

Relationship between acoustic measures and speech naturalness ratings in Parkinson's disease: A within-speaker approach

Marie Klopfenstein

To cite this article: Marie Klopfenstein (2015) Relationship between acoustic measures and speech naturalness ratings in Parkinson's disease: A within-speaker approach, Clinical Linguistics & Phonetics, 29:12, 938-954, DOI: [10.3109/02699206.2015.1081293](https://doi.org/10.3109/02699206.2015.1081293)

To link to this article: <https://doi.org/10.3109/02699206.2015.1081293>



Published online: 24 Sep 2015.



Submit your article to this journal [↗](#)



Article views: 408



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 5 View citing articles [↗](#)

Relationship between acoustic measures and speech naturalness ratings in Parkinson's disease: A within-speaker approach

MARIE KLOPFENSTEIN

Department of Special Education and Communication Disorders, Southern Illinois University, Edwardsville, IL, USA

(Received 27 May 2015; revised 22 July 2015; accepted 5 August 2015)

Abstract

This study investigated the acoustic basis of across-utterance, within-speaker variation in speech naturalness for four speakers with dysarthria secondary to Parkinson's disease (PD). Speakers read sentences and produced spontaneous speech. Acoustic measures of fundamental frequency, phrase-final syllable lengthening, intensity and speech rate were obtained. A group of listeners judged speech naturalness using a nine-point Likert scale. Relationships between judgements of speech naturalness and acoustic measures were determined for individual speakers with PD. Relationships among acoustic measures also were quantified. Despite variability between speakers, measures of mean F0, intensity range, articulation rate, average syllable duration, duration of final syllables, vocalic nucleus length of final unstressed syllables and pitch accent of final syllables emerged as possible acoustic variables contributing to within-speaker variations in speech naturalness. Results suggest that acoustic measures correlate with speech naturalness, but in dysarthric speech they depend on the speaker due to the within-speaker variation in speech impairment.

Keywords: Fundamental frequency, intensity, phrase-final syllable lengthening, pitch accent, prosody, speech naturalness, speech rate, vocalic nucleus

Introduction

Naturalness is an overall measure of an individual's speech that is closely related to the individual's prosody and intelligibility. As one recent definition of naturalness states, "Speech is natural if it conforms to the listener's standard of rate, rhythm, intonation and stress patterning, and if it conforms to the syntactic structure of the utterance being produced" (Yorkston, Beukelman, Strand, & Hakel, 2010, p. 288). The concept of naturalness is used widely with various communication disorders in clinical and research applications, dysarthria included. For example, naturalness has been used in the investigation of perceptual differences between the speech of non-stutterers and post-treatment stutterers (Ingham & Onslow, 1985; Onslow & Ingham, 1987; Runyan, Hames, & Proseck, 1982; Sacco, Metz, & Schiavetti, 1992), the relationship between acoustic variables and natural speech (Metz, Schiavetti, & Sacco, 1990), and

the effect of finger spelling and perceived naturalness of speech during simultaneous communication (Schiavetti, Whitehead, Whitehead, & Metz, 1998), to name only a few.

Speech naturalness is commonly assessed on a nine point Likert scale (Ingham, Gow, & Costello, 1985; Onslow, Adams, & Ingham, 1992). While this procedure yields a numerical score that can quantify the difference between impaired and unimpaired speech, a naturalness rating alone cannot provide insight on which production variables are contributing to unnaturalness. There have been calls for more research on what aspects of speech contribute to perceived naturalness (Linebaugh & Wolfe, 1984; Metz et al., 1990; Southwood & Weismer, 1993; Yorkston, Beukelman, Minifie, & Sapis, 1984; Yorkston et al., 2010), but there have been very few studies that have investigated the relationship between speech naturalness and acoustic measures and then only English speech. Metz et al. (1990) found that voice onset time (VOT) and sentence duration correlated with and predictive of speech naturalness in a study comparing the speech of individuals who stutter and controls. Linebaugh and Wolfe (1984) examined the relationship between mean syllable duration, intelligibility and naturalness in individuals with either spastic or ataxic dysarthria. Mean syllable duration correlated significantly with intelligibility and naturalness for the spastic dysarthric speakers, but not for the ataxic group. To date, no other study has been published on the relationship between speech naturalness and speech production measures or acoustic measures in other dysarthria subtypes. This study addresses this gap by investigating the speech of individuals with hypokinetic dysarthria secondary to Parkinson's disease (PD).

Numerous studies have investigated the speech characteristics of hypokinetic dysarthria that are of interest to this study. Hypokinetic dysarthria is associated with a cluster of speech characteristics that have been coined as prosodic insufficiency: short rushes of speech, variable speech rate, breathy or harsh voice, low pitch, reduced loudness or hypophonia, monopitch or monoloudness and hypernasality (Darley, Aronson, & Brown, 1969a,b; Duffy, 1995; Kent & Rosenbek, 1982; Ludlow & Bassich, 1984). As several studies have suggested that individuals with hypokinetic dysarthria secondary to Parkinson's disease (PD) have perceptually more rapid speech, researchers have looked at the duration of speech segments as confirmation of this observation. These reports indicate that for a small proportion of PD subjects, phrase, syllable or segment durations are indeed shorter than those of normal speakers (Canter, 1963; Forrest, Weismer, & Turner, 1989; Uziel, Bohe, Cadilhac, & Passouant, 1975; Weismer, 1984; Weismer, Kimelman, & Gorman, 1985). Another study (Ludlow, Connor, & Bassich, 1987) looked at sentence and syllable duration at different speech rates in PD patients. Those who had the least amount of change in sentence duration between the two rates were those least able to change their syllable durations, suggesting that the perceptually abnormal speaking rates of individuals with hypokinetic dysarthria are related to problems controlling production time during speech execution. This short review illustrates some of the prosodic aspects of hypokinetic dysarthria. It has been argued that although naturalness and intelligibility are related dimensions of speech, prosody may serve as a distinguishing feature between the two (Dagenais & Wilson, 2002; Whitehill & Chun, 2002). Therefore, prosodic characteristics of speech were chosen as a starting point for the investigation of what speech features correlate with perceived naturalness.

Summary and purpose

Despite the importance of speech naturalness to the assessment and intervention of dysarthria, there is little known about what aspects of speech relate to perceived naturalness, let alone what acoustic variables might correlate with naturalness (Yorkston et al., 2010). One challenge in attempting this type of research is the heterogeneous nature of the dysarthrias, but some studies

have used a within-speaker, across utterance design to avoid the third-variable effect (Feenaughty et al., 2014; Yunusova et al., 2005). For example, there are a variety of measures that could relate to speech naturalness and may also serve as global metrics of severity. However, some of these measures could be integral to naturalness, meaning that when modified by treatment improved naturalness results, or they covary with severity and only relate because they are associated with naturalness by means of severity as a third variable. Therefore, this study uses a single subject design to investigate potential acoustic variables related to within-speaker variations in naturalness in hypokinetic dysarthria. As it is unlikely to find a group of speakers that are matched for overall speech naturalness or across-utterance variations in naturalness, each speaker serves as their own control.

