2, 17)$ = 2.82, p = .09.

Predictors of Naturalness

Linear regressions with each independent variable evaluated separately found that both timing typicality, β = .87, $t(18)$ = 12.39, p < .001, R^2 = .90, and articulatory precision, β = .7, $t(18)$ = 5.91, p < .001, R^2 = .66, were significantly associated with naturalness (see Figure 4). Using a linear regression model with both timing typicality and articulatory precision as the independent variables and naturalness as the dependent variable, both timing typicality, β = .71, $t(17)$ = 7.32, p < .001, and articulatory precision, β = .22, $t(17)$ = 2.18, p = .04, were found to be significant after controlling for the other. The entire model explained almost all the variance in naturalness, R^2 = .92, $F(2, 17)$ = 95.2, p < .001.

Figure 1. Mean perceived naturalness plotted as a function of mean perceived intelligibility. Circles represent average ratings across 20 listeners. The dashed line indicates the least squares fit line for people with Parkinson's disease only. Control = neurologically healthy control speakers; PD = speakers with Parkinson's disease.

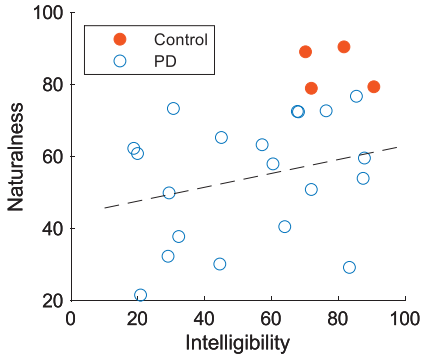
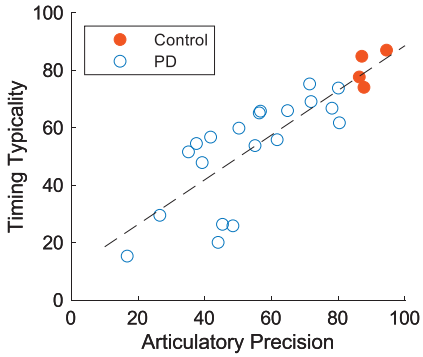


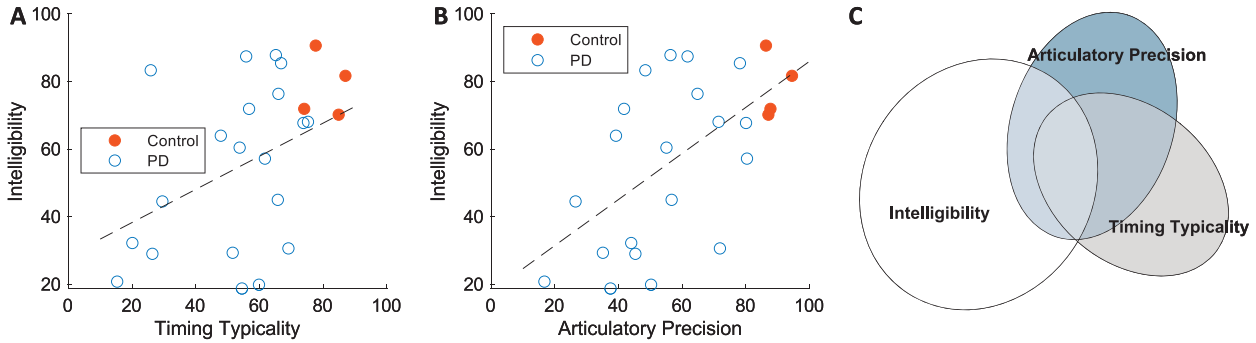
Figure 2. Mean perceived timing typicality plotted as a function of mean perceived articulatory precision. Circles represent average ratings across 20 listeners. The dashed line indicates the least squares fit line for people with Parkinson's disease only. Control = neurologically healthy control speakers; PD = speakers with Parkinson's disease.



