

Research Article

Investigating Acoustic Correlates of Intelligibility Gains and Losses During Slowed Speech: A Hybridization Approach

Frits van Brenk,^a Alexander Kain,^b and Kris Tjaden^a

Purpose: This exploratory study sought to identify acoustic variables explaining rate-related variation in intelligibility for speakers with dysarthria secondary to multiple sclerosis.

Method: Seven speakers with dysarthria due to multiple sclerosis produced the same set of Harvard sentences at habitual and slow rates. Speakers were selected from a larger corpus on the basis of rate-related intelligibility characteristics. Four speakers demonstrated improved intelligibility and three speakers demonstrated reduced intelligibility when rate was slowed. A speech analysis resynthesis paradigm termed *hybridization* was used to create stimuli in which segmental (i.e., short-term spectral) and suprasegmental variables (i.e., sentence-level fundamental frequency, energy characteristics, and duration) of sentences produced at the slow rate were donated individually or in combination to habitually produced sentences. Online crowdsourced orthographic transcription was used to quantify intelligibility for six

hybridized sentence types and the original habitual and slow productions.

Results: Sentence duration alone was not a contributing factor to improved intelligibility associated with slowed rate. Speakers whose intelligibility improved with slowed rate showed higher intelligibility scores for duration spectrum hybrids and energy hybrids compared to the original habitual rate sentences, suggesting these acoustic cues contributed to improved intelligibility for sentences produced with a slowed rate. Energy contour characteristics were also found to play a role in intelligibility losses for speakers with decreased intelligibility at slowed rate. The relative contribution of speech acoustic variables to intelligibility gains and losses varied considerably between speakers.

Conclusions: Hybridization can be used to identify acoustic correlates of intelligibility variation associated with slowed rate. This approach has further elucidated speaker-specific and individualized speech production adjustments when slowing rate.

Behavioral communication intervention for individuals with dysarthria frequently aims to maintain or improve speech intelligibility, commonly defined as the degree to which an individual's acoustic signal is understood by a listener (Weismer, 2008). An ongoing goal of dysarthria research is to identify specific speech production characteristics or combinations

of characteristics, including phonation, articulation, prosody, and resonance, that underlie or drive variation in intelligibility (De Bodt et al., 2002; Kim et al., 2011). Research identifying the relevant production characteristics contributing to intelligibility gains or losses for an individual potentially helps to shape targeted and patient-tailored behavioral treatments (Lansford et al., 2011; Yorkston, 2007).

In building the evidence base for using certain treatment techniques that may improve intelligibility and address deviant production characteristics, researchers have taken an approach in which intelligibility judgments and acoustic speech measures are obtained and correlated (Feenaughty et al., 2014; Kim et al., 2011; Neel, 2009). Acoustic measures reflecting impairments of the speech subsystems (i.e., respiration, phonation, articulation, resonance, and prosody) mediating intelligibility decline have been of particular interest (Chiu & Neel, 2020; Kent et al., 1989; McAuliffe et al., 2017; Weismer, 2008).

^aDepartment of Communicative Disorders and Sciences, University at Buffalo, NY

^bDepartment of Pediatrics, Oregon Health & Science University, Portland

Correspondence to Frits van Brenk: brenk@buffalo.edu

Editor-in-Chief: Jessica E. Huber

Editor: Angela C. Roberts

Received June 13, 2020

Revision received October 28, 2020

Accepted March 18, 2021

https://doi.org/10.1044/2021_AJSLP-20-00172

Publisher Note: This article is part of the Special Issue: Selected Papers From the 2020 Conference on Motor Speech—Clinical Science and Implications.