Currently, the literature only hints at a relationship between prosody and naturalness and does not address whether some aspects of prosody are more important in listener’s perceptions of speech naturalness. This study endeavours to address this gap by investigating several acoustic correlates of prosody and their relationship to naturalness. Acoustic measures of fundamental frequency, phrase-final syllable lengthening, intensity and speech rate were selected as suprasegmental acoustic measures for investigation.

Methods

Speakers

Four individuals with hypokinetic dysarthria secondary to Parkinson’s disease participated in this study. The research plan was reviewed and approved by the author’s Institutional Review Board (IRB) and informed consent was obtained from subjects prior to data collection. Characteristics of these subjects are included in Table 1. The subjects were all recruited from a support group for individuals with PD in East Texas. All had at least 2 years of post-secondary education and were monolingual English speakers. Two of the four had a history of hearing impairment including S2, who wore hearing aids in both ears, and S3, who reported a mild hearing impairment in his left ear and a hearing aid that he currently does not use. Speakers reported no history of other neurological disease, neurosurgical treatment or cognitive impairment.

The severity of each subjects’ dysarthria was determined by two experienced speech-language pathologists and the author. Each listened to a short sample of connected speech from each subject and independently assigned an estimate of severity to each speaker on a scale of mild, moderate or severe, given the lack of a standard metric for determining severity. All judgments were arrived at independently and judges were in agreement on the severity of each speaker’s dysarthria.

Table 1. Subject characteristics are summarised.

Subject	Age	Sex	Years post-diagnosis	History of speech therapy	Severity	Average naturalness rating
S1	78	M	5	Yes	Mild	4.05
S2	76	F	4	Yes	Moderate	6.09
S3	79	M	18	Yes	Mild	3.12
S4	67	M	9	No	Severe	6.27
Mean	75		9			4.68
SD	5		6			1.63
Range	67–79		4–18		Mild-Severe	3.12–6.27

The average naturalness rating is the average perceptual judgment of speech naturalness on a scale of 1 (most natural) to 9 (least natural) from 66 listeners.

Procedures for obtaining speech naturalness ratings are explained in more detail in the section concerning the perceptual task below.

Speech task and procedure

All speech samples were collected in a sound-isolated booth. Speakers were recorded using a cardioid condenser microphone (Audio Technica AT2020) placed at a distance of 30 cm from subjects' lips. Samples were directly digitised using a digital audio solid state recorder (Marantz PMD660) at a sampling rate of 44.1 kHz with 16-bit quantisation, and were later re-sampled at 22.05 kHz for acoustic analysis. Anti-aliasing low-pass filtering was performed prior to re-sampling at the Nyquist frequency, 9.8 kHz. A 1000 Hz calibration tone also was recorded for use in calculating SPL.

Three different kinds of speech sample were collected from each subject: individual sentences, a short story and spontaneous speech. For the individual sentences, subjects read four sets of sentences containing either words from a set of 25 monosyllabic or a set of 25 disyllabic words. Words were phonetically balanced as much as possible so that duration of stressed and unstressed syllables could be compared across words and for ease of measuring duration aspects of the signal. When possible, long vowels ([ɑ], [i], [ɔ], [u]) before voiced stops were included in stressed syllables so that length in these syllables would be consistent between words. The first two sets of sentences contained the words in the carrier sentence, "No, he said ____." Subjects were asked to read the sentences with emphasis on the last word, as if correcting someone who had misheard a statement they had made. The final two sets of sentences contained the same words used previously in phrase-final position, but in 50 short sentences at a simple reading level, i.e. containing 5–10 words. Subjects were given no instructions to emphasise any particular words in these sets.

Subjects were also asked to read a short children's story. Speakers were instructed to read the story aloud as if they were reading to a child in order to maximise prosodic aspects of their speech. The story was divided up approximately into thirds and the middle third, containing 404 words, was selected for analysis in order to control for the effects of warming up and vocal fatigue. One subject was only able to read half of the story due to the respiratory involvement of her PD. However, due to the repetition involved in the story, sentences matching the selected section used for analysis for the other three speakers were found within what the subject could read and were included in the data to be analysed.

Finally, two samples of spontaneous speech were collected. In the first spontaneous speech task, subjects were asked to watch a short (6 min) Pink Panther cartoon, "Pickled Pink." Subjects were then asked to recall what occurred in the cartoon to the best of their ability and retell the action, as if telling a friend about what they saw. In the second sample, subjects were recorded in conversation dyads or triads with either the author, or family members and caretakers. Topics of conversation were chosen based on background information gathered at the beginning of data collection. Conversations were recorded for at least 5 min. Sentences were selected for analysis from the middle of the conversation to control for the effects of vocal fatigue.

Perceptual task and procedure

Listeners were 69 speech-language pathology students. The listeners included one PhD student, 20 second year Master's level students who had completed coursework in motor speech disorders, 22 first year Master's level students enrolled in a craniofacial anomalies course and 26 undergraduate students enrolled in a course on articulation and phonological disorders.

These participants served as naturalness judges during the latter half of the study. None of the students reported having any history of hearing impairment.

In order to rate speech samples, the Experiment MFC function in Praat (Boersma & Weenink, 2014) was run through a script to provide samples for rating and record the rating given for each item by each participant. Speech samples presented to listeners came from the sentence, short story and spontaneous speech tasks described previously. Each speech sample was a single sentence taken from each of these contexts. Sixty-five sentences per speaker were used from the sentence reading task, for a total of 260 sentences from all four participants. Fourteen sentences per speaker, for a total of 56 sentences, were also included from the short story reading task. All of these sentences were matched across speakers. Naturally, sentences from the spontaneous speech tasks could not be matched exactly between speakers. The length of the spontaneous speech sample for the video recall task varied between speakers, so five sentences per speaker were included for naturalness rating, for 20 sentences total. Sentences that described the same action in the video were chosen as much as possible. The recorded conversation dyads and triads were divided up approximately into thirds for each speaker, in order to obtain a representative sample from the beginning, middle and end of a conversation. Five sentences were selected for naturalness judgments from each third for each speaker, for a total of 15 sentences per speaker and 60 sentences overall. In total, 396 sentences were selected from the previous speaking tasks for naturalness ratings. Of those sentences, 40 (10%) were repeated in the rating task to test for intra-judge reliability. Listeners rated a total of 436 sentences from all four speakers or 109 sentences per speaker. Sentences were presented individually and in a random order generated by the program for each listener.

Listeners read instructions on how to play each sample and provide a rating; immediately afterward they completed a practice set of five samples unrelated to the experimental samples in order to familiarise themselves with the procedure¹. Each listener was instructed to wear headphones adjusted to his or her most comfortable listening level in order to reduce ambient noise in the room and maximise audio reception of the speech signal. Listeners were allowed to repeat a speech sample as many times as they wished before rating each sentence and had the option of going back one sample if they felt they had made a mistake. Sentences were rated on a 1–9 Likert scale, following the procedures of Martin Haroldson, and Triden (1984). Listeners were prompted to take a short break after every 100 samples.

