Research Article

Predicting Intelligibility Gains in Individuals With Dysarthria From Baseline Speech Features

Annalise R. Fletcher,[a,b] Megan J. McAuliffe,[a,b] Kaitlin L. Lansford,[c] Donal G. Sinex,[b] and Julie M. Liss[d]

Purpose: Across the treatment literature, behavioral speech modifications have produced variable intelligibility changes in speakers with dysarthria. This study is the first of two articles exploring whether measurements of baseline speech features can predict speakers' responses to these modifications.
Methods: Fifty speakers (7 older individuals and 43 speakers with dysarthria) read a standard passage in habitual, loud, and slow speaking modes. Eighteen listeners rated how easy the speech samples were to understand. Baseline acoustic measurements of articulation, prosody, and voice quality were collected with perceptual measures of severity.
Results: Cues to speak louder and reduce rate did not confer intelligibility benefits to every speaker. The degree

to which cues to speak louder improved intelligibility could be predicted by speakers' baseline articulation rates and overall dysarthria severity. Improvements in the slow condition could be predicted by speakers' baseline severity and temporal variability. Speakers with a breathier voice quality tended to perform better in the loud condition than in the slow condition.
Conclusions: Assessments of baseline speech features can be used to predict appropriate treatment strategies for speakers with dysarthria. Further development of these assessments could provide the basis for more individualized treatment programs.

This study is the first of two (Fletcher, Wisler, McAuliffe, Lansford, & Liss, 2017) that investigate whether assessments of dysarthric speech can be used to predict speakers' intelligibility gains following behavioral speech modification. The larger goal of this project is to develop evidence-based protocols for determining whether a treatment technique is appropriate for a given speaker (Fletcher & McAuliffe, 2017). In the present article, we explore whether manually obtained acoustic and perceptual measurements of dysarthria can be used to statistically model differences in speakers' intelligibility gains. In the companion article (see Fletcher, Wisler, et al., 2017), we assess the predictive power of these models on cross-validation

and examine whether automated acoustic analyses can account for similar variation in treatment outcomes.

For speakers with dysarthria, the perception of reduced intelligibility can be associated with lower levels of communication participation in life situations (McAuliffe, Baylor, & Yorkston, 2016). Hence, improving intelligibility is a common goal of speech therapy. Several studies have reported that behavioral speech modifications result in improved intelligibility for speakers with a range of dysarthria subtypes (e.g., Cannito et al., 2012; Wenke, Theodoros, & Cornwell, 2011). However, these improvements have not been universal. Among treatment groups, individual speakers differ markedly in their responses to different strategies (e.g., Lowit, Dobinson, Timmins, Howell, & Kröger, 2010; Mackenzie & Lowit, 2007; Mahler & Ramig, 2012; Wenke et al., 2011), and few data have been generated to explain why some speakers experience more improvement in intelligibility in response to treatment modifications. To advocate for a particular treatment approach, we must better define the speech characteristics of those clients who do and do not achieve intelligibility gains in response to treatment cues.

Cues to speak louder and reduce speech rate form the foundation of many well-established treatment programs (e.g., the Lee Silverman voice treatment program) and strategies (e.g., pacing boards and delayed auditory feedback). These behavioral modifications can result in changes to

[a]Department of Communication Disorders, University of Canterbury, Christchurch, New Zealand
[b]New Zealand Institute of Language, Brain & Behaviour, Christchurch
[c]School of Communication Science & Disorders, Florida State University, Tallahassee
[d]Department of Speech and Hearing Science, Arizona State University, Tempe

Correspondence to Annalise Fletcher: annalise.fletcher@canterbury.ac.nz

rate and loudness parameters and in more diffuse global acoustic changes to articulation, prosody, and voice characteristics (Baumgartner, Sapir, & Ramig, 2001; Tjaden & Wilding, 2004). For this reason, cues to speak loudly or slowly are used in the management of a large range of dysarthria etiologies (e.g., Fox & Boliek, 2012; Sapir et al., 2003; Van Nuffelen, De Bodt, Vanderwegen, Van de Heyning, & Wuyts, 2010). However, the success of these techniques is unclear. Not all speakers with dysarthria have gained intelligibility benefits from programs that use these cueing strategies (Cannito et al., 2012). Partly as a result of this variation in individual response, treatment studies often fail to demonstrate intelligibility improvements across speaker groups (e.g., Lowit et al., 2010; Mahler & Ramig, 2012).

To predict whether a client will benefit from a larger program of treatment, it is important to first determine that these clients are able to change their speech in response to cues and that these changes confer greater intelligibility. Thus far, cueing strategies applied directly to dysarthric speech have resulted in promising improvements in blinded listeners' understanding of dysarthric phrases (Hammen, Yorkston, & Minifie, 1994; McAuliffe, Kerr, Gibson, Anderson, & LaShell, 2014; Neel, 2009; Patel, 2002; Patel & Campellone, 2009; Pilon, McIntosh, & Thaut, 1998; Tjaden & Wilding, 2004; Turner, Tjaden, & Weismer, 1995; Van Nuffelen et al., 2010; Van Nuffelen, De Bodt, Wuyts, & Van de Heyning, 2009; Yorkston, Hammen, Beukelman, & Traynor, 1990). However, even among participants with the same etiology and dysarthria subtype, there has been considerable variation in treatment effects observed across studies—and not all individuals have responded positively to loud and slow treatment cues (McAuliffe, Fletcher, Kerr, O'Beirne, & Anderson, 2017). As with more complex programs of treatment, this variation across participants can prevent group outcomes from reaching statistical significance (e.g., Pilon et al., 1998; Van Nuffelen et al., 2010).

It is unclear why certain speakers respond differently to treatment cues. Studies examining the effects of loud speech on intelligibility have mainly focused on hypokinetic dysarthria, which is closely associated with Parkinson's disease (PD). Neel (2009) examined the effect of loud cued speech on intelligibility in five people with PD and hypokinetic dysarthria. Four of these speakers exhibited significantly improved intelligibility when cued to speak loudly. Tjaden and Wilding (2004) also found significant intelligibility gains in a group of 12 speakers with PD. However, for their 15 participants with multiple sclerosis, the cue to speak loudly did not significantly improve speech intelligibility. Although these results seem promising for individuals with PD, even within this group it seems treatment decisions may not be straightforward. For example, Tjaden and Wilding showed that loud speech exacted greater intelligibility gains than did the cue to speak slowly, whereas McAuliffe et al. (2014) reported the opposite; slowed cued speech resulted in a significantly higher proportion of correct listener responses compared with increased vocal loudness

(although speakers were able to increase their accuracy in the loud condition over time, thereby reducing these differences).

The second cueing strategy, reduced speech rate, can be elicited using a variety of techniques. Studies directly comparing techniques to slow rate of speech have indicated that voluntary rate reduction (e.g., speaking slowly on demand) may benefit more speakers than rigid rate control techniques (Van Nuffelen et al., 2009, 2010). However, it is unclear whether this finding is due to differences in the manner in which speakers pause and extend the duration of syllables or to the overall extent to which speakers change their speech rate. Across rate control techniques, studies have revealed significant variation in intelligibility outcomes. For example, Turner et al. (1995) examined the impact of reduced speech rate, elicited through magnitude production, in nine speakers with amyotrophic lateral sclerosis and dysarthria. Improved intelligibility was found in only four participants. In contrast, across paced reading conditions, Yorkston et al. (1990) found consistent intelligibility improvements among participants with hypokinetic ($n = 4$) and ataxic ($n = 4$) dysarthria. Hammen et al. (1994) also found consistent improvement in five speakers with hypokinetic dysarthria, although the degree of improvement differed considerably between the two studies. When speakers with hypokinetic dysarthria were prompted to reduce their speech to 60% of their habitual rate, the average increase in intelligibility was 26% in the study by Yorkston et al. but only 9% in the study by Hammen et al. Other studies have also produced conflicting findings for speakers with hypokinetic dysarthria. For example, McAuliffe et al. (2014) reported that slowed speech (elicited with magnitude production) resulted in significant improvements to intelligibility in five speakers with hypokinetic dysarthria. However, Tjaden and Wilding (2004), using the same elicitation technique, found that slow speech did not significantly enhance intelligibility in their group of 12 speakers with PD or 15 speakers with multiple sclerosis. Overall, little consistent evidence has emerged regarding the effects of reduced speech rate on intelligibility outcomes across individuals with different subtypes of dysarthria.