Discussion

The aim of this study was to examine the contributions of timing typicality and articulatory precision to intelligibility and naturalness in PwPD. It was hypothesized that, on their own, both timing typicality and articulatory precision would be significantly correlated with both intelligibility and naturalness. The present findings partially upheld this hypothesis: Significant regression coefficients were found when timing typicality and articulatory precision were regressed separately on naturalness and when articulatory precision was regressed on intelligibility. However, when timing typicality was regressed on intelligibility, the coefficient was not significant. Furthermore, it was hypothesized that after accounting for the effect of articulatory precision, timing typicality would be significantly related to both intelligibility and naturalness. The present findings only partially supported this hypothesis, with a nonsignificant coefficient in the multiple regression model of intelligibility and a significant coefficient in the multiple regression model of naturalness. Similarly, it was hypothesized that after accounting for the effect of

Figure 3. (A and B) Mean perceived intelligibility plotted as a function of (A) perceived timing typicality and (B) perceived articulatory precision. Circles represent average ratings across 20 listeners. The dashed line indicates the least squares fit line for people with Parkinson's disease only. (C) An approximate visual representation of shared variance between the independent and dependent variables generated using the *eulerr* package in R. Sum of squares was used to scale the size of the ellipses representing each variable. Control = neurologically healthy control speakers; PD = speakers with Parkinson's disease.

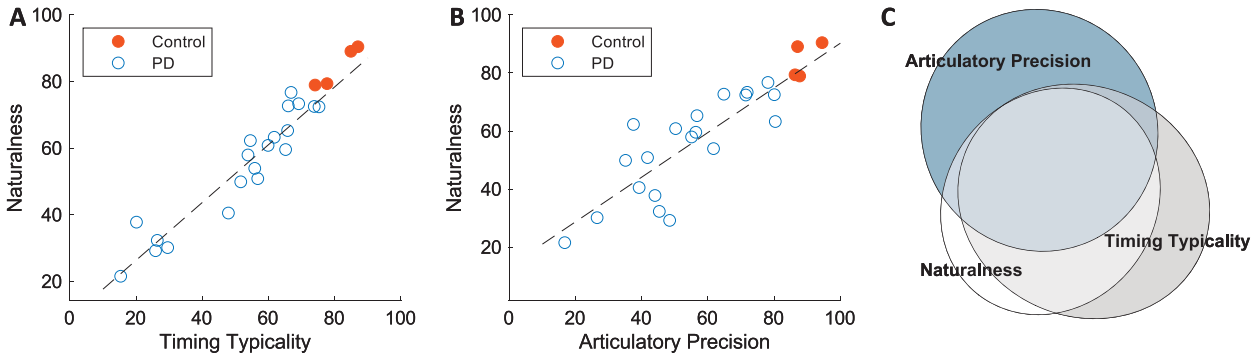


timing typicality, articulatory precision would be significantly related to both intelligibility and naturalness. This hypothesis was, again, only partially upheld by the present study: Articulatory precision was not uniquely associated with intelligibility, but it was uniquely associated with naturalness. These results and their implications are discussed further below.

Predictors of Intelligibility

Consistent with prior literature (Darley et al., 1969; Plowman-Prine et al., 2009), articulatory precision was a significant predictor of intelligibility when examined by itself, although perhaps to a lesser extent (R^2 = .25/Pearson's r = .5 in the present study compared to R^2 = .83/r = .91 in Darley et al., 1969, and Spearman's r = .81 in Plowman-Prine et al., 2009). In contrast, a global measure of timing typicality was not found to be significantly correlated with intelligibility on its own, contrasting with individual measures of distinct aspects of timing found by Darley et al. (1969) but similar to that of Plowman-Prine et al. (2009), in which no significant relationships between speech timing variables and intelligibility were found. Together, these results match the broader literature: Acoustic measures and behavioral modifications related to articulatory precision are consistently correlated with perceived intelligibility (Kearney et al., 2017; Stipancic et al., 2022; Tjaden et al., 2014; Tjaden & Wilding, 2004; van Brenk et al., 2022; Weismer et al., 2012). In contrast, modifying speech timing has inconsistent effects on intelligibility (e.g., Blanchet & Snyder, 2010; Tjaden et al., 2014; Van Nuffelen et al., 2009). One reason for the differences between the present study and prior perceptual studies could be that the prior work was rated by trained speech-language pathologists including, in the case of Darley et al., the authors themselves. These raters may have their own

Figure 4. (A and B) Mean perceived naturalness plotted as a function of (A) perceived timing typicality and (B) perceived articulatory precision. Circles represent average ratings across 20 listeners. The dashed line indicates the least squares fit line for people with Parkinson's disease only. (C) An approximate visual representation of shared variance between the independent and dependent variables generated using the *eulerr* package in R. Sum of squares was used to scale the size of the ellipses representing each variable. Control = neurologically healthy control speakers; PD = speakers with Parkinson's disease.



biases based on prior exposure, compared to the present sample of inexperienced listeners. At the same time, evidence suggests that experienced and inexperienced raters do not behave differently in rating dysarthric speech (e.g., Bunton et al., 2007), and several other variables such as the content of the speech sample (van Brenk et al., 2022) impact ratings of intelligibility. There may have been differences in overall dysarthria severity and variability across these studies; a wider range of intelligibility scores may have allowed for the detection of significant relationships with this variable. Furthermore, unlike previous studies, our intelligibility ratings were completed in the presence of multitalker babble. As a result, the features leading to variability in intelligibility scores may have been different across studies. Finally, the timing typicality construct may have incorporated aspects of timing that were less correlated with intelligibility in prior studies, thus yielding a weaker relationship in the present study.