Naturalness ratings for each of the four PD subjects were compiled for the 69 student raters. Intra-rater reliability based on each listener's consistency when rating repeated sentences was calculated using Cronbach's alpha. Cronbach's alpha is commonly used to determine internal consistency between items in Likert-type scales and employs a reliability coefficient that ranges between zero and one. Following the guidelines of George and Mallery (2003), the ratings from three listeners were eliminated from further analysis because their alpha values were below the minimum reliability level of 0.6 selected for this study.

Acoustic measures and procedures

All acoustic measurements were made using Praat version 5.3.80 (Amsterdam, The Netherlands). They can be roughly grouped into two categories: temporal and spectral measures. The procedures used to obtain measurements are described in detail below.

Temporal measures

¹Practice samples were sentences taken from the Rainbow Passage (Fairbanks, 1960) as recorded by the author.

Each utterance was annotated in a Praat TextGrid object; temporal landmarks at the sentence, word and syllable levels were created. Measurement of syllable duration was achieved by viewing the spectrogram in conjunction with the acoustic waveform according to criteria adopted from Flipsen (2003). Specifically, syllables were judged to begin at the onset of F1 energy for vowels and resonant consonants, the onset of broadband noise for fricative consonants and the onset of the burst release for stop and affricate consonants. Words and sentences were annotated on separate tiers in a similar fashion. Duration measures were collected for syllable duration, including phrase-final syllable duration and utterance length. Articulation rate was calculated for each utterance as the number of syllables per second, minus pauses. Average syllable duration (ASD) was calculated as speaking time (pauses excluded) per number of syllables.

Spectral measures

Pitch accent range within falling pitch contours at the conclusion of declarative sentences were also measured in procedures adapted from Snow (1998). The acoustic measures used to quantify pitch contour were duration and F0 contour of the vocalic nucleus of the final syllable of each declarative sentence. For the purposes of this study, the vocalic nucleus was defined as the vowel plus any sonorant segments in the syllable rime, excluding any obstruents in the syllable onset or rime. In order to measure duration of the vocalic nucleus, the beginning and ending boundaries were defined as the first or last periodic cycle that was visually distinct in the waveform display. F0 was measured at the beginning and ending of each vocalic nucleus. The range of the pitch contour was calculated as the absolute difference between the maximum F0 and minimum F0 and expressed in cents (one octave = 12 semitones = 1200 cents). This scale was used instead of the frequency measure of Hertz in order to closer approximate perceptually equivalent units. The pitch contour was identified as falling if the maximum F0 preceded the minimum and rising if the maximum F0 followed the minimum. In addition, F0 contours were calculated for each utterance and max F0, mean F0 and F0 range values were extracted. Similarly, intensity contours were generated in Praat; max, min and mean values were obtained.

Statistical analysis

Levene's test for homogeneity of variance was used to assess the equality of variance between speakers for each acoustic measure. This assumption was not met for many of the variables; therefore, non-parametric tests were used to compare acoustic measures between PD speakers. Spearman rank order correlations were used to evaluate the strength of association between the various acoustic variables and speech naturalness within each PD subject. According to Cohen (1988), variables with a correlation of ± 0.30 or stronger have a moderately strong association. Therefore, moderate correlations among acoustic measures and speech naturalness are considered meaningful in this study in addition to correlations that met the standard criterion for statistical significance ($p < 0.05$). Past studies of dysarthria have considered similar magnitudes of correlations between acoustic variables and intelligibility meaningful (Feenaughty, Tjaden, & Sussman, 2014; Kim, Kent, & Weismer, 2011; Tjaden & Wilding, 2011; Weismer, Jeng, Laures, Kent, & Kent, 2001; Yunusova et al., 2005, 2012). To the best of the author's knowledge, no study has attempted to correlate acoustic variables with speech naturalness in dysarthric speakers to date. Consequently, a conservative approach to data interpretation was taken in order to not miss potential trends of interest. Correlation analysis was also used to investigate the relationship between all acoustic measures.

Results

Comparison of speech naturalness and acoustic measurements for PD speakers

Descriptive characteristics for naturalness judgments of each utterance type for each participant are summarised in Table 2. Average speech naturalness for all participants was 4.68 (SD = 1.63). A Kruskal–Wallis H test was run to determine if there were differences in speech naturalness ratings between speakers. Distributions of naturalness scores were not similar for all speakers, as assessed by visual inspection of a boxplot. The distributions of naturalness ratings were statistically significantly different between groups, $\chi^2(3) = 235.662$, $p < 0.001$. Pairwise comparisons were performed using Dunn’s (1964) procedure with a Bonferroni correction for multiple comparisons. Statistical significance was accepted at the $p < 0.0083$ level. Values are mean ranks unless otherwise stated. This *post hoc* analysis revealed statistically significant differences in naturalness ratings between all speakers: S3 (2.97) and S1 (4.00; $p < 0.001$), S3 (2.97) and S2 (5.98; $p < 0.001$), S3 (2.97) and S4 (6.41; $p < 0.001$), S1 (4.00) and S2 (5.98; $p < 0.001$), S1 (4.00) and S4 (6.41; $p < 0.001$) and S2 (5.98) and S4 (6.41; $p < 0.001$).

Figure 1 shows participant means and standard deviations for the following acoustic measures from speech runs: mean F0, F0 range, intensity range, articulation rate and average syllable duration. Speech runs were defined as a stretch of speech bounded by silent periods between words of at least 200 ms (Turner & Weismer, 1993). A Kruskal–Wallis H test was run to determine if there were differences in acoustic measures taken in speech runs between each participant. Distributions of acoustic measures were similar for all subjects, as assessed by visual inspection of a boxplot. Median scores were statistically significantly different between subjects for all measures: mean F0, $\chi^2(3) = 279.380$, $p < 0.001$; F0 range, $\chi^2(3) = 237.302$, $p < 0.001$; intensity range, $\chi^2(3) = 48.256$, $p < 0.001$; articulation rate, $\chi^2(3) = 46.278$, $p < 0.001$ and average syllable duration, $\chi^2(3) = 20.020$, $p < 0.001$.

Figure 2 shows speaker means and standard deviations for the following syllable level acoustic measures: duration of final stressed and unstressed syllables, vocalic nucleus length of final stressed and unstressed syllables and pitch accent of final stressed and unstressed syllables. Mann–Whitney *U* tests were run for each subject to determine if there were differences in syllable-level acoustic measures between stressed and unstressed syllables. Distributions of each acoustic measure were similar, as assessed by visual inspection. For S1, median syllable duration was statistically significantly longer in final stressed syllables (404) than in final unstressed syllables (163.5), $U = 622.5$, $z = -9.882$, $p < 0.001$. Median vocalic nucleus length was statistically significantly longer in final stressed syllables (220) than in final unstressed syllables (109.5), $U = 1,089.5$, $z = -8.696$, $p < 0.001$. Median pitch accent was statistically significantly longer in final stressed syllables (1006.22) than in final unstressed syllables (418.69), $U = 2,236$, $z = -5.236$, $p < 0.001$. For S2, median syllable duration was statistically significantly longer in final stressed syllables (483) than in final unstressed syllables (277.5), $U = 250.5$, $z = -7.606$, $p < 0.001$. Median vocalic nucleus length was statistically significantly longer in final stressed syllables (371.5) than in final unstressed syllables (256), $U = 689.5$, $z = -5.037$, $p < 0.001$.