Differences in intelligibility gains between studies should be interpreted with some caution. Intelligibility can be estimated in many ways, across both predictable and unpredictable speech stimuli. Rating scales may have a slight advantage in their ability to measure subtle speech changes because they do not exhibit the same ceiling effects observed in transcription-based measures (Sussman & Tjaden, 2012). However, generally speaking, listener rating scales and subjective estimates of intelligibility tend to rank speakers in the same order as objective measures of the percentage of words correct (Yorkston & Beukelman, 1978). Hence, although we cannot directly compare intelligibility gains that have been established using different measurement techniques, we can assume that a speaker who exhibited intelligibility gains in one study would not exhibit reduced intelligibility in another study. For this

reason, the lack of agreement regarding which strategies are of most benefit to speakers is a concern.

The Relationship Between Speech Features and Treatment Outcomes

On the basis of the evidence presented, no "one size fits all" strategy exists for improving intelligibility for individuals with dysarthria. At present, neurologic etiology and the Mayo classification system are generally used to group participants with dysarthria in treatment studies (e.g., Cannito et al., 2012; Lowit et al., 2010). Researchers have long presumed that similar patterns of muscle or movement disorders will contribute to similar speech symptoms, making it easier to generalize speech treatments to a group with "the same" type of neurological impairment (Duffy, 2013). However, the Mayo classification system does not imply or assume homogeneity of speech characteristics within a given neurological etiology or dysarthria subtype. Rather, speakers within a single dysarthria subtype can present quite differently across levels of severity, and many aberrant perceptual features are common across multiple dysarthria subtypes (e.g., imprecise consonants or slow rate of speech; Duffy, 2013; Kim, Kent, & Weismer, 2011; Weismer, 2006). The purpose and power of the Mayo classification system is differential diagnosis and lesion localization afforded by specific clusters of perceptual symptoms that overlap minimally with other diagnoses; however, this system may not be optimal for grouping participants in treatment studies.

Although a common hypothesis is that certain treatment strategies are more appropriate for some dysarthria subtypes than others (Duffy, 2013, pp. 421–423), the research evidence underpinning this hypothesis is limited. Fox, Morrison, Ramig, and Sapir (2002), in reviewing the literature on the Lee Silverman voice treatment, suggested that loudness training was most beneficial for speakers exhibiting reduced respiratory and laryngeal drive, with speech that presents as quiet and breathy and has less variation in pitch and amplitude. However, to our knowledge, no studies have been conducted to examine the relationship between the presence of these speech features and intelligibility gains following the use of loudness as a cue. In individuals with spastic dysarthria, increasing loudness is not traditionally recommended to improve intelligibility because this increase is thought to exacerbate hyperadduction of the vocal folds (Duffy, 2013, pp. 421–423). Yet, case study evidence indicates that strategies focused on increasing loudness can increase the intelligibility of speakers diagnosed with dysarthrias containing some spastic components (e.g., D'Innocenzo, Tjaden, & Greenman, 2006). Furthermore, although a reduced speech rate is commonly recommended across dysarthria subtypes to improve intelligibility, Van Nuffelen et al. (2010), in their examination of rate control techniques across six dysarthria subtypes (including unspecified mixed dysarthrias), found that many participants experienced no change in intelligibility in response to a slow speech cue. There was no indication that speakers of any

one subtype were more likely to increase their intelligibility than were speakers of another type.

If knowing a person's dysarthria subtype does not offer reliable information about the presence and severity of various disordered speech features, we are limited in what we can infer from studies that group together speakers in this manner. The result of this approach is limited evidence of which diagnostic speech features are most important for selecting a particular treatment approach (e.g., does the presence of a certain speech feature mean that a loud speech strategy will facilitate improvements to intelligibility?). To determine the appropriateness of a treatment technique for any given individual, clinicians should utilize more information about the unique combinations of features that occur in speech output when making treatment decisions. By approaching treatment decisions in this way, clinicians disregard any assumptions about the speech signal that are based purely on neurologic etiology.

The severity of an individual speaker's dysarthria is also likely to affect their response to cueing strategies (e.g., Hammen et al., 1994; Pilon et al., 1998). For example, speakers with more severe dysarthria may have more "potential" to increase their intelligibility. That is, speakers who exhibit severely reduced intelligibility in their baseline speech will not experience a ceiling effect in the treatment gains that can be made (Hammen et al., 1994). Pilon et al. posited that slow cued speech might negatively impact a speaker's "naturalness" and, for this reason, may be less effective for speakers who already have high intelligibility. However, evidence supporting improved treatment outcomes for speakers with more severe dysarthria has been inconsistent (Van Nuffelen et al., 2009). The effect of increased intelligibility gains as dysarthria severity increases probably also depends on the treatment strategy being tested (Hunter, Pring, & Martin, 1991). Thus, although we suspect that baseline severity will affect speakers' responses to treatment, it is unclear in exactly what direction these effects might occur.

The Current Investigation

Overall, we do not have clear evidence demonstrating which baseline characteristics are indicative of speakers who will benefit from a particular form of speech cue and which symptoms may act as contraindications for this treatment. Current evidence suggests the need to look beyond broad subtype categories to fully understand a person's response to treatment. One hypothesis is that to identify the types of speakers who will achieve success with certain behavioral strategies (in addition to the types of speakers who will not), a deeper understanding of these participants' baseline speech features is needed. These features should be measurable across participants with dysarthria, so that researchers can make predictions about whether a treatment strategy is appropriate for any speaker, regardless of their dysarthria etiology and subtype.

The current study was conducted to explore whether acoustic and perceptual features of participants' baseline speech can be used to predict speaker responses to treatment techniques. Baseline speech refers to participants' habitual speech patterns when they are not using any compensatory strategies. The three aims of this investigation were (a) to compare how cues to speak loudly and reduce speech rate changed speakers' intelligibility, (b) to determine whether features of speakers' baseline speech could account for the variation observed in intelligibility gains, and (c) to model, in speakers whose intelligibility improved, which of the two treatment strategies was most appropriate for each individual.

In the current study, targeted assessments of baseline speech were selected because they were reported to be useful in predicting intelligibility deficits and distinguishing among different types of dysarthria. In our associated companion article (Fletcher, Wisler, et al., 2017), we investigate whether automated acoustic assessments can also be used to accurately model speakers' intelligibility gains.

Method

Speakers

Fifty speakers of New Zealand English (NZE; 35 men and 15 women) ages 43 to 89 years contributed speech recordings. Of these speakers, 43 were diagnosed with dysarthria ranging from mild to severe (see Table 1 for biographical information). Dysarthria severity was determined via consensus judgments of the first three authors. The remaining seven speakers, who reported no history neurological impairment, acted as controls. The group diagnosed with dysarthria had a mean age of 65 years, and the control group had a mean age of 70 years.

Speech Stimuli

Speakers attended a single recording session. Recordings took place in a quiet room, with an investigator present. Digital audio recordings were made via an Audix HT2 headset condenser microphone positioned approximately 5 cm from the mouth with a sampling rate of 48 kHz and 16 bits of quantization.

The "Grandfather Passage" was used to elicit a sample of participants' baseline speech and samples simulating two common treatment strategies. For the baseline condition, speakers were asked to read the passage in their everyday speaking voice after they had familiarized themselves with passage. Two participants with dysarthria required assistance reading the passage. In these instances, the first author would read full sentences from the passage, with the speaker repeating the sentences immediately afterwards. To create the treatment simulations, a magnitude scaling procedure was used to elicit louder and slower speech. For the slow condition, speakers were asked to say each phrase at "what feels like half your normal speed." For the loud condition, speakers were asked to read each phrase "at a level

that feels like twice as loud as normal" (Tjaden & Wilding, 2004). For approximately half of the participants (*n* = 20; some of whom were included in previous studies in our lab; see McAuliffe et al., 2014), stimuli were produced in the habitual condition first, followed by the slow condition and then the loud condition. The intention was to reduce the carryover of loud speech into the slow condition. However, this carryover effect was not observed in readings of longer passages; hence, for the remaining 30 participants the order that the loud and slow cues were presented was randomized (although the baseline habitual condition was always produced first). Because order in which speakers produced loud and slow speech was not balanced, we investigated whether the order of stimuli production had any effect on our outcome measures. We found no significant relationship between the order in which speakers produced loud and slow speech and any of the outcomes examined in this study.