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Rate Reduction: Clinical Rationale, Perceptual and Acoustic Correlates

Rate reduction is a prominent behavioral management technique for improving intelligibility in dysarthria (Yorkston, 2007). A number of specific rate reduction methods have found their place in clinical practice, including pacing boards, alphabet boards, visual and auditory feedback, and custom cueing and pacing strategies (Blanchet & Snyder, 2009; Van Nuffelen et al., 2009; Yorkston et al., 2007). A variety of mechanisms have been proposed to explain the improved intelligibility that may accompany a slower than normal rate. For example, slow rate may increase articulatory precision by facilitating the achievement of more extreme vocal tract configurations, enabling greater acoustic distinctiveness during running speech (Blanchet & Snyder, 2010; Yorkston et al., 2007). Slow rate may also facilitate a speaker's ability to simultaneously manipulate multiple speech subsystems and enable speakers to produce more appropriate breath group units (Hardcastle & Tjaden, 2008; Yorkston et al., 2010). Research further suggests the influence of rate reduction on listener perceptual strategies. Slow rate may improve lexical processing by allowing the listener more time to decode speech (Liss, 2007; McAuliffe et al., 2014). However, speech produced too slowly may tax the listener's short-term memory, resulting in retrieval difficulties (Lim et al., 2019; Liss, 2007). Slow rate also has the potential to hinder speech segmentation by reducing the natural timing of strong–weak syllables (Utianski et al., 2011). Relatedly, atypical grouping of words could potentially hamper syntactic processing (McAuliffe et al., 2014).

Rate control as a behavioral management technique has shown treatment effectiveness across many neurological diagnoses and clinical dysarthria subtypes (Duffy, 2019; Yorkston et al., 2010). However, even within relatively homogeneous populations (i.e., speakers with the same neurological diagnosis, presumed underlying pathophysiology, and comparable severity), the impact of rate reduction on intelligibility has been found to vary. For example, Tjaden, Sussman, and Wilding (2014) investigated scaled intelligibility for sentences produced by speakers with dysarthria secondary to Parkinson's disease (PD) or multiple sclerosis (MS) in a variety of speaking conditions, including habitual and slow rates. When data were aggregated over speakers in each group, there was no significant difference in intelligibility between the habitual and slow rate conditions for either group. Instead, both groups included speakers with divergent responses to rate reduction, with intelligibility for some speakers improving, worsening, or showing no change for the slow rate relative to habitual (Tjaden, Sussman, and Wilding, 2014). The variable effects of rate reduction on intelligibility were further confirmed in a follow-up study employing orthographic transcription to quantify intelligibility (Stipancic et al., 2016). Similar trends were found by McAuliffe et al. (2017), who reported both intelligibility gains and losses following rate reduction for a group of speakers with hypokinetic dysarthria of matching severity.

To identify candidate speech production variables potentially responsible for intelligibility change associated with slow rate, a number of dysarthria studies have examined acoustic changes accompanying slow rate. For example, slow rate is associated with larger acoustic working spaces for vowels and consonants for individuals with a variety of dysarthrias and neurological diagnoses, including amyotrophic lateral sclerosis, MS, and PD (McRae et al., 2002; Mefferd, 2015; Turner et al., 1995; Weismer et al., 2000). Slow speech rates are also associated with shallower second formant (F2) slopes of diphthongs in speakers with dysarthria secondary to PD (Walsh & Smith, 2012) and, to a lesser extent, in speakers with dysarthria secondary to MS (Tjaden & Wilding, 2004). In addition, slow rate is associated with prosodic adjustments in dysarthria in the form of reduced phrase-level fundamental frequency (F0) range and reduced phrase-level intensity (McAuliffe et al., 2014; Tjaden & Wilding, 2004, 2011). A closer inspection of these studies suggests considerable individual speaker variation both within and between groups. For example, with respect to measures of articulatory–acoustic working space, within-group variation was primarily characterized by differences in the magnitude of acoustic change for slow rate relative to the habitual condition (McRae et al., 2002; Tjaden & Wilding, 2004; Turner et al., 1995). Thus, acoustic adjustments accompanying slowed rate may even diverge for speakers with the same dysarthria diagnosis or underlying etiology. These findings further highlight that the acoustic changes underlying intelligibility variation accompanying rate reduction in dysarthria are not fully understood. Moreover, studies reporting that slow rate is accompanied by segmental and suprasegmental acoustic changes cannot be taken as evidence that these speech production changes are causing intelligibility variations that may accompany slow rate. As discussed in the following section, digital signal-processing techniques are one approach for addressing this challenge.