Table 2. Descriptive characteristics are summarised of naturalness judgments from 66 listeners for each PD speaker.

Subject	S1	S2	S3	S4	Total
Mean	4.05	6.09	3.12	6.27	4.68
Min	2.43	4.27	1.75	3.81	1.75
Max	6.00	8.32	6.32	8.46	8.46
SD	0.86	0.76	0.92	1.24	1.63

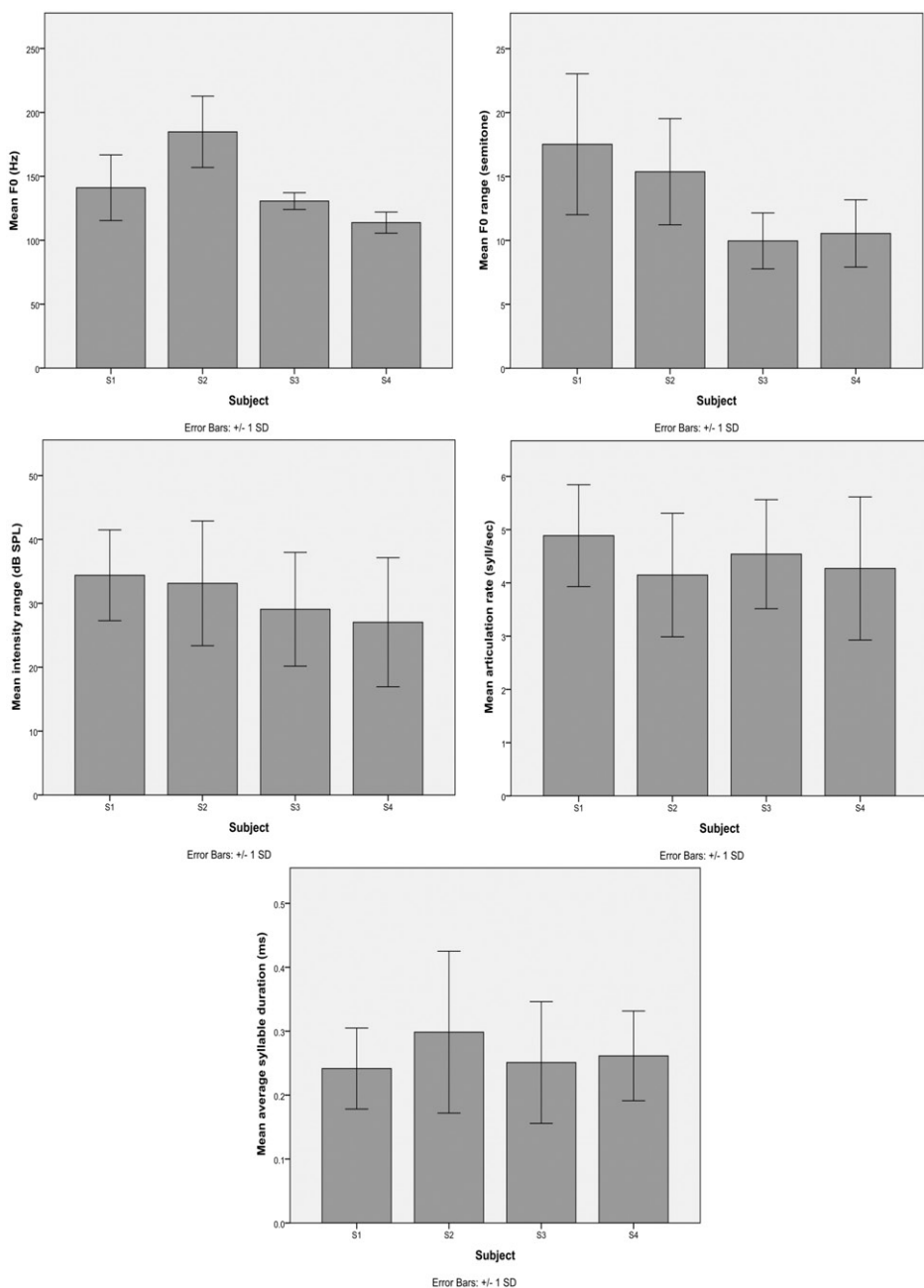


Figure 1. Mean and standard deviation bars are reported for mean F0 in Hz (upper left panel), mean F0 range in semitone (upper right panel), mean intensity rate in decibels (middle left panel), mean articulation rate in syllables per second (middle right panel) and mean average syllable duration in milliseconds (bottom panel) for each speaker.

Median pitch accent was statistically significantly longer in final stressed syllables (1017.81) than in final unstressed syllables (588.15), $U = 646$, $z = -5.291$, $p < 0.001$. For S3, median syllable duration was statistically significantly longer in final stressed syllables (521) than in final