Despite a significant bivariate relationship between articulatory precision and intelligibility, when both timing typicality and articulatory precision were included as predictors of intelligibility in a single model, neither was significant. This indicates that neither variable contributes unique information about intelligibility beyond what is explained by the other variable, which may be due to the high correlation between the predictors (see Figure 3C for an illustration of the shared variance among each of the predictors). One potential clinical implication is that targeting both articulatory precision and timing in speech therapy may not provide more benefit to intelligibility than addressing articulatory precision alone. Of course, speech therapy strategies such as using clear, overenunciated speech will inevitably alter both articulation and speech timing, potentially to the detriment of naturalness. In this context, the present results suggest that these speech timing changes are less important if the primary therapeutic goal is increasing intelligibility. However, as these findings reflect across-speaker correlations, future analyses evaluating these relationships within speakers and a specific investigation of the effects of therapy would be needed to confirm this suggestion.

Predictors of Naturalness

The present study confirmed the implicit relationship between timing typicality and naturalness, with ratings of timing typicality accounting for over 90% of the variance in naturalness rating. This somewhat corroborates the significant correlations found previously between features of timing such as reduced stress, variable rate, short rushes, and inappropriate silences as well as an overall judgment of bizarreness (Darley et al., 1969), which might be comparable to the opposite of naturalness. Furthermore, the strength of the correlation in the present study may have exceeded that found by Darley et al. (1969) because timing typicality likely reflects a combination of these variables. In addition, the present study explicitly examined the relationship between articulatory precision and naturalness for the first time and found a significant correlation (with articulatory precision accounting for about 66% of the variance in naturalness). The strength of this relationship is somewhat comparable to the unreported findings in Brown and Spencer (2023) and the relationship between consonant precision and bizarreness found by Darley et al., indicating a stable correlation between these percepts, even across speech-language pathologists (Brown & Spencer, 2023; Darley et al., 1969) and inexperienced listeners (the present study).

Taken on its own, one might presume that this relationship between articulatory precision and naturalness is only so strong because of the high correlation between articulatory precision and timing typicality. However, even after accounting for the effects of timing typicality, articulatory precision is still a significant predictor of naturalness. This suggests that, in contrast with intelligibility, targeting both timing and articulation in speech therapy would be important for maximizing speech naturalness. Nonetheless, articulatory precision only explains an additional 2% of the variance in naturalness above and beyond that of timing typicality (R^2 = .90 for the model that only included timing typicality vs. R^2 = .92 for the model that included both predictors), so although the contribution is statistically significant, it may not be clinically meaningful. Similar to intelligibility, additional research into the relationship between timing typicality, articulatory precision, and naturalness across speaking tasks within speakers, particularly in response to therapy, will be needed to determine the precise clinical impact of modifying timing, articulation, or both on perceived naturalness.

Potential for Timing Typicality as an Outcome Measure

The present study introduced a new auditory-perceptual construct, namely, timing typicality, to characterize the overall perception of disruption to speech timing in PwPD and its relationship with intelligibility and naturalness. Reduced loudness and pitch variability are aspects of impaired prosody that are consistently referenced in the literature (e.g., Bowen et al., 2013; Darley et al., 1969; Logemann et al., 1978). In contrast, the precise characterization of speech timing deficits in PwPD is not as consistently reported. This may be due to the multifactorial nature of speech timing deviations in PwPD comprising speech rate, rhythm, pausing, and disfluency. Indeed, there were 10 auditory-perceptual characteristics related to

timing evaluated by Darley et al. (1969). Having so many factors may have precluded a clear understanding of the nature of speech timing deficits in PD.