Procedure

This study was conducted in two steps. Step 1 was determination of intelligibility gains—a perception experiment in which listeners rated the intelligibility of speakers' baseline, loud, and slow speech recordings. Step 2 was an analysis of participants' baseline speech features.

Step 1: Determining Intelligibility Gains

Intelligibility gains in response to loud and slow speaking cues were measured using a listener rating task. Intelligibility gains for the seven control speakers were examined alongside those for the 43 speakers with dysarthria to include a large range of possible speech features and severity of dysarthria. Initially, 25 listeners made judgments of intelligibility along a visual analog scale. However, only 18 listeners were selected to be included in the final measurement of intelligibility gain on the basis of satisfactory intrarater agreement across repeated trials. All listeners were NZE speakers between the ages of 18 and 35 years, and all passed a hearing screening before beginning the experiment. None of the listeners were familiar with dysarthric speech. In each trial, the listeners were presented with three identical phrases extracted from the "Grandfather Passage" that were produced in the baseline, loud, and slow conditions by one of the study's speakers. The listeners heard different sets of phrases for different speakers, but all speech stimuli were 11–14 syllables in length, and the first and last sentences of the reading passage were not included. The length of these stimuli was consistent with previous work involving perceptual ratings of a large number of speakers (e.g., Fletcher, McAuliffe, Lansford, & Liss, 2017; Lansford, Liss, & Norton, 2014).

To determine which speech stimuli were played, each speaker was randomly assigned to one phrase, and the same phrase was presented in the habitual, loud, and slow conditions for that speaker. When interruptions in the reading

Table 1. Demographic information for speakers with dysarthria.

| Sex	Age (years)	Medical etiology	Severity of disorder
F	46	Brain tumor	Mild–moderate
M	58	Brainstem stroke	Moderate
M	56	Cerebellar ataxia	Mild
F	69	Cerebral palsy	Severe
M	60	Cerebral palsy	Severe
F	68	Freidreich's ataxia	Mild
F	47	Huntington's disease	Moderate–severe
M	55	Huntington's disease	Severe
M	43	Hydrocephalus	Severe
F	53	Multiple sclerosis	Mild–moderate
F	60	Multiple sclerosis	Moderate–severe
F	79	Parkinson's disease	Mild
M	76	Parkinson's disease	Mild
M	77	Parkinson's disease	Mild
M	67	Parkinson's disease	Mild
F	83	Parkinson's disease	Mild
M	68	Parkinson's disease	Mild
M	89	Parkinson's disease	Mild
M	58	Parkinson's disease	Mild
M	73	Parkinson's disease	Mild
M	79	Parkinson's disease	Mild
M	69	Parkinson's disease	Mild
M	68	Parkinson's disease	Mild
F	70	Parkinson's disease	Mild–moderate
M	67	Parkinson's disease	Mild–moderate
M	71	Parkinson's disease	Mild–moderate
F	73	Parkinson's disease	Mild–moderate
M	65	Parkinson's disease	Mild–moderate
M	75	Parkinson's disease	Moderate
M	79	Parkinson's disease	Moderate
M	71	Parkinson's disease	Moderate
M	69	Parkinson's disease	Moderate
M	81	Parkinson's disease	Moderate–severe
M	77	Parkinson's disease	Moderate–severe
F	67	Progressive supranuclear palsy	Mild
M	64	Spinocerebellar ataxia	Severe
M	72	Stroke	Severe
F	48	Traumatic brain injury	Mild–moderate
M	55	Traumatic brain injury	Mild–moderate
M	60	Traumatic brain injury	Moderate
M	47	Traumatic brain injury	Severe
M	53	Undetermined neurological disease	Moderate
F	45	Wilson's disease	Mild

Note. F = female; M = male.

of the phrase occurred that were unrelated to dysarthria (e.g., speakers coughing, laughing, or misreading words in one of the three speech conditions), the speaker was reassigned to another phrase. This phrase assignment process ensured that repetitions of phrases in the baseline, loud, and slow conditions could be fairly compared.

In each trial, the speech stimuli from the habitual, loud, and slow conditions were represented on-screen graphically with an icon (e.g., a loud phrase was represented by a triangle and a slow phrase was represented by a circle). The icons were randomly assigned in each trial, so no pattern in the order in which stimuli were presented to listeners was discernable. Recordings were all scaled to the same average dB SPL and were presented at the same volume throughout the experiment. Intelligibility was defined as the ease with which speech could be understood,

consistent with previous work (Fletcher, Wisler, et al., 2017; Tjaden & Wilding, 2004). Listeners were prompted to "rate how easy you find the phrase to understand" and then place the specific phrase's icon at a point along the scale (i.e., when the slow condition was easier to understand than the baseline condition, the slow icon would be deposited further along the scale). An example of the visual presentation is shown in Figure 1. Listeners were able to replay the stimuli as often as they wished before making a decision. Each listener rated two identical blocks of habitual, loud, and slow phrases for every speaker, completing a total of 100 trials. Trials were presented in a random order for each listener.

A rating scale was used to measure intelligibility gains for several reasons. Acoustic analysis required a set of identical speech samples across speakers, which made orthographic transcription inappropriate because repeating

Figure 1. Example of a trial in the perceptual experiment to determine intelligibility gain. The screen shows three icons that were randomly assigned to play a matching phrase from a participant's habitual, loud, and slow speaking condition. Copies of these icons were placed along a visual analog scale to indicate how difficult the three phrases were to understand.
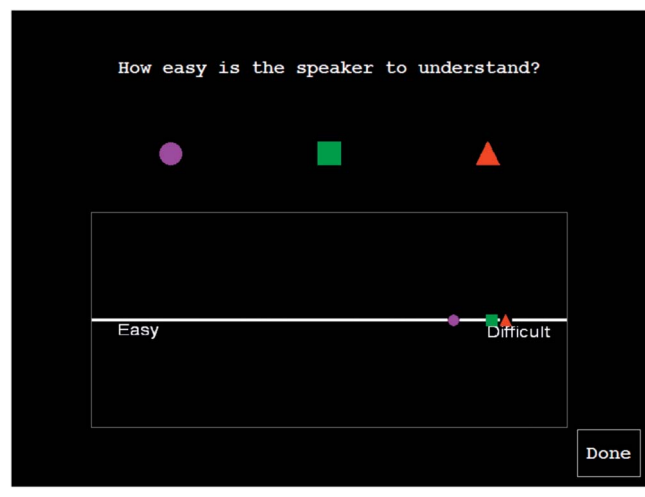
the same speech stimuli in a transcription task would result in perceptual learning and improved transcription across the experiment. Visual analog scales tend to be sensitive to subtle speech changes associated with dysarthria, and ratings of intelligibility generally agree closely with orthographic measures (Sussman & Tjaden, 2012; Yorkston & Beukelman, 1978). For this reason, we determined that visual analog scaling would allow listeners to efficiently index small intraspeaker changes in speech production.

To measure intelligibility gain, the average distance that the loud and slow icons were placed from the baseline speech icon was calculated for each listener. This average was used to compute a z score for the amount of change (in either the negative or positive direction) that occurred as a result of the two treatment conditions. For example, the z score for Listener 1's rating of Speaker 1 in the loud condition would be derived in the following manner. The absolute value of the difference between the listener's rating of Speaker 1's baseline speech and loud cued speech would be calculated. Then, the average difference in ratings that this listener gave to all speakers' baseline speech and their loud and slow cued speech would be subtracted from this value. The resultant number would then be divided by the standard deviation of the difference in ratings that this listener gave to all speakers' baseline speech and cued speech. This process was used for each rating reported by the listener participants and created a z score for each trial.