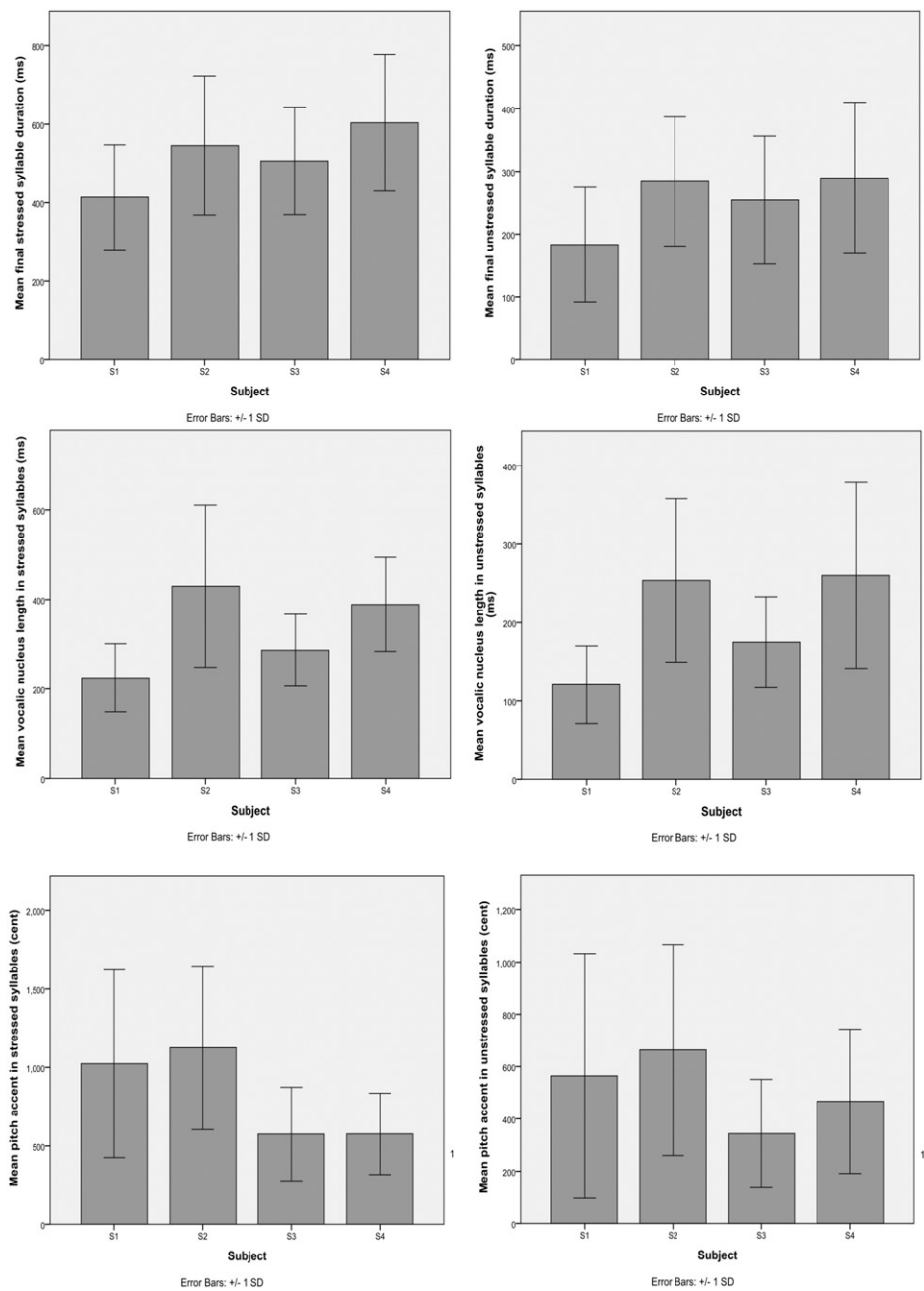


Figure 2. Mean and standard deviation bars are reported for final syllable duration in milliseconds (upper row), mean vocalic nucleus length in milliseconds (middle row) and mean pitch accent in cent (bottom row) for each speaker.

unstressed syllables (228.5), $U = 370$, $z = -8.443$, $p < 0.001$. Median vocalic nucleus length was statistically significantly longer in final stressed syllables (293) than in final unstressed syllables (169.5), $U = 640$, $z = -7.330$, $p < 0.001$. Median pitch accent was statistically significantly longer

Table 3. Spearman rank order correlation results are reported.

Speaker	Mean F0 (Hertz)	F0 range (Semitones)	Intensity range (dB SPL)	Articulation rate (syllables/s)	Average syllable duration (ms)
S4	-0.217	-0.012	0.244*	0.503*	- 0.387*
S2	-0.195	0.137	0.082	-0.109	0.290*
S1	0.279*	0.032	0.276*	0.226*	0.053
S3	-0.429*	0.057	0.322*	0.023	0.118

Results summarise the association between acoustic variables for speech runs and speech naturalness within speakers. The sign of the correlation indicates the direction of the association. Correlations of ± 0.30 or greater are in bold. An asterisk indicates significant correlations ($p < 0.05$).

Table 4. Spearman rank order correlation results are reported.

Speaker	Duration of final stressed syllables (ms)	Vocalic nucleus length of stressed final syllables (ms)	Pitch accent of final stressed syllables (cents)	Duration of final unstressed syllables (ms)	Vocalic nucleus length of unstressed final syllables (ms)	Pitch accent of final unstressed syllables (cents)
S4	-0.275	-0.231	- 0.465*	-0.285	- 0.347	- 0.531*
S2	-0.232	-0.285	0.039	0.153	0.223	0.537*
S1	- 0.357*	-0.298*	-0.151	0.014	-0.020	0.077
S3	-0.153	-0.102	- 0.419*	0.536*	-0.020	0.121

Results summarise the association between acoustic variables specific to either stressed or unstressed syllables and speech naturalness within speakers. The sign of the correlation indicates the direction of the association. Correlations of 0.30 or greater are in bold. An asterisk indicates significant correlations ($p < 0.05$).

in final stressed syllables (525.70) than in final unstressed syllables (291.81), $U = 1240.5$, $z = -4.854$, $p < 0.001$. For S4, median syllable duration was statistically significantly longer in final stressed syllables (613) than in final unstressed syllables (317), $U = 71.5$, $z = -6.208$, $p < 0.001$. Median vocalic nucleus length was statistically significantly longer in final stressed syllables (382) than in final unstressed syllables (264), $U = 262$, $z = -4.183$, $p < 0.001$. Median pitch accent was statistically significantly longer in final stressed syllables (571.19) than in final unstressed syllables (430.52), $U = 451$, $z = -2.174$, $p < 0.001$.

Correlations between naturalness and acoustic measures for speakers with PD

Table 3 reports Spearman rank order correlations reflecting the strength of association between rated naturalness and acoustic variables for entire speech runs for each speaker. Speakers are listed from least (S4 = 6.27) to most (S3 = 3.12) natural, according to their average naturalness rating. Text in bold indicates correlations of ± 0.30 or stronger; statistically significant correlations ($p < 0.05$) are indicated with asterisks. At least one acoustic variable was significantly correlated with speech naturalness in each speaker.

Table 4 reports Spearman rank order correlations between acoustic variables in stressed and unstressed syllables and rated naturalness for each speaker, following the same conventions as Table 3. At least one acoustic variable for stressed or unstressed syllables was significantly correlated with speech naturalness in each speaker.

Mean F0, intensity range, articulation rate and average syllable duration

As reported in Table 3, mean F0 was significantly correlated with two of the four speakers; for S3, there was a moderate and significant correlation. The absolute magnitude of correlations ranged from 0.195 to 0.429. The direction of the effect varied between speakers, but for S3 higher F0 was associated with increased speech naturalness.

Intensity range was significantly correlated with speech naturalness for three out of four speakers, but the strength of the correlation for all speakers was weak. The absolute magnitude of correlations ranged from 0.082 to 0.322. The direction of the relationship was the same for all speakers such that reduced intensity range was associated with higher naturalness ratings.

Articulation rate was moderately correlated with speech naturalness for one out of four speakers and the correlation was significant for the same plus one other speaker. The absolute magnitude of correlations ranged from 0.023 to 0.503. The direction of the relationship varied across speakers, but for S4 increased articulation rate was associated with lower naturalness ratings.

Average syllable duration was significantly correlated with speech naturalness for two out of four speakers. The absolute magnitude of correlations ranged from 0.053 to 0.387. The direction of the relationship varied across speakers, but for the speaker with the strongest correlation that was also significant, S4, increased syllable duration was associated with higher naturalness ratings.

Syllable duration, vocalic nucleus and pitch accent of final stressed and unstressed syllables

As reported in Table 4, duration of final stressed syllables was moderately strong and significantly correlated with speech naturalness for one of four speakers. The absolute magnitude of correlations ranged from 0.153 to 0.357. The direction of the relationship indicates that increased final stressed syllable duration was associated with higher naturalness ratings. The duration of final unstressed syllables was moderately strong and significantly correlated with naturalness with one out of four speakers as well. The absolute magnitude of the correlations ranged from 0.014 to 0.536. The direction of the relationship varied across speakers. For S3, decreased final unstressed syllable duration was associated with higher naturalness ratings.