Because of its relative efficiency, timing typicality could be a useful auditory-perceptual attribute that can be quickly assessed and used to track progression and therapy-related improvement. Currently, in both research and clinical settings, judging rate and/or prosody would be most similar. However, rate on its own does not capture the variety of temporal atypicalities in speech in PwPD, and prosody encompasses other variables, such as the modulation of pitch or loudness, that are related to, but potentially separate from, the control of speech timing. In order for this measure to be useful, its validity and reliability need to be evaluated. The present study was a first step toward this, evaluating intra- and interrater reliability and the relationships between timing typicality and other clinically relevant metrics such as intelligibility and naturalness. Convergent and discriminant validity were demonstrated to some degree since timing typicality would be expected to have a closer relationship with naturalness than intelligibility given their operational definitions. In terms of reliability, this construct had comparable, if not slightly better, intrarater reliability than the other measures and similar interrater reliability. Together, these factors suggest promise for using timing typicality for research or clinical purposes, but further examination of both validity and reliability will be necessary.

Limitations and Future Considerations

Although this study provided the first evidence of the interdependencies of speech timing and articulation on intelligibility and naturalness in PwPD, there are certain limitations that should be acknowledged. First, certain trade-offs were made when choosing the speech sample material. Listeners heard the same passage repeated for all speakers and constructs, which could have led to a positive bias in ratings for intelligibility (i.e., all speakers rated as more intelligible) than if speech samples were varied as in other studies (e.g., Abur et al., 2019; Stipancic et al., 2022). In addition, knowing the content may have led listeners to subconsciously fill in articulatory deviations such that they were less able to perceive them and rated articulatory precision more highly. Alternatively, it could have led them to perceive minor deviations as more salient, resulting in a negative bias. Relatedly, the first several samples may have been rated as less intelligible than later samples as the listeners became familiarized with the passage. As previously noted, the same passage was presented repeatedly to standardize the phonetic and prosodic contexts, which was imperative for ratings of timing and articulation. In addition, asking listeners to rate the six-sentence passage as a whole rather than in smaller chunks could have biased listeners' attention to "memorable" anomalies. This could have led speakers to reduce their scores for the whole passage more than if the sentences were rated individually. Future research will be needed to evaluate whether using shorter, less predictable samples (e.g., Sentence Intelligibility Test sentences; Yorkston et al., 1996) has a meaningful effect on these ratings.

In addition, the present study found only moderate intrarater reliability across all four auditory-perceptual constructs using standard ICC metrics. For intelligibility, intrarater reliability as measured using ICC(A, 1) was reduced (.59) compared to that (.71) in the only prior study using this metric (van Brenk et al., 2022). However, we suspected that the ICC values were an underestimate due to the low number of repeated samples. Using mean absolute difference scores, intrarater reliability for intelligibility was similar to one previous study that used inexperienced listeners and a VAS (Abur et al., 2019). Comparatively fewer studies have examined intrarater reliability for ratings of naturalness and articulatory precision in PwPD, and none have evaluated inexperienced raters. In the absence of comparable ratings, comparing mean absolute difference scores between intelligibility and the other constructs is one way to assess reliability for previously unevaluated constructs. Using this approach, intrarater reliability similar to intelligibility was found across constructs (intelligibility: 14.8, naturalness: 16.2, timing typicality: 12.3, and articulatory precision: 13.9), with potentially better scores for timing typicality and articulatory precision. Furthermore, for interrater reliability, excellent reliability (> .95) was found for all constructs using a standard ICC metric. This was similar to or better than that reported by prior work in which inexperienced raters used a VAS for intelligibility (.89 in Knowles et al., 2021; .83 in Tjaden et al., 2014; and .85 in van Brenk et al., 2022), naturalness (.93 in Brown & Spencer, 2023), and articulatory precision (.98 in Chiu et al., 2021).

As discussed above, making strong claims about the present results as they pertain to clinical decisions is preliminary at present. As with any correlational analysis across individuals, there may have been underlying differences between speakers that led to changes in multiple variables and explain their relationships to one another. In the context of PD, one underlying variable that may explain some of the results is differences in severity, whether defined as disease stage, extent of global motor symptoms, or overall dysarthria severity. If overall severity was the main driver of variation in intelligibility, naturalness, timing typicality, and articulatory precision, their significant associations would not be particularly unexpected. This issue contributed to the decision to ensure that naturalness and intelligibility were not correlated in

the sample—if both were related to overall severity, generating such a sample would reduce the potential effects of this underlying variable. This issue also highlights a benefit of using multiple regression to examine relationships between speech features and outcomes such as intelligibility and naturalness: Including multiple speech features in a single model can help account for these underlying variables. For example, the association between articulatory precision and naturalness is evaluated after taking into account the effects of timing typicality on naturalness, including any component of timing typicality related to severity. Clearly, however, if overall severity impacts some variables more than others, this benefit would be reduced.