By definition, the z score was lowest when the listener gave the baseline and cued speech conditions the same rating. Therefore, when we extracted the difference between the lowest possible z score and all other z scores reported by a listener, we produced a series of positive

scores (when the baseline and treatment conditions were rated differently) and scores of 0 (when the conditions were rated the same). The final ratings given to each speaker were then made positive or negative depending on the direction of the change. After completing these procedures, ratings from the 18 most internally reliable listeners were averaged to give one group average rating per speaker.

Comparing speech samples within the same trial to determine whether a change in intelligibility and/or speech quality has occurred is common practice in dysarthria treatments (see Park, Theodoros, Finch, & Cardell, 2016; Sapir et al., 2002, 2003; Wenke et al., 2011). In the present study, we did not attempt to quantify the precise degree of intelligibility change caused by each treatment simulation. Instead, we were primarily interested in determining whether cues to speak loudly and slowly changed the intelligibility of different speakers. Hence, the purpose of comparing samples within the same speaker was to provide a measurement that could be used across people who exhibited large differences in baseline intelligibility and would be sensitive to the relatively small changes produced following cues to speak loudly and slowly.

On average, listener z score ratings ranged from -4.9 to 5.2 across the speakers they were asked to rate, producing a scale ranging from 10 to 11 points. Although listeners usually agreed about which sample they preferred, the distance between where they placed the baseline and treatment icons was highly variable. For loud cued speech, listeners' standard deviation in z scores was 0.50–1.68 for different speakers. In the slow condition, variability in listeners' ratings was even higher, with standard deviations in z scores of 0.61–2.03 for different speakers. Thus, for certain speakers, listeners were not in close agreement about the amount of improvement the treatment strategy produced.

To best represent the general opinion of the listener population, we averaged z scores for the 18 most internally reliable listeners. The models developed were built using these average values. Issues regarding the reliability of listeners' ratings of "ease of understanding" are further addressed in the Discussion section.

Reliability

Because the experiment required sustained focus by the listener, checks were included to ensure that listeners were consistent in their ratings throughout the experiment. Trials from each speaker were presented twice for this purpose, at random stages throughout the experiment. Although the order of all trials was randomized for each listener, trials from the same speaker were never allowed to occur in immediate succession. Only listeners who placed the baseline and treatment icons in the same order in over 65% of the repeated trials were included in the final results. After removing the less reliable listeners, 18 remained. On average, the 18 listeners were consistent in the order in which they placed the icons in 77% of the repeated trials. Interrater reliability was assessed in a similar way by examining the proportion of listeners who placed

the baseline and treatment icons in the same order within trials (trials in which listeners gave two samples identical ratings were excluded). The proportion of listeners who agreed on the "most intelligible" sample differed considerably across speakers. For example, when a speaker did not produce noticeable changes between the treatment conditions, listeners were often inconsistent in their preferences. For this reason, listeners' agreement across trials ranged from 50% (indicating no consensus) to 100% (indicating total agreement about which sample was most intelligible). On average, 76% of listeners agreed in their preferences for the treatment versus baseline speech in each trial.

Step 2: Analyses of Participants' Baseline Speech Features

Perception Task: Severity of Dysarthria

A second perceptual task was undertaken, in this case to obtain a speech severity rating value for each speaker (both those with dysarthria and healthy controls) in the baseline condition. These data were originally reported by Fletcher, Wisler, et al. (2017). In this task, a separate group of 14 listeners rated speakers' baseline speech precision along a visual analog scale. All listeners were speakers of NZE (ages 18–31 years), passed a hearing screening, and did not have training in the assessment of dysarthria. The listeners were exposed to one identical phrase from each speaker and were prompted to rate "the speaker's speech precision." Listeners rated one speaker at a time, with each trial repeated once for reliability purposes. As with the previous perceptual experiment, the order of trials was randomized for each listener so repeated trials were spread throughout the experiment. However, two trials from the same speaker were never allowed to occur in immediate succession. The raw ratings were converted to z scores for each listener before being averaged across the listener group. Further details of this rating protocol were previously published by Fletcher, Wisler, et al. (2017). Using this protocol, the average intrarater correlation (across ratings of the same speech samples) was $r(852) = .90$, $p < .001$. Intraclass correlation coefficients (2, 1) were .835, indicating high interrater reliability.

Acoustic Analysis

To complete the acoustic analysis, each speaker's baseline recording of the "Grandfather Passage" was transcribed and then automatically segmented at the phoneme level using the hidden Markov model toolkit (Young et al., 2002). All automatically derived phoneme boundaries were then visually checked for accuracy by a team of trained analyzers using standard criteria (Peterson & Lehiste, 1960). When uncertainty arose in discriminating boundaries for consecutive phonemes (e.g., /t/ and /s/) the boundary selected through the automatic segmentation was retained. For further information about this process see Fletcher, McAuliffe, Lansford, and Liss (2015).

After manual checking, seven acoustic metrics were extracted from across each speaker's baseline speech recordings. The first five were used to index articulation

and speech prosody; we measured articulation rates, the pairwise variability index for vocalic intervals (vPVI), formant centralization ratios, and the standard deviations of speakers' fundamental frequency and intensity from across the "Grandfather Passage." These measurements are similar to those examined by Kim et al. (2011) and were selected because they were reported to be useful either for predicting intelligibility deficits or for distinguishing among different types of dysarthria. For example, certain dysarthria subtypes are thought to have quantifiable differences in their articulation rates and vPVI scores (Liss et al., 2009), whereas variations in speakers' fundamental frequency and intensity are thought to differ based on their dysarthria severity (Bunton, Kent, Kent, & Rosenbek, 2000; Metter & Hanson, 1986; Schlenck, Bettrich, & Willmes, 1993). However, Lowit (2014) provided evidence of variability in the vowel vPVI measures among speakers with hypokinetic dysarthria.

In addition to these five acoustic metrics, two indices of voice quality were also included: smoothed cepstral peak prominence (Hillenbrand & Houde, 1996) and the amplitude of the first harmonic. These measurements were included because aspects of voice quality are thought to differ considerably among speakers with dysarthria (Darley, Aronson, & Brown, 1975), and preliminary research indicated that these measurements may be useful for indexing differences in listeners' perceptions of breathiness and strain (Cannito, Buder, & Chorna, 2005; Hillenbrand & Houde, 1996).

To summarize, these seven acoustic measures were chosen to gain an objective account of differences in the speech signal across speakers. Further details of the measurements are included in Table 2.

Results

Effect of Loud and Slow Cued Speech on Intelligibility

Speakers' intelligibility changes in response to different treatment simulations are summarized in Figure 2. Visual analysis indicated considerable variation in the amount of intelligibility gain speakers achieved in the loud and slow conditions. However, speakers who had large intelligibility gains in the one condition were more likely to have intelligibility gains in the other (i.e., intelligibility gains in the two conditions were positively correlated across speakers), $r(48) = .55$, $p < .001$. Overall, 35 speakers showed some degree of improvement using one or more of the treatment strategies. The baseline speech of the remaining 15 speakers was rated as more intelligible than their speech in either treatment condition. Among the speakers that did not benefit from the treatment cues, healthy control speakers were disproportionally represented. Five of seven healthy controls had baseline speech samples that were rated as the most intelligible. In contrast, only 10 of the 43 speakers with dysarthria were given the highest rating for the baseline condition.

Table 2. Description of acoustic measures.