Digital Signal Processing to Identify Acoustic Variables Causing Intelligibility Variation

In order to identify possible causal links between speech acoustic variables and intelligibility, digital signal-processing techniques have been employed to alter aspects of the acoustic signal and subsequently measuring the change in intelligibility (Bunton et al., 2001; Hammen et al., 1994; Hertrich & Ackermann, 1998; Turner et al., 2008). These approaches enable precise and targeted modifications of certain parameters of the speech signal, such as duration or intensity, while holding other acoustic properties constant. In this way, the contribution of the acoustic property of interest to intelligibility decrease or increase may be assessed independently of other acoustic parameters.

A digital signal-processing approach has been applied to neurotypical speech and to speech produced by speakers with dysarthria (Kain et al., 2007). Studies employing modifications of neurotypical speech have manipulated duration (Liu & Zeng, 2006), F0 contour (Binns & Culling, 2007;

Miller et al., 2010), spectral information (Krause & Braida, 2009), or a combination of these characteristics (Spitzer et al., 2007). These techniques may establish causality, as opposed to merely association, between acoustic modifications and intelligibility variation. Such an approach potentially aids the identification of relevant acoustic features that could be targeted in dysarthria treatment.

Hertrich and Ackermann (1998) made some preliminary inroads with dysarthria by using a pitch-synchronous linear predictive coding (LPC) analysis–manipulation–resynthesis approach to manipulate the temporal structure of utterances produced by two patients with ataxic dysarthria. Rhythm-manipulated versions of utterances were generated from baseline utterances by adjusting unstressed syllable durations, interword pauses, and overall speech rate. Listeners judged whether these acoustic alterations yielded improved perceptual judgments of naturalness, intelligibility, slowness, fluency, and rhythm. The results showed that the alterations of rate and rhythm improved ratings of most perceptual dimensions, with the exception of overall intelligibility. Results were interpreted to suggest that variables other than rhythm need to be considered as predictors of intelligibility in ataxic dysarthria.

In a similar manner, Bunton et al. (2001) employed LPC resynthesis to reduce the F0 range in sentence-level stimuli produced by speakers with hypokinetic dysarthria or unilateral upper motor neuron dysarthria and by neurotypical speakers. Subsequently, transcription and scaling scores were used to evaluate the change in speech intelligibility. Flattening the F0 range resulted in significant decreases in intelligibility for all speakers, but decreases in intelligibility were found to be larger in the speakers with dysarthria, indicating an increased detrimental effect of loss of perceptual F0 information to an already compromised acoustic signal. Kain et al. (2007) also employed LPC resynthesis to manipulate the formants of vowels in consonant–vowel–consonant words produced by a speaker with Friedreich’s ataxia. Intelligibility was significantly improved by manipulating vowel duration and formant reference points to resynthesized values approximating neurotypical productions. Rudzicz (2013) also applied a speech signal transformation system to sentence-level utterances produced by eight male speakers with dysarthria due to cerebral palsy as well as amyotrophic lateral sclerosis and age- and gender-matched neurotypical speakers, with the objective of increasing overall intelligibility. The signal-processing system used a serialized sequence of acoustic transformations, including the removal of repeated sounds and insertion of deleted sounds, devoicing of unvoiced phonemes, as well as temporal and spectral adjustments of the speech signal. Transcription scores indicated a significant improvement in word intelligibility when removing repeated sounds and inserting deleted sounds.

It is important to note that these prior studies predominantly manipulated acoustic characteristics of utterances produced by speakers with dysarthria, with the intention of approximating the acoustic characteristics of neurotypical speech (Bunton et al., 2001; Hertrich & Ackermann, 1998; Rudzicz, 2013). This approach does not consider whether