A potential area of future research would be assessing the relationships among these variables during conversational speech rather than read speech. There are several auditory-perceptual and acoustic differences that have been found between speaking contexts, including understandability and naturalness (Weir-Mayta et al., 2017), timing (Lowit et al., 2018; Maffia et al., 2021), and articulation (Rusz et al., 2013). As a result, the relationships among these perceptual dimensions during conversational speech may be different from those found in the present study. Furthermore, they may provide important insights for addressing everyday communication challenges in PwPD during speech therapy. At the same time, using more natural speech samples allows for less control over the content, which could introduce unnecessary variability. Assessing these relationships across several contexts would provide the most complete and informative understanding for clinical translation.

Finally, timing and articulatory precision are not the only factors that impact intelligibility and naturalness (Anand & Stepp, 2015; Darley et al., 1969; Plowman-Prine et al., 2009). As a result, there may be more complex associations that are not captured by these factors. Carrying out a regression with other predictor variables such as monopitch and monoloudness would allow for a more complete examination of the unique effects of each of these variables on intelligibility and naturalness.

Conclusions

In this study, we sought to examine the contributions of the percepts of speech timing and articulatory precision on naturalness and intelligibility in individuals with PD. Both timing typicality and articulatory precision contributed unique information to explain variation in naturalness during a reading passage, but not intelligibility. These results provide an initial examination of the complex interactions that various perceptual features may have related to their contributions to intelligibility and naturalness. By better understanding the unique contributions of speech timing and articulation to intelligibility and naturalness, this study may aid in the developing and evaluating novel therapies that address the diverse impacts of PD on speech production.

Data Availability Statement

The data sets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grant R01 DC020867 (C.E.S. and D.D.M.). The authors wish to thank Daniel Buckley for his assistance in providing expert auditory-perceptual ratings; Defne Abur for the auditory-perceptual presentation scripts and multitalker babble; and Julia Toto, Kaitlyn Siedman, and Megan Cushman for their assistance with recruitment.

References

Abur, D., Enos, N. M., & Stepp, C. E. (2019). Visual analog scale ratings and orthographic transcription measures of sentence intelligibility in Parkinson's disease with variable listener exposure. *American Journal of Speech-Language Pathology, 28*(3), 1222–1232. https://doi.org/10.1044/2019_AJSLP-18-0275

Abur, D., Subaciute, A., Daliri, A., Lester-Smith, R. A., Lupiani, A. A., Cilento, D., Enos, N. M., Weerathunge, H. R., Tardif, M. C., & Stepp, C. E. (2021). Feedback and feedforward auditory-motor processes for voice and articulation in Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 64*(12), 4682–4694. https://doi.org/10.1044/2021_JSLHR-21-00153

Ackermann, H., & Ziegler, W. (1991). Articulatory deficits in parkinsonian dysarthria: An acoustic analysis. *Journal of Neurology, Neurosurgery & Psychiatry, 54*(12), 1093–1098. https://doi.org/10.1136/jnnp.54.12.1093

American Speech-Language-Hearing Association. (2005). *Guidelines for manual pure-tone threshold audiometry* [Guidelines]. https://doi.org/10.1044/policy.GL2005-00014

Anand, S., & Stepp, C. E. (2015). Listener perception of monopitch, naturalness, and intelligibility for speakers with Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 58*(4), 1134–1144. https://doi.org/10.1044/2015_JSLHR-S-14-0243

Blanchet, P. G., & Snyder, G. J. (2010). Speech rate treatments for individuals with dysarthria: A tutorial. *Perceptual and Motor Skills, 110*(3), 965–982. https://doi.org/10.2466/PMS.110.3.965-982

Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual learning of dysarthric speech: A review of experimental studies. *Journal of Speech, Language, and Hearing Research, 55*(1), 290–305. https://doi.org/10.1044/1092-4388(2011/10-0349)

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.

Brown, K. A., & Spencer, K. A. (2023). The relationship between listener ratings of speech, primary motor symptoms, and global cognitive functioning in idiopathic Parkinson's disease. *Archives in Neurology & Neuroscience, 14*(2), 1–9.