Measure	Description
Articulation rate	Number of syllables per second. Pauses > 50 ms were excluded from the calculation (see Robb, Maclagan, & Chen, 2004).
vPVI	Normalized pairwise variability index for vocalic intervals. The mean of the differences between successive vocalic intervals divided by their sum (×100) (see Liss et al., 2009).
Pitch variation	Standard deviation of fundamental frequency in Hz.
Intensity variation	Standard deviation of intensity in dB.
FCR	Formant centralization ratio, adapted for New Zealand vowels. Measured as a ratio of formant frequencies in Bark; (F2[oː] + F2[eː] + F1[iː] + F1 [oː]) ÷ (F2[iː] + F1[eː]) (see Fletcher, Wisler, et al., 2017).
CPPS	Smoothed cepstral peak prominence. The amplitude of the CP corresponding to the fundamental period (in dB), normalized for overall signal amplitude (see Hillenbrand & Houde, 1996). Cepstrums were measured from within vowel segments, with a Hanning window applied to the 60-ms segment at the temporal center of each vowel and shorter vowels excluded from the analysis.
H1A	First harmonic amplitude, measured in dB. The amplitude of the first harmonic minus the amplitude of the second harmonic (see Hillenbrand & Houde, 1996). As described for CPPS, harmonics were measured from within vowel segment windows.

Predicting Intelligibility Improvement

A series of linear regression models were used to analyze the effect of speakers' baseline speech features on their intelligibility improvement in the two treatment conditions. Our aim was to better characterize both the types of speakers who made large gains in response to treatment cues and the types of speakers who did not. For this reason, Models 1 and 2 included all speakers in our data set.[1] Before the models were run, speakers' acoustic and perceptual features were evaluated for sources of multicollinearity. Correlations between different speech features ranged from .06 to .64. Although many significant correlations were found, none were high enough to raise concern (see Table 3 for more detail). Because of the large set of acoustic variables, backward stepwise regression was conducted to identify a subset of speech features that were predictive of speakers' improvement in intelligibility. Model selection proceeded in a backward stepwise iterative fashion seeking to create a predictive model that contained only significant effects (with α = .05). This process resulted in the creation of two models: one to predict improvement in the loud condition and one to predict improvement in the slow condition. To provide an additional check for multicollinearity, we calculated variance inflation factors for all

---

[1]A separate set of linear regression models were calculated with the seven healthy control speakers removed from the data set. For Model 1, which predicted improvement in the slow condition, the exact same significant effects were found. The main effect for listener ratings of speech precision was only marginally weaker, β = −0.53 (0.16), *p* = .001, whereas the effect of vPVI increased slightly, β = 0.36 (0.13), *p* = .007. In Model 2, which predicted improvement in the loud condition, the effect of listener ratings of speech precision was no longer significant when included in the model (*p* = .09). Hence, the final model contained only one main effect for articulation rate, β = 0.42 (0.12), *p* = .001. Removal of listener ratings from this model resulted in a clear reduction in the effect size of articulation rate, indicating that controlling for speech severity increased the strength of the relationship between articulation rate and intelligibility gains.

statistical models that contained more than one independent variable. All factors were below 1.5.

Model 1: Intelligibility Gains in Response to Cues to Speak More Slowly

Model 1 examined the degree that speakers changed their intelligibility in the slow condition relative to their baseline speech sample. With Model 1, the level of improvement in intelligibility made in the slow condition was best predicted by both speakers' baseline speech severity and their vocalic pairwise variability index. To compare the relative effect of these features, all regression coefficients

Figure 2. Perceived improvement in loud and slow speaking conditions. The origin (zero) indicates the speaker had the same average intelligibility rating across the habitual and treatment conditions. The *x*- and *y*-axes indicate the number of standard deviations of intelligibility change that occurred from this point.
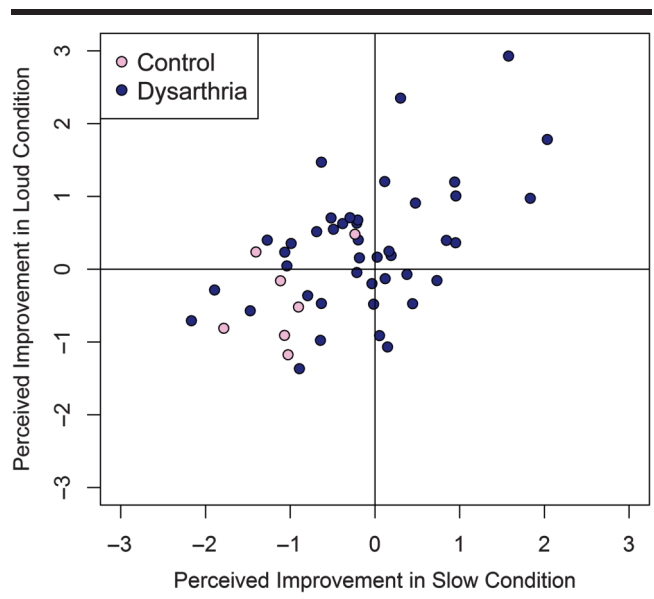
Table 3. Correlations between baseline speech measurements across speakers.

Feature	1	2	3	4	5	6	7
1. Perceptual ratings (higher rating = less severe)							
2. Formant centralization ratio	−.61**						
3. Articulation rate	.44**	−.21					
4. First harmonic amplitude	.18	−.06	.64**				
5. Smoothed cepstral peak prominence	−.23	.15	−.56**	−.47**			
6. Pitch variation	−.10	−.02	−.43**	−.40**	.19		
7. Intensity variation	−.15	.22	−.49**	−.38**	.51**	.36*	
8. Normalized vocalic pairwise variability index	.31*	−.19	.35*	.07	.08	−.22	−.06

*$p < .05$. **$p < .01$.

were standardized. The final model included a main effect for listener ratings of speech precision, $\beta = -0.59$ (0.13), $p < .001$, indicating that speakers with more severe dysarthria experienced significantly greater intelligibility improvements when cued to slow down. There was also a significant effect for measurements of vPVI, $\beta = 0.33$ (0.11), $p = .006$, which suggests that speakers with a greater degree of temporal variability in their speech segments were better able to utilize slow speech as a strategy for increasing intelligibility. However, this relationship was only apparent once baseline speech severity was held constant. Overall, Model 1 accounted for 34% of the variance in speakers' responses to treatment cues. Interactions between the variables were not significant.

Model 2: Intelligibility Gains in Response to Cues to Speak Louder

Model 2 examined the degree that speakers changed their intelligibility when cued to speak loudly. With Model 2, improvements in the loud condition were best predicted by speakers' articulation rate and their baseline speech severity. The final model had a main effect for articulation rate, $\beta = 0.52$ (0.12), $p < .001$, suggesting that speakers with a faster rate of speech experienced significantly greater improvements in intelligibility when cued to speak loudly. There was also a significant effect for ratings of speech precision, $\beta = -0.41$ (0.13), $p = .003$, indicating that speakers with more severe dysarthria had greater gains in intelligibility, although this relationship was apparent only when speakers' articulation rates were held constant. All regression coefficients were standardized, and interactions between variables were not significant. Overall, Model 2 accounted for 31% of the variance in speakers' responses to treatment cues.

Choosing Between Treatment Cues

Models 1 and 2 provided insight into speech features that can be used to determine the appropriateness of loud or slow treatment cues for any given person. However, Figure 2 shows that some speakers benefited considerably more from one speech modification as opposed to the other. Hence, the question remains: Which characteristics

of dysarthria can be used to identify speakers who will perform better with one strategy over another? To answer this question, we examined a subset of 35 speakers who experienced a positive change in intelligibility in response to either the loud or slow treatment strategy. These participants were divided into two groups: Group 1 had greater intelligibility gains in the loud condition ($n = 24$), and Group 2 had greater gains in the slow condition ($n = 11$). The participants who were included in each group are listed in Appendix A.

Intelligibility improvement was measured as an average of listeners' z scored ratings (with each unit representing 1 SD of change between the baseline and treatment conditions). Participants in Group 1 experienced an average improvement in the loud condition of 0.76 SD. Participants in Group 2 experienced an average improvement in the slow condition of 0.70 SD. The two groups were coded separately, and group membership was used as the dependent variable for a series of binomial regression models developed to determine whether there were differences in the speech features of participants who did better in one treatment condition as opposed to the other. Model selection proceeded in a backward stepwise iterative fashion seeking to create a predictive model that contained only significant effects (with $\alpha = .05$).