speakers with dysarthria are indeed capable of making such adjustments. Thus, even if meaningful changes in intelligibility were observed in the manipulated utterances and related acoustic features were identified, it is not guaranteed that behavioral treatment approaches designed to address these acoustic features are feasible and effective for the individual. A digital signal-processing approach that takes these limitations into account has been developed by Kain et al. (2008). This approach employs an analysis–resynthesis paradigm termed *hybridization* in which an acoustic parameter of interest, such as the F0 time history, is extracted from a sentence produced in one speech style (e.g., clear speech) and replaces the F0 time history of a habitually produced sentence, while holding the remaining acoustic parameters constant. The resulting “hybrid” sentence thus contains acoustic information from sentences actually produced by the speaker. The hybridization technique has been shown to have potential in identifying acoustic features underlying intelligibility gains in both neurotypical and disordered speech. In their 2008 study, Kain et al. created hybrid stimuli from sentences produced in a habitual and clear speaking style by a neurotypical speaker. Transcription scores indicated that a hybrid variant in which durational and spectral information from the clear speaking style sentences was imposed on the habitually produced sentences was found to be significantly more intelligible compared to the habitually produced sentences. In contrast, hybrids involving sentence-level energy and F0 did not show an increase in intelligibility, indicating that changes in short-term spectrum and duration explained the improved intelligibility for clearly produced sentences (Kain et al., 2008).

Thus far, the hybridization approach has been applied to a single published study investigating dysarthria, where it was used to identify acoustic variables explaining intelligibility variation in speech produced in conversational and clear speaking styles by two speakers with mild hypokinetic dysarthria secondary to PD (Tjaden, Kain, & Lam, 2014). A series of hybrids were created by extracting selected acoustic parameters from sentences produced in clear speaking style and imposing these on habitually produced sentences. Transcription scores indicated that hybrids involving short-term spectrum (i.e., segmental articulation) information for one speaker and energy trajectory information for another speaker yielded significant improvements in intelligibility. These results indicate that hybridization may be used to identify acoustic variables that cause intelligibility variation in mild dysarthria. These findings further suggest that acoustic adjustments underlying improved intelligibility may differ among speakers, even for speakers sharing similar dysarthria type, severity, or underlying neuropathology (Tjaden, Kain, & Lam, 2014).

Purpose

Against this background, the purpose of the present exploratory study was to use Kain et al.’s (2008) hybridization signal-processing technique to identify properties of

the acoustic signal explaining both increased and decreased intelligibility for sentences produced at a slow rate by speakers with dysarthria secondary to MS. MS is an appropriate population to study for several reasons. First, rate reduction has been recommended as a potential treatment technique for dysarthria in MS, and its feasibility in this population is suggested in published studies (Blanchet & Snyder, 2009; Duffy, 2019; Yorkston et al., 2010). The spastic-ataxic dysarthria typical of MS (e.g., strain-strangled harsh voice, slow rate, and irregular articulatory breakdown) also differs from hypokinetic dysarthria of PD (e.g., breathy voice, monotone, and segmental imprecision) and thus presents a different signal challenge for hybridization. In this manner, the current study addresses the viability of extending hybridization to a new clinical population and speech mode (i.e., slow rate). Sourced from an existing clinical speaker database (Stipanovic et al., 2016), we identified four speakers demonstrating improved intelligibility when slowing rate and three speakers demonstrating reduced intelligibility when slowing rate. By hybridizing segmental and suprasegmental aspects of speech produced at slow and habitual rates and subsequently examining the effect on crowdsourced transcription intelligibility, we sought to identify broad categories of acoustic variables (i.e., sentence-level energy, F₀, short-term spectrum, and duration) explaining or causing rate-related intelligibility gains and losses. Intelligibility measures were supplemented with speech production (i.e., acoustic) measures. Acoustic metrics were chosen to align with the broad acoustic categories represented in the hybrid variants and were intended to identify more specific speech production changes explaining the changes in intelligibility during the slow rate. This approach has the potential to inform clinical use of rate reduction and may also advance conceptual understanding of the relationship between intelligibility and acoustic features in dysarthria.

Method

Speakers

Speakers were five females and two males with a medical diagnosis of MS. All participants were native speakers of American English, had achieved at least a high school diploma, did not use a hearing aid, reported no other history of neurological disease, and demonstrated no more than mild cognitive impairment, as evidenced by a score of at least 26 points on the Standardized Mini-Mental State Examination (Molloy, 1999). Pure-tone thresholds were obtained by an audiologist at the University at Buffalo Speech-Language and Hearing Clinic for the purpose of providing a global indication of auditory status but were not used to exclude speakers from participation. Pure-tone thresholds averaged across 500, 1000, and 4000 Hz were 25 dB HL or better in both ears. An exception was speaker MSF03, who displayed elevated average pure-tone thresholds of 37 dB HL in the left ear.