Bunton, K., Kent, R. D., Duffy, J. R., Rosenbek, J. C., & Kent, J. F. (2007). Listener agreement for auditory-perceptual ratings of dysarthria. *Journal of Speech, Language, and Hearing Research, 50*(6), 1481–1495. https://doi.org/10.1044/1092-4388(2007/102)

Chiu, Y.-F., Neel, A., & Loux, T. (2021). Exploring the acoustic perceptual relationship of speech in Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 64*(5), 1560–1570. https://doi.org/10.1044/2021_JSLHR-20-00610

Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience, 30*(1–2), 1–31. https://doi.org/10.1080/23273798.2014.963130

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12*(2), 246–269. https://doi.org/10.1044/jshr.1202.246

De Bodt, M. S., Hernández-Díaz Huici, M. E., & Van De Heyning, P. H. (2002). Intelligibility as a linear combination of dimensions in dysarthric speech. *Journal of Communication Disorders, 35*(3), 283–292. https://doi.org/10.1016/s0021-9924(02)00065-5

Duffy, J. (2020). *Motor speech disorders: Substrates, differential diagnosis, and management* (4th ed.). Elsevier.

Fabbri, M., Guimarães, I., Cardoso, R., Coelho, M., Guedes, L. C., Rosa, M. M., Godinho, C., Abreu, D., Gonçalves, N., Antonini, A., & Ferreira, J. J. (2017). Speech and voice response to a levodopa challenge in late-stage Parkinson's disease. *Frontiers in Neurology, 8,* Article 432. https://doi.org/10.3389/fneur.2017.00432

Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed.). Harper & Row.

Fletcher, A. R., McAuliffe, M. J., Lansford, K. L., & Liss, J. M. (2015). The relationship between speech segment duration and vowel centralization in a group of older speakers. *The Journal of the Acoustical Society of America, 138*(4), 2132–2139. https://doi.org/10.1121/1.4930563

Fletcher, A. R., McAuliffe, M. J., Lansford, K. L., Sinex, D. G., & Liss, J. M. (2017). Predicting intelligibility gains in individuals with dysarthria from baseline speech features. *Journal of Speech, Language, and Hearing Research, 60*(11), 3043–3057. https://doi.org/10.1044/2016_JSLHR-S-16-0218

Hammer, M. J. (2013). Aerodynamic assessment of phonatory onset in Parkinson's disease: Evidence of decreased scaling of laryngeal and respiratory control. *Journal of Parkinson's Disease, 3*(2), 173–179. https://doi.org/10.3233/JPD-130180

Ho, A. K., Bradshaw, J. L., & Iansek, R. (2008). For better or worse: The effect of levodopa on speech in Parkinson's disease. *Movement Disorders, 23*(4), 574–580. https://doi.org/10.1002/mds.21899

Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1999). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology, 11*(3), 131–137. https://doi.org/10.1155/1999/327643

Ip, E. H. S. (2001). Visualizing multiple regression. *Journal of Statistics Education, 9*(1), Article 4. https://doi.org/10.1080/10691898.2001.11910646

Jiao, Y., Berisha, V., Liss, J., Hsu, S.-C., Levy, E., & McAuliffe, M. (2017). Articulation entropy: An unsupervised measure of articulatory precision. *IEEE Signal Processing Letters, 24*(4), 485–489. https://doi.org/10.1109/LSP.2016.2633871

Kearney, E., Giles, R., Haworth, B., Faloutsos, P., Baljko, M., & Yunusova, Y. (2017). Sentence-level movements in Parkinson's disease: Loud, clear, and slow speech. *Journal of Speech, Language, and Hearing Research, 60*(12), 3426–3440. https://doi.org/10.1044/2017_JSLHR-S-17-0075

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54*(4), 482–499. https://doi.org/10.1044/jshd.5404.482

Klopfenstein, M. (2015). Relationship between acoustic measures and speech naturalness ratings in Parkinson's disease: A within-speaker approach. *Clinical Linguistics & Phonetics, 29*(12), 938–954. https://doi.org/10.3109/02699206.2015.1081293

Klopfenstein, M., Bernard, K., & Heyman, C. (2020). The study of speech naturalness in communication disorders: A systematic review of the literature. *Clinical Linguistics & Phonetics, 34*(4), 327–338. https://doi.org/10.1080/02699206.2019.1652692

Knowles, T., Adams, S. G., & Jog, M. (2021). Variation in speech intelligibility ratings as a function of speech rate modification in Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 64*(6), 1773–1793. https://doi.org/10.1044/2021_JSLHR-20-00593

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Liu, C.-T., Chu, S.-W., & Chen, Y.-S. (2022). On the relationship between speech intelligibility and fluency indicators among English-speaking individuals with Parkinson's diseases. *Behavioural Neurology, 2022*(1), Article 1224680. https://doi.org/10.1155/2022/1224680

Logemann, J. A., & Fisher, H. B. (1981). Vocal tract control in Parkinson's disease. *Journal of Speech and Hearing Disorders, 46*(4), 348–352. https://doi.org/10.1044/jshd.4604.348

Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *Journal of Speech and Hearing Disorders, 43*(1), 47–57. https://doi.org/10.1044/jshd.4301.47

Lowit, A., Marchetti, A., Corson, S., & Kuschmann, A. (2018). Rhythmic performance in hypokinetic dysarthria: Relationship between reading, spontaneous speech and diadochokinetic tasks. *Journal of Communication Disorders, 72,* 26–39. https://doi.org/10.1016/j.jcomdis.2018.02.005

Maffia, M., De Micco, R., Pettorino, M., Siciliano, M., Tessitore, A., & De Meo, A. (2021). Speech rhythm variation in early-stage Parkinson's disease: A study on different speaking tasks. *Frontiers in Psychology, 12,* Article 668291. https://doi.org/10.3389/fpsyg.2021.668291

Marras, C., Beck, J. C., Bower, J. H., Roberts, E., Ritz, B., Ross, G. W., Abbott, R. D., Savica, R., Van Den Eeden, S. K., Willis, A. W., & Tanner, C. M. (2018). Prevalence of Parkinson's disease across North America. *npj Parkinson's Disease, 4*(1), Article 21. https://doi.org/10.1038/s41531-018-0058-0

Martens, H., Van Nuffelen, G., Dekens, T., Hernández-Díaz Huici, M., Kairuz Hernández-Díaz, H. A., De Letter, M., & De Bodt, M. (2015). The effect of intensive speech rate and intonation therapy on intelligibility in Parkinson's disease. *Journal of Communication Disorders, 58,* 91–105. https://doi.org/10.1016/j.jcomdis.2015.10.004

McAuliffe, M. J., Baylor, C. R., & Yorkston, K. M. (2017). Variables associated with communicative participation in Parkinson's disease and its relationship to measures of health-related quality-of-life. *International Journal of Speech-Language Pathology, 19*(4), 407–417. https://doi.org/10.1080/17549507.2016.1193900

McAuliffe, M. J., Kerr, S. E., Gibson, E. M. R., Anderson, T., & LaShell, P. J. (2014). Cognitive–perceptual examination of remediation approaches to hypokinetic dysarthria. *Journal of Speech, Language, and Hearing Research, 57*(4), 1268–1283. https://doi.org/10.1044/2014_JSLHR-S-12-0349

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

McRae, P. A., Tjaden, K., & Schoonings, B. (2002). Acoustic and perceptual consequences of articulatory rate change in Parkinson disease. *Journal of Speech, Language, and Hearing Research, 45*(1), 35–50. https://doi.org/10.1044/1092-4388(2002/003)

Metz, D. E., Schiavetti, N., & Sacco, P. R. (1990). Acoustic and psychophysical dimensions of the perceived speech naturalness of nonstutterers and posttreatment stutterers. *Journal of Speech and Hearing Disorders, 55*(3), 516–525. https://doi.org/10.1044/jshd.5503.516

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

Moya-Galé, G., & Levy, E. S. (2019). Parkinson's disease-associated dysarthria: Prevalence, impact and management strategies. *Research and Reviews in Parkinsonism, 9,* 9–16. https://doi.org/10.2147/JPRLS.S168090

Netsell, R., Daniel, B., & Celesia, G. G. (1975). Acceleration and weakness in parkinsonian dysarthria. *Journal of Speech and Hearing Disorders, 40*(2), 170–178. https://doi.org/10.1044/jshd.4002.170

Plowman-Prine, E. K., Okun, M. S., Sapienza, C. M., Shrivastav, R., Fernandez, H. H., Foote, K. D., Ellis, C., Rodriguez, A. D., Burkhead, L. M., & Rosenbek, J. C. (2009). Perceptual characteristics of Parkinsonian speech: A comparison of the pharmacological effects of levodopa across speech and non-speech motor systems. *NeuroRehabilitation, 24*(2), 131–144. https://doi.org/10.3233/NRE-2009-0462

Pu, T., Huang, M., Kong, X., Wang, M., Chen, X., Feng, X., Wei, C., Weng, X., & Xu, F. (2021). Lee Silverman voice treatment to improve speech in Parkinson's disease: A systemic review and meta-analysis. *Parkinson's Disease, 2021,* Article 3366870. https://doi.org/10.1155/2021/3366870

Rusz, J., Cmejla, R., Tykalova, T., Ruzickova, H., Klempir, J., Majerova, V., Picmausova, J., Roth, J., & Ruzicka, E. (2013). Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task. *The Journal of the Acoustical Society of America, 134*(3), 2171–2181. https://doi.org/10.1121/1.4816541