Model 3: Speakers' Most Successful Treatment Strategy

Model 3 examined whether speakers had greater intelligibility gains in the loud or slow treatment condition. With Model 3, speakers' better treatment strategy was best predicted by the baseline measurement of their first harmonic amplitude (H1A). The final model contained only one main effect for H1A, $\beta = -1.36$ (0.53), $p = .01$, suggesting that a speaker's baseline voice quality was the best determinant of whether one treatment condition would be more successful than another. As in previous models, the regression coefficient was standardized, and there were no significant interactions. The odds ratio revealed that for each standard deviation increase in a speaker's H1A scores, speakers were 3.9 times more likely to perform better in the loud condition than in the slow condition. In comparison to the null model, the inclusion of

H1A significantly improved the fit of the model, $\chi^2(1) = 9.13$, $p = .003$.

Discussion

In the present study, we investigated measurements of speakers' baseline speech features to determine whether there were objective speech assessment data that could be used to make predictions about individuals' responses to treatment. The study included a range of speakers with dysarthria, and each of their baseline speech features was analyzed in the same way, regardless of their underlying etiology. The three main aims were (a) to compare how cues to speak loudly and reduce speech rate changed speakers' intelligibility, (b) to determine whether features of speakers' baseline speech could account for the variation observed in their intelligibility gains, and (c) to model which of the two treatment strategies was most appropriate for each person.

Intelligibility Gains Following Cues to Speak Loudly and Slow the Rate of Speech

Across speakers, variable intelligibility gains were observed following cues to speak loudly and reduce the rate of speech. Overall, behavioral cues had a positive effect on ratings of intelligibility for most speakers with dysarthria, and these improvements were linked across speakers. When a speaker exhibited large improvement in response to one treatment, they were likely to also exhibit improvement in response to the other treatment. The variation in participants' responses to speech cues was consistent with findings from previous investigations (Hammen et al., 1994; Neel, 2009; Pilon et al., 1998; Tjaden & Wilding, 2004; Turner et al., 1995; Van Nuffelen et al., 2009, 2010; Yorkston et al., 1990). The level of variation in ratings of loud cued speech and slow cued speech also was reasonably similar (see Figure 2). However, the slow cued speech samples were more likely to be perceived as causing a reduction in listeners' ease of understanding.

This finding may, in part, reflect the manner in which intelligibility gains were measured in the perceptual experiment. In this experiment, listeners were asked to rate how easy each speaker was to understand. Thus, the measurements taken were not an objective tally of the words each listener was able to correctly transcribe. Instead, the protocol allowed the listener to express his or her own preferences and biases for certain speech samples. In previous studies, listeners had a natural preference for speech samples of a similar or slightly faster rate than their own speech (Street, Brady, & Putman, 1983). Street et al. found that when a healthy person speaks in this range, they are rated by listeners as significantly more socially attractive and competent than when speaking at a slower rate. For this reason, listeners probably were naturally predisposed to prefer faster speech samples where the objective intelligibility of the two samples was similar.

Although the majority of speakers with dysarthria were perceived to be somewhat easier to understand in at least one of the cued speech conditions, this result was not found for the healthy older speakers, which may indicate that when a speaker is relatively healthy (i.e., a control group member) and intelligible (as indicated by high ratings of baseline speech precision) listeners are less likely to view positively any behavioral changes to the speech signal. The tendency for negative ratings of control speakers in the slow condition indicate that this cueing strategy may be perceived as less natural than loud cued speech or may require more listener effort when the speech signal is relatively unimpaired.

A significant correlation was found between the intelligibility gains made by speakers in each of the treatment conditions. This result was not particularly surprising—speakers who are able to use one strategy with great success are also more likely to be able to successfully employ other speech modification strategies. However, because in most dysarthria treatment studies the effects of only one program at a time have been explored, researchers often do not consider whether a particular group would achieve similar results with a different speech therapy approach. The correlation between speakers' intelligibility gains following cues to speak loudly and reduce speech rate indicate the importance of remaining open minded in approaches to treating dysarthria. Although one strategy is successful for a particular participant, it may not be the only speech modification that will produce positive outcomes.

Predictors of Intelligibility Gain in the Slow Condition

In combination, an individual's speech precision in the baseline condition and their temporal vowel variation (as measured with vPVI) were significant predictors of intelligibility improvement when prompted to speak slowly. Specifically, participants with more severe speech imprecision and greater levels of temporal variation in their syllables were more likely to be given higher intelligibility ratings in the slow speech condition. Regarding the perceptual speech precision ratings, this result was not unexpected. Baseline severity has previously been hypothesized to affect speakers' intelligibility improvement in exactly this manner (Hammen et al., 1994; Pilon et al., 1998). Several reasons have been proposed for why speakers with more severe dysarthria might benefit more from rate control strategies. For example, intelligibility gains could exhibit a ceiling effect in speakers with highly intelligible baseline speech (Hammen et al., 1994). Thus, among speakers who improved their intelligibility, those with more severe dysarthria could have larger differences in their ratings. Slow cued speech also could negatively impact the "naturalness" of a person's speech (Pilon et al., 1998). Because listeners have a tendency to prefer speech that is of a similar rate to their own, people with highly intelligible speech who significantly slow down their natural speaking rate may be perceived as requiring more effort to understand.

When perceptual ratings of speech severity were held constant, speakers who had a larger vPVI tended to make greater intelligibility improvements. A high normalized vPVI occurs as a result of increased durational differences from one syllable to the next and is thought to be associated with speech that has more variation in stress (Liss et al., 2009). Syllabic stress helps listeners in their segmentation of the dysarthric speech signal (Liss, Spitzer, Caviness, Adler, & Edwards, 1998). Our model suggests that speakers who have high temporal differentiation of stressed and unstressed syllables may be able to employ rate control strategies more effectively. When these speakers extend the duration of their speech sounds, they may be able to produce clearer differences in the length of their vowels, which may enable listeners to better detect their stressed syllables despite the alteration to speech prosody.

Predictors of Intelligibility Gain in the Loud Condition

Intelligibility gains in the loud condition were best predicted by articulation rate and a perceptual measure of the severity of the dysarthria in baseline speech. This model (Model 2) suggests that when baseline severity is controlled for people who speak with a faster articulation rate tend to exhibit greater intelligibility gains in the loud condition. In contrast, when baseline articulatory rates are held constant, speakers with more severe dysarthria tend to make greater intelligibility gains. The severity of the dysarthria in baseline speech predicted intelligibility improvement in a similar manner for both loud and slow cued speech (although the effect was stronger for slow cued speech). The reasons for this effect of severity in the loud condition are likely to be similar to those in the slow condition. For example, cues to speak loudly may have a negative impact on perceived naturalness of the speech to some degree.

A faster articulatory rate was predictive of intelligibility gains in the loud condition but not in the slow condition. One explanation is that speakers with a faster articulatory rate may exhibit a range of related characteristics that make loud speech an appropriate treatment strategy. For example, cues to speak loudly have been theorized to specifically address breathiness by improving vocal fold adduction in speakers with dysarthria (Baumgartner et al., 2001). In the present study, measures of articulatory rate were strongly correlated with measures of acoustic voice quality (see Table 3). Specifically, H1A had a strong and significant relationship with articulatory rate. In previous studies, H1A was positively correlated with the perception of breathiness (Hillenbrand & Houde, 1996) but was hypothesized to be negatively correlated with measures of creakiness, vocal strain, and the perception of roughness (when breathiness is controlled for; Cannito et al., 2005). The relationship between H1A and articulatory rate suggests that speakers with a slow rate of speech may exhibit a more strained speech quality, whereas speakers with a normal or increased rate might be more likely to exhibit breathiness. Articulatory rate may be more sensitive than our measurements of cepstral peak prominence and H1A to differences between breathy versus strained and effortful speech. Hence, positive relationship between speakers' articulatory rate and improvement in the loud condition may be related to a number of covarying factors.

Comparing Treatment Cues in the Same Speakers

The final issue investigated was whether characteristics of dysarthria could be used to identify speakers who performed better with one strategy than another. Measures of H1A were a significant predictor of whether speakers would demonstrate greater success with cues to speak loudly as opposed to cues to speak slowly. No further variables that accounted for significant variation in this model (Model 3) were identified. Measures of H1A may be differently affected in speakers with a breathy voice quality and speakers who are perceived to sound tense or strained (Cannito et al., 2005; Hillenbrand & Houde, 1996). In this model, participants who had breathier speech were more likely to be rated higher in the loud than in the slow speech condition.