Vocalic nucleus length was moderately correlated with speech naturalness in unstressed final syllables for one out of four speakers, but was weakly, yet significantly correlated with naturalness for one subject in stressed final syllables. The absolute magnitude of correlations ranged from 0.20 to 0.347 in the former group and 0.102 to 0.298 in the latter. The direction of the relationship in stressed final syllables was such that increased vocalic nucleus length was associated with higher naturalness ratings. The direction of the relationship varied across speakers for unstressed final syllables.

Pitch accent was moderately and significantly correlated with speech naturalness in both unstressed and stressed final syllables for three out of four speakers, with one speaker (S4) showing an moderate association for both syllable types. The absolute magnitude of correlations ranged from 0.077 to 0.537 in the unstressed syllables and 0.039 to 0.465 in the stressed syllables. The direction of the association varied across speakers in both groups. As indicated in the left panel of Figure 3, decreased pitch accent was correlated with increased naturalness for S2 ($\rho = 0.537$) while the right panel indicates increased pitch accent was correlated with increased naturalness for S4 ($\rho = -0.531$) in final unstressed syllables. In final stressed syllables, the direction of the relationship for S3 ($\rho = -0.419$) and S4 ($\rho = -0.465$) was such that increased pitch accent was correlated with higher naturalness ratings.

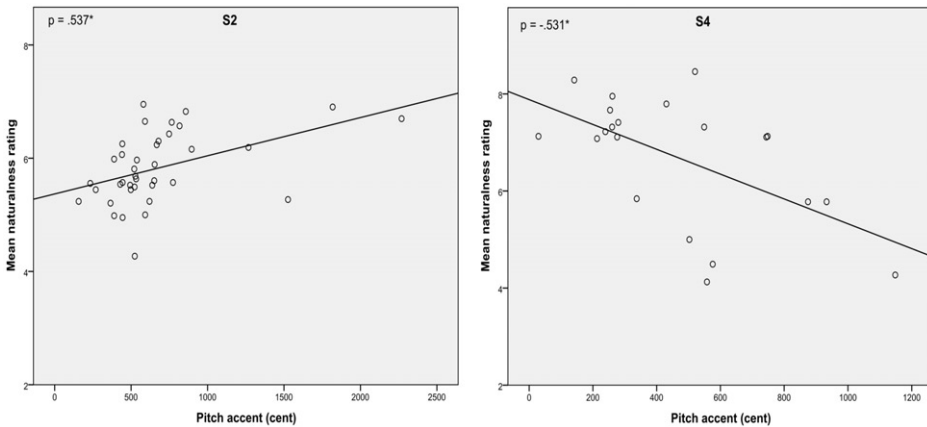


Figure 3. The relationship between mean naturalness rating and pitch accent in final unstressed syllables is reported for speakers S2 (left panel) and S4 (right panel). Each circle represents a data point from an utterance.

Table 5. Spearman rank order inter-correlations among speech run acoustic variables are reported.

Acoustic variables	(1) M F0 (Hz)	(2) F0 range (Semitones)	(3) Intensity range (dB SPL)	(4) Articulation rate (syllables/ second)	(5) Average syllable duration (ms)
(1) M F0 (Hz)	1				
(2) F0 range (Semitones)	0.177*	1			
(3) Intensity range (dB SPL)	0.197*	0.341*	1		
(4) Articulation rate (syllables/second)	-0.009	0.066	-0.064	1	
(5) Average syllable duration (ms)	0.050	0.116*	0.335*	-0.838*	1

Alpha level = 0.05. The sign of the correlation indicates the direction of the association. An asterisk indicates significant correlations ($p < 0.05$).

Summary of within-speaker correlations

The strength and significance of the relationship between most acoustic measures and speech naturalness varied. The direction of the relationship also varied. Measures of mean F0, intensity range, articulation rate, average syllable duration, duration of final stressed syllables, pitch accent of final stressed syllables, duration of final unstressed syllables, vocalic nucleus length of final unstressed syllables and pitch accent of final unstressed syllables emerged as possible acoustic variables contributing to within-speaker variations in speech naturalness. For pitch accent in final unstressed syllables, the direction of the relationship varied between speakers. In sum, all of the speakers had at least one moderately strong correlation between an acoustic variable in speech runs or in a final syllable and speech naturalness.

Correlations between acoustic variables

Correlations between acoustic variables from speech runs are reported in Table 5. This table suggests significant associations among most of the acoustic variables when data were pooled across speakers, with the exception of articulation. Overall, only average syllable duration was

unsurprisingly strongly correlated with other acoustic variable, average syllable duration, while mean F0, F0 range and intensity range were less strongly correlated with other acoustic variables.

Discussion

The purpose of this study was to investigate potential relationships between speech naturalness and acoustic variables connected to prosodic aspects of speech in speakers with dysarthria secondary to PD. The strength of the associations as well as the direction of the relationships between most acoustic measures and naturalness varied. Further consideration of these results and their implications are discussed below.

Relationship between mean F0 and speech naturalness

For two of the four speakers, there was a statistically significant relationship between mean F0 and speech naturalness, but the direction of the relationship varied. For S3, higher mean F0 was related to higher ratings of speech naturalness. For S1, the opposite was true and lower mean F0 was associated with increased naturalness. Both speakers are male, so the explanation cannot lie with gender differences. As seen in Figure 1, the mean F0 values for S1 and S3 are roughly the same, but there is more variability in S1's values. In addition, both speakers are on the high end of the male range for mean F0, but this not atypical for older males (Awan, 2001). Therefore, it may be the increased variability of S1's voice that is responsible for the differing results. There also may be other aspects of voice quality that vary along with pitch for each individual. This could result in higher pitch improving voice quality and naturalness in one individual, whereas in another the opposite may be true.

Relationship between intensity range and speech naturalness

For three out of four speakers, **reduced intensity range was statistically significant and weakly to moderately associated with higher naturalness ratings**. Given that reduced loudness is a common feature of hypokinetic dysarthria (Darley et al., 1969a,b; Duffy, 1995; Kent & Rosenbek, 1982; Ludlow & Bassich, 1984), at first glance one may expect that a speaker with a low intensity range would be rated less natural. However, the naturalness ratings and acoustic measurements discussed here were made for each speech run, not for longer stretches of speech. Data indicate that speakers who have large average intensity ranges overall in their speech are not necessarily perceived as more unnatural than other speakers. S1, a speaker with mild dysarthria rated more natural than other speakers in this study [see Klopfenstein (2012) for further discussion], had a large intensity range overall. This may indicate that a speaker who uses a large intensity overall in connected speech may not be perceived as unnatural; only someone who uses a large range within a single utterance would be perceived as less natural sounding.

Consequently, the reduction of naturalness is mostly likely associated with a decrease in intensity over a single speech run, rather than an increase. The larger intensity ranges may have been the result of subjects speaking at a normal loudness level, but then losing breath support over the utterance, ending with reduced loudness. This may have implications for an intervention such as the Lee Silverman Voice Treatment (LSVT), which aims to achieve overall improvements in voice and speech through increased vocal effort and intensity. LSVT has been shown by numerous studies to result in overall increased vocal intensity (Dromey, Ramig, & Johnson, 1995; Ramig,

Countryman, Thompson, & Horii, 1995; Ramig et al., 2001; Ramig, Sapir, Fox, & Countryman, 2001; Sapir, Ramig, Hoyt, O'Brien, & Hoehn, 2002); therefore, if LSVT results in a lack of a decrease in intensity over each utterance, speakers may gain increased naturalness through the therapy, although to date this has not been investigated.