Skodda, S., Visser, W., & Schlegel, U. (2011). Vowel articulation in Parkinson's disease. *Journal of Voice, 25*(4), 467–472. https://doi.org/10.1016/j.jvoice.2010.01.009

Stipancic, K. L., van Brenk, F., Kain, A., Wilding, G., & Tjaden, K. (2022). Clear speech variants: An investigation of intelligibility and speaker effort in speakers with Parkinson's disease. *American Journal of Speech-Language Pathology, 31*(6), 2789–2805. https://doi.org/10.1044/2022_AJSLP-22-00189

Tjaden, K. (2000). A preliminary study of factors influencing perception of articulatory rate in Parkinson disease. *Journal of Speech, Language, and Hearing Research, 43*(4), 997–1010. https://doi.org/10.1044/jslhr.4304.997

Tjaden, K., Sussman, J. E., & Wilding, G. E. (2014). Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in Parkinson's disease and multiple sclerosis. *Journal of Speech, Language, and Hearing Research, 57*(3), 779–792. https://doi.org/10.1044/2014_JSLHR-S-12-0372

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

Tysnes, O.-B., & Storstein, A. (2017). Epidemiology of Parkinson's disease. *Journal of Neural Transmission, 124*(8), 901–905. https://doi.org/10.1007/s00702-017-1686-y

van Brenk, F., Stipancic, K., Kain, A., & Tjaden, K. (2022). Intelligibility across a reading passage: The effect of dysarthria and cued speaking styles. *American Journal of Speech-Language Pathology, 31*(1), 390–408. https://doi.org/10.1044/2021_AJSLP-21-00151

Van Nuffelen, G., De Bodt, M., Vanderwegen, J., Van de Heyning, P., & Wuyts, F. (2010). Effect of rate control on speech production and intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica, 62*(3), 110–119. https://doi.org/10.1159/000287209

Van Nuffelen, G., De Bodt, M., Wuyts, F., & Van de Heyning, P. (2009). The effect of rate control on speech rate and intelligibility of dysarthric speech. *Folia Phoniatrica et Logopaedica, 61*(2), 69–75. https://doi.org/10.1159/000208805

Walsh, B., & Smith, A. (2012). Basic parameters of articulatory movements and acoustics in individuals with Parkinson's disease. *Movement Disorders, 27*(7), 843–850. https://doi.org/10.1002/mds.24888

Weir-Mayta, P., Spencer, K. A., Eadie, T. L., Yorkston, K., Savaglio, S., & Woollcott, C. (2017). Internally versus externally cued speech in Parkinson's disease and cerebellar disease. *American Journal of Speech-Language Pathology, 26*(2S), 583–595. https://doi.org/10.1044/2017_AJSLP-16-0109

Weismer, G. (1984). Acoustic descriptions of dysarthric speech: Perceptual correlates and physiological inferences. *Seminars in Speech and Language, 5*(04), 293–314. https://doi.org/10.1055/s-2008-1064291

Weismer, G., Yunusova, Y., & Bunton, K. (2012). Measures to evaluate the effects of DBS on speech production. *Journal of Neurolinguistics, 25*(2), 74–94. https://doi.org/10.1016/j.jneuroling.2011.08.006

Whitfield, J. A., & Goberman, A. M. (2014). Articulatory–acoustic vowel space: Application to clear speech in individuals with Parkinson's disease. *Journal of Communication Disorders, 51,* 19–28. https://doi.org/10.1016/j.jcomdis.2014.06.005

Whitfield, J. A., & Goberman, A. M. (2017). Articulatory–acoustic vowel space: Associations between acoustic and perceptual measures of clear speech. *International Journal of Speech-Language Pathology, 19*(2), 184–194. https://doi.org/10.1080/17549507.2016.1193897

Yorkston, K. M., Beukelman, D. R., Strand, E. A., & Hakel, M. (Eds.). (2010). *Management of motor speech disorders in children and adults* (3rd ed.). Pro-Ed.

Yorkston, K. M., Beukelman, D. R., & Tice, R. (1996). *Sentence Intelligibility Test* [Computer software]. Tice Technology Services.

Yorkston, K. M., Hammen, V. L., Beukelman, D. R., & Traynor, C. D. (1990). The effect of rate control on the intelligibility and naturalness of dysarthric speech. *Journal of Speech and Hearing Disorders, 55*(3), 550–560. https://doi.org/10.1044/jshd.5503.550