This finding is not entirely surprising. Breathiness occurs when adduction of the vocal folds is reduced, and previous studies have reported improved vocal fold adduction following treatment that focuses on speaking loudly (Smith, Ramig, Dromey, Perez, & Samandari, 1995). In contrast, data obtained in the present study for the loud condition suggest that loud speech may be contraindicated in speakers who have a very slow speech and who perhaps also exhibit an effortful or strained voice quality. This effect may account for the significance of the H1A measure for distinguishing the speakers who are more likely to be successful with cues to speak more slowly. This model was developed to predict a broad, binary outcome and—in contrast to Models 1 and 2—had less speaker data available, which may account to some degree for its relative simplicity.

Limitations

The limited sample of speakers with dysarthria included in this study was the largest limitation to model development and the generalizability of the models' findings. For example, the group of speakers included is unlikely to represent the full range of speech characteristics and dysarthria severity present in the larger population. Therefore, the models generated probably cannot account for all combinations of speech features found in speakers with dysarthria. An overrepresentation of certain dysarthria etiologies (e.g., those with PD) may have skewed the patterns of speech features observed in this study—and therefore may have influenced the size of the effects reported in the models. Speakers with the same etiology do not necessarily share more similar acoustic speech features (Kim et al., 2011). However, certain cardinal features of dysarthria have been commonly documented in this group,

including breathiness, reduced loudness, and a faster rate of speech (Duffy, 2013). In the loud condition, articulatory rate was a significant predictor of intelligibility gains. H1A was a significant predictor of speakers' most successful cueing strategy. However, the speakers with PD may have affected the strength of these effects. For example, speakers with PD and unusually fast rates or atypically breathy voices may have had more influence on the results because they represented a relatively large proportion of the sample. This study was not designed to assess relationships between neurological etiologies and treatment outcomes. However, the most successful cueing strategy for each speaker is listed in Appendix A, and many speakers with PD produced their highest rated sample in the loud condition.

This study did not include an attempt to measure to what degree participants accurately followed the cues to speak loudly or reduce their speech rate. For example, some speakers made considerably more effort to speak loudly than did others. In addition, some participants may have been unable to produce noticeably louder speech or to maintain this speech pattern throughout the reading passage. For slow cued speech, there was no control for the reduction in speaking or the manner in which the speakers reduced their rate. For example, some speakers were more inclined to insert pauses between words and others extended the duration of each syllable. The manner and degree to which each speaker enacted production changes very likely influenced the individual's resultant intelligibility gain (an issue explored further by McAuliffe et al., 2017). However, the intention of the present study was to establish whether it was possible to predict speakers' intelligibility gains before they completed the two treatment simulations. For this reason, information about the manner in which they enacted loud and slow speech—including the extent to which they changed their speech patterns—was not controlled for in any way.

This study was also limited in its characterization of the full range of listener responses to treatment strategies. The reliability of listeners' preferences for a particular strategy was not particularly high. For a couple of participants, listeners completely disagreed about whether a particular strategy raised the speaker's intelligibility. When speech samples sounded very similar, the average z score outcome was close to zero, indicating little positive or negative change. However, this outcome also occurred when listeners were rating the two samples differently but were unable to agree on whether the treatment strategy was beneficial; these two results are difficult to disentangle. Generally speaking, average listener ratings provided useful insight into the types of speakers who benefited most from particular strategies. However, the exact amount that different speakers improved in the loud and slow conditions should be interpreted with some caution because improvement is likely to be perceived differently across listener groups. In future studies, a larger range of outcome measures will be compared to investigate the relationships between listener ratings, transcriptions, and a

speaker's ability to maintain speech modifications over a longer period of time.

Summary

In this study, features of a speaker's baseline speech —including information about the severity of dysarthria and acoustic measures of the dysarthric speech signal— were useful for predicting the speaker's level of success in response to different treatment strategies. As expected, features of the dysarthric speech signal were not able to account for all the variation observed in speakers' intelligibility gains. Factors related to the participants' cognitive abilities, motivation, and fatigue probably significantly affected their responses to the speech modification strategies. However, the ability to account for around one-third of the variance in listeners' perceptions of their intelligibility gains has considerable clinical importance. These preliminary data indicate that new assessment methods could be used to select and group participants for future treatment studies. The ability to specifically target the types of speakers who are likely to make large intelligibility gains has the potential to promote much stronger group outcomes in these studies.

Data from this study also provide the beginnings of an evidence base for clinical decision making that can account for a wider variety of presenting dysarthrias. The assessment protocol used in this investigation can be applied to any speaker, regardless of the underlying etiology. If more speakers had been added to this analysis, the models probably would have been able to incorporate even more baseline speech variables and hence account for increasingly individualized presentations of dysarthric speech. These models could provide a pathway to more individually targeted and adaptive approaches to speech modification and motor learning in dysarthria therapy.

Acknowledgments

References

Baumgartner, C. A., Sapir, S., & Ramig, L. O. (2001). Voice quality changes following phonatory-respiratory effort treatment (LSVT®) versus respiratory effort treatment for individuals with Parkinson disease. *Journal of Voice, 15,* 105–114.

Bunton, K., Kent, R. D., Kent, J. F., & Rosenbek, J. C. (2000). Perceptuo-acoustic assessment of prosodic impairment in dysarthria. *Clinical Linguistics & Phonetics, 14,* 13–24.

Cannito, M. P., Buder, E. H., & Chorna, L. B. (2005). Spectral amplitude measures of adductor spasmodic dysphonic speech. *Journal of Voice, 19,* 391–410.

Cannito, M. P., Suiter, D. M., Beverly, D., Chorna, L., Wolf, T., & Pfeiffer, R. M. (2012). Sentence intelligibility before and after voice treatment in speakers with idiopathic Parkinson's disease. *Journal of Voice, 26,* 214–219.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor speech disorders* (Vol. 304). Philadelphia, PA: Saunders.

D'Innocenzo, J., Tjaden, K., & Greenman, G. (2006). Intelligibility in dysarthria: Effects of listener familiarity and speaking condition. *Clinical Linguistics & Phonetics, 20,* 659–675.

Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management.* New York, NY: Elsevier.

Fletcher, A., & McAuliffe, M. (2017). Examining variation in treatment outcomes among speakers with dysarthria. *Seminars in Speech and Language, 38*(03), 191–199.

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,* 2132–2139.

Fletcher, A. R., McAuliffe, M. J., Lansford, K. L., & Liss, J. M. (2017). Assessing vowel centralization in dysarthria: A comparison of methods. *Journal of Speech, Language, and Hearing Research, 60*(2), 341–354.

Fletcher, A. R., Wisler, A. A., McAuliffe, M. J., Lansford, K. L., & Liss, J. M. (2017). Predicting intelligibility gains in dysarthria through automated speech feature analysis. *Journal of Speech, Language, and Hearing Research, 60,* 3058–3068. https://doi.org/10.1044/2017_JSLHR-S-16-0453

Fox, C. M., & Boliek, C. A. (2012). Intensive voice treatment (LSVT LOUD) for children with spastic cerebral palsy and dysarthria. *Journal of Speech, Language, and Hearing Research, 55,* 930–945.

Fox, C. M., Morrison, C. E., Ramig, L. O., & Sapir, S. (2002). Current perspectives on the Lee Silverman Voice Treatment (LSVT) for individuals with idiopathic Parkinson disease. *American Journal of Speech-Language Pathology, 11,* 111–123.

Hammen, V. L., Yorkston, K. M., & Minifie, F. D. (1994). Effects of temporal alterations on speech intelligibility in parkinsonian dysarthria. *Journal of Speech, Language, and Hearing Research, 37,* 244–253.

Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research, 39,* 311–321.

Hunter, L., Pring, T., & Martin, S. (1991). The use of strategies to increase speech intelligibility in cerebral palsy: An experimental evaluation. *International Journal of Language & Communication Disorders, 26,* 163–174.

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.