Relationship between speech rate and speech naturalness

Two of the four speakers had a statistically significant relationship between measures of speech rate, articulation rate and average syllable duration and naturalness ratings. The strength of the relationship varied from weak to moderate. For articulation rate, the direction of the relationship was the same for both speakers for whom it was significant and indicated that for these speakers increased articulation rate was associated with lower naturalness ratings. By examining average syllable duration, a statistically significant relationship was found for one speaker, S2, which was not found with articulation rate. The direction of the relationship varied, so that for S4 longer syllable duration was associated with higher naturalness ratings, while for S2 shorter syllable duration was associated with higher naturalness ratings.

The mixed results on speech rate and naturalness are interesting, because conventional wisdom suggests that reducing speaking rate may reduce speech naturalness (Yorkston et al., 2010). There is a tendency for PD subjects to have a lot of inter-subject variation in speech rate (Skodda, 2011). It could be, therefore, that there is no simple one-to-one relationship between speech rate and naturalness. In other words, it could depend on the speaker whether speeding up or slowing down speech rate will affect naturalness because it could either highlight or diminish other suprasegmentals aspects of speech contributing to naturalness, such as intonation, monotone or other voice qualities. Also, intelligibility or severity may also serve as confounding variables, such that individuals who are more unintelligible or whose dysarthrias are more severe may gain some naturalness from slower rates rather than perceived unnaturalness. This explanation would fit for S4, who had the most severe dysarthria and who has a moderate association between slower speech rates and higher naturalness ratings. Finally, the data suggest that various measurements of speech rate may be more sensitive to a given analysis than others, as articulation rate and average syllable duration was only significant for one speaker. For two other speakers, S1 and S2, only one of the two measures was significant respectively.

Relationship between phrase-final lengthening variables and speech naturalness

Looking at final stressed syllables, three of four speakers show statistically significant associations between longer final syllables and increased naturalness, regardless of whether the metric used for final syllable length was duration, vocalic nucleus length or pitch accent. This is what we would expect, because it fits this particular prosodic pattern of English, phrase-final syllable lengthening. This suggests that speakers whose speech does not fit this pattern may be perceived as having less natural speech.

However, when examining the data on unstressed final syllables, a more complex pattern emerges. For S4, acoustic measures of lengthening were moderately associated with increased naturalness. In contrast, acoustic measurements of final syllables showed a relationship between shorter final syllables and increased naturalness for S3 and decreased pitch accent and increased naturalness for S2. Even though these syllables are unstressed, they are still expected to follow the pattern of final lengthening. It may be that some of the subjects' voice quality issues played a role;

S2 in particular tended to descend into creak at the end of her utterances, a quality which may be been more noticeable in longer final syllables. This highlights the difficulty in examining the relationship between acoustic measures and naturalness in real speech, as a number of confounding variables may come into play.

Limitations and future directions

Although several statistically significant correlations were found, several factors should be kept in mind. The speakers in this study had varying levels of severity and may not be representative of dysarthric speakers with their respective severity. Also, while the within-speaker analysis was useful to identifying acoustic variables that may contribute to reduced naturalness, the findings may not generalise to the larger population of individuals with PD. A larger data set collected from a bigger subject pool could allow for results that represent the population of individuals with hypokinetic dysarthria. Finally, the heterogeneity of the speakers in this study, while typical for speakers with PD, makes it difficult to control for additional acoustic variables that may also contribute to speech naturalness.

In summary, results suggest that a within-speaker analysis can aid in identifying acoustic variables that have a relationship with speech naturalness in speakers with PD. Certain acoustic variables, like measures of mean F0, intensity range, articulation rate, average syllable duration, duration of final syllables, vocalic nucleus length of final unstressed syllables and pitch accent of final syllables show an association with within-speaker variation in naturalness for the speakers in the current study. One possible clinical implication is that speakers with PD with similar speech characteristics may benefit from therapy focused on the underlying acoustic variables associated with naturalness. As suggested by Yorkston et al. (1984), the assessment of acoustic aspects of dysarthric speech and naturalness can be used to determine the sequencing of treatment. Future studies could confirm and expand on the results presented here by experimentally manipulating samples in terms of length, frequency or other acoustic characteristics in order to determine whether certain thresholds for the naturalness of these variables exist.

Declaration of interest

The author reports no conflict of interest. The author alone is responsible for the content and writing of the paper.

References

- Awan, S. N. (2001). *The voice diagnostic protocol: A practical guide to the diagnosis of voice disorders*. Gaithersburg, MD: Aspen Publishers.
- Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer. Retrieved from <http://www.praat.org/>.
- Canter, G. J. (1963). Speech characteristics of patients with Parkinson's disease: I. Intensity, pitch, and duration. *Journal of Speech and Hearing Disorders*, 28, 221–229.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.
- Dagenais, P. A., & Wilson, A. F. (2002). Acceptability and intelligibility of moderately dysarthric speech by four types of listener. In F. Windsor, M. L. Kelly, & N. Hewlett (Eds.), *Investigations in clinical phonetics and linguistics* (pp. 363–372). Mahwah, NJ: Lawrence Erlbaum Associates.
- Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research*, 12, 462–496.
- Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, 12, 246–269.

- Dromey, C., Ramig, L. O., & Johnson, A. B. (1995). Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: A case study. *Journal of Speech and Hearing Research*, 38, 751–764.
- Duffy, J. R. (1995). *Motor speech disorders: Substrates, differential diagnosis, and management*. St. Louis: Mosby.
- Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6, 241–252.
- Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed.). New York: Harper & Row.
- Feenaughty, L., Tjaden, K., & Sussman, J. (2014). Relationship between acoustic measures and judgments of intelligibility in Parkinson's disease: A within-speaker approach. *Clinical Linguistics & Phonetics*, 28, 857–878.
- Flipsen, P. (2003). Articulation rate and speech-sound normalization failure. *Journal of Speech, Language, and Hearing Research*, 46, 724–737.
- Forrest, K., Weismer, G., & Turner, G. S. (1989). Kinematic, acoustic, and perceptual analyses of connected speech produced by parkinsonian and normal geriatric adults. *The Journal of the Acoustical Society of America*, 85, 2608–2622.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston, MA: Allyn & Bacon.
- Ingham, R. J., & Onslow, M. (1985). Measurement and modification of speech naturalness during stuttering therapy. *Journal of Speech and Hearing Disorders*, 50, 261–281.
- Ingham, R. J., Gow, M., & Costello, J. M. (1985). Stuttering and Speech Naturalness: Some Additional Data. *Journal of Speech and Hearing Disorders*, 50, 217–219.
- Kent, R. D., & Rosenbek, J. C. (1982). Prosodic disturbance and neurologic lesion. *Brain and Language*, 15, 259–291.
- 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, 417–429.
- Klopfenstein, M. (2012). *Prosodic Features and Speech Naturalness in Individuals with Dysarthria*. (Unpublished doctoral dissertation). University of Louisiana at Lafayette, Lafayette, LA.
- Linebaugh, C. E., & Wolfe, V. E. (1984). Relationships between Articulation Rate, Intelligibility, and Naturalness in Spastic and Ataxic Speakers. In M. R. McNeil, J. C. Rosenbek, & A. E. Aronson (Eds.), *The dysarthrias: Physiology, acoustics, perception, management* (pp. 197–205). San Diego, CA: College-Hill Press.
- Ludlow, C. L., & Bassich, C. J. (1984). Relationships between perceptual ratings and acoustic measures of hypokinetic speech. In M. R. McNeil, J. C. Rosenbek, & A. E. Aronson (Eds.), *The dysarthrias: Physiology, acoustics, perception, management* (pp. 163–192). San Diego: College-Hill Press.
- Ludlow, C. L., Connor, N. P., & Bassich, C. J. (1987). Speech timing in Parkinson's and Huntington's disease. *Brain and Language*, 32, 195–214.
- Martin, R. R., Haroldson, S. K., & Triden, K. A. (1984). Stuttering and speech naturalness. *Journal of Speech and Hearing Disorders*, 49, 53–58.
- Metz, D. E., Schiavetti, N., & Sacco, P. R. (1990). Acoustic and psychophysical dimensions of the perceived speech naturalness of nonstutterers and posttreatment stutterers. *The Journal of Speech and Hearing Disorders*, 55, 516–525.
- Onslow, M., & Ingham, R. J. (1987). Speech quality measurement and the management of stuttering. *Journal of Speech and Hearing Disorders*, 52, 2–17.
- Onslow, M., Adams, R., & Ingham, R. J. (1992). Reliability of speech naturalness ratings of stuttered speech during treatment. *Journal of Speech and Hearing Research*, 35, 994–1001.
- Ramig, L. O., Countryman, S., Thompson, L. L., & Horii, Y. (1995). Comparison of two forms of intensive speech treatment for Parkinson disease. *Journal of Speech and Hearing Research*, 38, 1232–1251.
- Ramig, L. O., Sapir, S., Countryman, S., Pawlas, A. A., O'Brien, C., Hoehn, M., & Thompson, L. L. (2001). Intensive voice treatment (LSVT®) for patients with Parkinson's disease: A 2 year follow up. *Journal of Neurology, Neurosurgery & Psychiatry*, 71, 493–498.
- Ramig, L. O., Sapir, S., Fox, C., & Countryman, S. (2001). Changes in vocal loudness following intensive voice treatment (LSVT®) in individuals with Parkinson's disease: A comparison with untreated patients and normal age-matched controls. *Movement Disorders*, 16, 79–83.
- Runyan, C. M., Hames, P. E., & Prosek, R. A. (1982). A perceptual comparison between paired stimulus and single stimulus methods of presentation of the fluent utterances of stutterers. *Journal of Fluency Disorders*, 7, 71–77.
- Sacco, P. R., Metz, D. E., & Schiavetti, N. (1992). *Speech naturalness of nonstutterers and treated stutterers: Acoustical correlates. Presented at the annual meeting of the American Speech-Language-Hearing Association*, San Antonio, TX.
- Sapir, S., Ramig, L. O., Hoyt, P., Countryman, S., O'Brien, C., & Hoehn, M. (2002). Speech loudness and quality 12 months after intensive voice treatment (LSVT®) for Parkinson's disease: A comparison with an alternative speech treatment. *Folia Phoniatrica et Logopaedica*, 54, 296–303.
- Schiavetti, N., Whitehead, R., Whitehead, B., & Metz, D. (1998). Effect of fingerspelling task on temporal characteristics and perceived naturalness of speech in simultaneous communication. *Journal of Speech, Language & Hearing Research*, 41, 5–17.

- Skodda, S. (2011). Aspects of speech rate and regularity in Parkinson's disease. *Journal of the Neurological Sciences*, 310, 231–236.
- Snow, D. (1998). Prosodic markers of syntactic boundaries in the speech of 4-year-old children with normal and disordered language development. *Journal of Speech, Language, and Hearing Research*, 41, 1158–1170.
- Southwood, M. H., & Weismer, G. (1993). Listener judgments of the bizarreness, acceptability, naturalness, and normalcy of the dysarthria associated with amyotrophic lateral sclerosis. *Journal of Medical Speech-Language Pathology*, 1, 151–161.
- Tjaden, K., & Wilding, G. (2011). Speech and pause characteristics associated with voluntary rate reduction in Parkinson's disease and multiple sclerosis. *Journal of Communication Disorders*, 44, 655–665.
- Turner, G. S., & Weismer, G. (1993). Characteristics of speaking rate in the dysarthria associated with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research*, 36, 1134–1144.
- Uziel, A., Bohe, M., Cadilhac, J., & Passouant, P. (1975). Les troubles de la voix et de la parole dans les syndromes parkinsoniens [The voice and speech disorders in parkinsonian syndromes]. *Folia Phoniatica et Logopaedica*, 27, 166–176.
- Weismer, G. (1984). Articulatory characteristics of Parkinsonian dysarthria: Segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination. In M. R. McNeil, J. C. Rosenbek, & A. E. Aronson (Eds.), *The dysarthrias: physiology, acoustics, perception, management* (pp. 101–130). San Diego, CA: College-Hill Press.
- Weismer, G., Jeng, J., Laures, J. S., Kent, R. D., & Kent, J. F. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatica et Logopaedica*, 53, 1–18.
- Weismer, G., Kimelman, M. D. Z., & Gorman, S. (1985). More on the speech production deficit associated with Parkinson's disease. *Journal of the Acoustical Society of America*, 78, S55.
- Whitehill, T. L., & Chun, J. C. (2002). Intelligibility and acceptability in speakers with cleft palate. In F. Windsor, M. L. Kelly, & N. Hewlett (Eds.), *Investigations in clinical phonetics and linguistics* (pp. 405–415). Mahwah, NJ: Lawrence Erlbaum Associates.
- Yorkston, K. M., Beukelman, D. R., Minifie, F. D., & Sapir, S. (1984). Assessment of stress patterning. In M. R. McNeil, J. C. Rosenbek, & A. E. Aronson (Eds.), *The dysarthrias: Physiology, acoustics, perception, management* (pp. 197–205). San Diego, CA: College-Hill Press.
- Yorkston, K. M., Beukelman, D. R., Strand, E. A., & Hakel, M. (2010). *Management of motor speech disorders in children and adults* (3rd ed.). Austin, TX: Pro-Ed.
- Yunusova, Y., Green, J. R., Greenwood, L., Wang, J., Pattee, G. L., & Zinman, L. (2012). Tongue movements and their acoustic consequences in Amyotrophic Lateral Sclerosis. *Folia Phoniatica et Logopaedica*, 64, 94–102.
- Yunusova, Y., Weismer, G., Kent, R. D., & Rusche, N. M. (2005). Breath-group intelligibility in dysarthria: Characteristics and underlying correlates. *Journal of Speech, Language, and Hearing Research*, 48, 1294–1310.