Lansford, K. L., Liss, J. M., & Norton, R. E. (2014). Free-classification of perceptually similar speakers with dysarthria. *Journal of Speech, Language, and Hearing Research, 57,* 2051–2064.

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *The Journal of the Acoustical Society of America, 104,* 2457–2466.

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., & Caviness, J. N. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of Speech, Language, and Hearing Research, 52,* 1334–1352.

Lowit, A. (2014). Quantification of rhythm problems in disordered speech: A re-evaluation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 369*(1658), 20130404.

Lowit, A., Dobinson, C., Timmins, C., Howell, P., & Kröger, B. (2010). The effectiveness of traditional methods and altered auditory feedback in improving speech rate and intelligibility in speakers with Parkinson's disease. *International Journal of Speech-Language Pathology, 12,* 426–436.

Mackenzie, C., & Lowit, A. (2007). Behavioural intervention effects in dysarthria following stroke: Communication effectiveness, intelligibility and dysarthria impact. *International Journal of Language & Communication Disorders, 42,* 131–153.

Mahler, L. A., & Ramig, L. O. (2012). Intensive treatment of dysarthria secondary to stroke. *Clinical Linguistics & Phonetics, 26,* 681–694.

McAuliffe, M. J., Baylor, C., & Yorkston, K. M. (2016). 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.* https://doi.org/10.1080/17549507.2016.1193900

McAuliffe, M. J., Fletcher, A. R., Kerr, S. E., O'Beirne, G. A., & Anderson, T. (2017). Effect of dysarthria type, speaking condition, and listener age on speech intelligibility. *American Journal of Speech-Language Pathology, 26*(1), 113–123.

McAuliffe, M. J., Kerr, S. E., Gibson, E. M., Anderson, T., & LaShell, P. J. (2014). Cognitive–perceptual examination of remediation approaches to hypokinetic dysarthria. *Journal of Speech, Language, and Hearing Research, 57,* 1268–1283.

Metter, E. J., & Hanson, W. R. (1986). Clinical and acoustical variability in hypokinetic dysarthria. *Journal of Communication Disorders, 19,* 347–366.

Neel, A. T. (2009). Effects of loud and amplified speech on sentence and word intelligibility in Parkinson disease. *Journal of Speech, Language, and Hearing Research, 52,* 1021–1033.

Park, S., Theodoros, D., Finch, E., & Cardell, E. (2016). Be clear: A new intensive speech treatment for adults with nonprogressive dysarthria. *American Journal of Speech-Language Pathology, 25,* 97–110.

Patel, R. (2002). Prosodic control in severe dysarthria: Preserved ability to mark the question–statement contrast. *Journal of Speech, Language, and Hearing Research, 45,* 858–870.

Patel, R., & Campellone, P. (2009). Acoustic and perceptual cues to contrastive stress in dysarthria. *Journal of Speech, Language, and Hearing Research, 52,* 206–222.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32,* 693–703.

Pilon, M. A., McIntosh, K. W., & Thaut, M. H. (1998). Auditory vs visual speech timing cues as external rate control to enhance verbal intelligibility in mixed spastic ataxic dysarthric speakers: A pilot study. *Brain Injury, 12,* 793–803.

Robb, M. P., Maclagan, M. A., & Chen, Y. (2004). Speaking rates of American and New Zealand varieties of English. *Clinical Linguistics & Phonetics, 18,* 1–15.

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*(6), 296–303.

Sapir, S., Spielman, J., Ramig, L. O., Hinds, S. L., Countryman, S., Fox, C., & Story, B. (2003). Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on ataxic dysarthria: A case study. *American Journal of Speech-Language Pathology, 12,* 387–399.

Schlenck, K.-J., Bettrich, R., & Willmes, K. (1993). Aspects of disturbed prosody in dysarthria. *Clinical Linguistics & Phonetics, 7,* 119–128.

Smith, M. E., Ramig, L. O., Dromey, C., Perez, K. S., & Samandari, R. (1995). Intensive voice treatment in Parkinson disease: Laryngostroboscopic findings. *Journal of Voice, 9,* 453–459.

Street, R. L., Brady, R. M., & Putman, W. B. (1983). The influence of speech rate stereotypes and rate similarity or listeners'

evaluations of speakers. *Journal of Language and Social Psychology, 2,* 37–56.

Sussman, J. E., & Tjaden, K. (2012). Perceptual measures of speech from individuals with Parkinson's disease and multiple sclerosis: Intelligibility and beyond. *Journal of Speech, Language, and Hearing Research, 55,* 1208–1219.

Tjaden, K., & Wilding, G. E. (2004). Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research, 47,* 766–783.

Turner, G. S., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research, 38,* 1001–1013.

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.

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.

Weismer, G. (2006). Philosophy of research in motor speech disorders. *Clinical Linguistics & Phonetics, 20,* 315–349.

Wenke, R. J., Theodoros, D., & Cornwell, P. (2011). A comparison of the effects of Lee Silverman voice treatment and traditional therapy on intelligibility, perceptual speech features, and everyday communication in nonprogressive dysarthria. *Journal of Medical Speech-Language Pathology, 19*(4), 1–24.

Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders, 11,* 499–512.

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.

Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, J., Ollason, D., . . . Woodland, P. (2002). *The HTK-Book 3.2.* Cambridge, England: Cambridge University.

Appendix A

Highest Rated Condition by Speaker

Sex	Age (years)	Medical etiology	Severity of disorder	Highest rated condition
F	46	Brain tumor	Mild–moderate	Slow
M	58	Brainstem stroke	Moderate	Slow
M	56	Cerebellar ataxia	Mild	Slow
F	69	Cerebral palsy	Severe	Habitual
M	60	Cerebral palsy	Severe	Habitual
F	68	Freidreich's ataxia	Mild	Slow
F	47	Huntington's disease	Moderate–severe	Loud
M	55	Huntington's disease	Severe	Slow
M	43	Hydrocephalus	Severe	Loud
F	53	Multiple sclerosis	Mild–moderate	Habitual
F	60	Multiple sclerosis	Moderate–severe	Habitual
F	79	Parkinson's disease	Mild	Loud
M	76	Parkinson's disease	Mild	Loud
M	77	Parkinson's disease	Mild	Loud
M	67	Parkinson's disease	Mild	Habitual
F	83	Parkinson's disease	Mild	Slow
M	68	Parkinson's disease	Mild	Loud
M	89	Parkinson's disease	Mild	Loud
M	58	Parkinson's disease	Mild	Loud
M	73	Parkinson's disease	Mild	Slow
M	79	Parkinson's disease	Mild	Loud
M	69	Parkinson's disease	Mild	Loud
M	68	Parkinson's disease	Mild	Loud
F	70	Parkinson's disease	Mild–moderate	Slow
M	67	Parkinson's disease	Mild–moderate	Habitual
M	71	Parkinson's disease	Mild–moderate	Loud
F	73	Parkinson's disease	Mild–moderate	Loud
M	65	Parkinson's disease	Mild–moderate	Loud
M	75	Parkinson's disease	Moderate	Slow
M	79	Parkinson's disease	Moderate	Loud
M	71	Parkinson's disease	Moderate	Loud
M	69	Parkinson's disease	Moderate	Loud
M	81	Parkinson's disease	Moderate–severe	Loud
M	77	Parkinson's disease	Moderate–severe	Loud
F	67	Progressive supranuclear palsy	Mild	Slow
M	64	Spinocerebellar ataxia	Severe	Loud
M	72	Stroke	Severe	Habitual
F	48	Traumatic brain injury	Mild–moderate	Loud
M	55	Traumatic brain injury	Mild–moderate	Habitual
M	60	Traumatic brain injury	Moderate	Slow
M	47	Traumatic brain injury	Severe	Habitual
M	53	Undetermined neurological disease	Moderate	Habitual
F	45	Wilson's disease	Mild	Loud
M	56	Healthy		Habitual
M	67	Healthy		Habitual
M	70	Healthy		Habitual
M	70	Healthy		Loud
F	73	Healthy		Habitual
M	75	Healthy		Habitual
F	80	Healthy		Loud

Note. F = female; M = male