The seven speakers were sourced from an existing database based on transcription intelligibility scores for the

same Harvard sentences (Institute of Electrical and Electronics Engineers, 1969) produced at both habitual and slow rates (Stipanovic et al., 2016). More specifically, speakers with at least a 10% difference in intelligibility between habitual and slow rate speech were studied. Previous research suggests this magnitude of intelligibility difference is likely clinically meaningful (Stipanovic et al., 2018; Van Nuffelen et al., 2010). Three speakers demonstrating “lower” intelligibility for slowed speech compared to habitual speech formed the “Low” group, and four speakers demonstrating “higher” intelligibility for slowed speech compared to habitual speech formed the “High” group. These speaker numbers exceed or are consistent with previously published studies using a digital signal-processing approach to determine the acoustic basis of intelligibility variation in dysarthria (Kain et al., 2008; Rudzicz, 2013; Tjaden, Kain, & Lam, 2014).

Speaker characteristics are detailed in Table 1. Sentence Intelligibility Test (SIT) scores indicate that intelligibility was generally preserved across speakers. However, Scaled Speech Severity scores, considered to be sensitive to mild dysarthria (Sussman & Tjaden, 2012), indicate the presence of a perceptually detectable dysarthria in all speakers. In addition, two certified speech-language pathologists employed a consensus approach to further characterize each speaker. Prominent deviant perceptual characteristics and accompanying dysarthria diagnoses reported in Table 1 were identified from clinical speech samples composed of diadochokinetic tasks, a vowel prolongation task, the Grandfather Passage, and a 2-min spontaneous monologue.

Speech Production Tasks

Each speaker read the same 25 Harvard sentences (Institute of Electrical and Electronics Engineers, 1969) at habitual and slow rates. Each sentence ranged in length between seven and nine words and contained five key words (i.e., the nouns, verbs, adjectives, and adverbs in each sentence). See the Appendix for the list of Harvard sentences used in this study. For detailed data collection procedures, see Tjaden, Sussman, and Wilding, (2014). Audio recordings were collected using an AKG C410 head-mounted microphone, with a constant mouth-microphone distance, positioned 10 cm and 45°–50° from the left oral angle. The acoustic signal was preamplified, low-pass filtered at 9.8 kHz, and sampled at 22 kHz. A calibration tone was recorded to measure vocal intensity offline. The sentences were first produced at a habitual rate followed by several other speech conditions, including a slow rate. For the slow rate, speakers were instructed to produce the sentences half as fast as their habitual rate, were encouraged to stretch out words rather than solely insert pauses, and were asked to produce each sentence on a single breath. Similar instructions have been used in other studies (e.g., McHenry, 2003; Mefferd, 2019). For each speaker, a random selection of the same 10 sentences produced at both habitual and slow rates was used for further analysis.

Table 1. Speaker characteristics for speakers with multiple sclerosis (MS).

Subject	Age (years)	Sex ^a	Years postdiagnosis	MS type ^b	Dysarthria diagnosis ^c	Clinical speech characteristics ^d	SIT scores (%) ^e	Scaled Speech Severity ^f	Harvard sentence intelligibility (habitual) ^g	Harvard sentence intelligibility (slow) ^g	Group
MSF02	66	F	18	PP	Spastic-ataxic	a, b, c	94.63	0.924	57.20	23.60	Low
MSF03	58	F	2	PP	Spastic	d, e, a	93.33	0.381	96.00	79.60	Low
MSF04	62	F	9	SP	Ataxic	f, c, g	94.63	0.448	61.20	79.60	High
MSF17	47	F	18	RR	Ataxic-spastic	f, c	93.22	0.561	78.00	61.20	Low
MSF20	55	F	26	SP	Spastic-ataxic	f, i, c	97.65	0.527	60.80	73.20	High
MSM05	55	M	5	SP	Spastic-ataxic	a, g, h, f	92.09	0.368	54.80	64.80	High
MSM07	55	M	4	RR	Ataxic-spastic	e, g, j, c	88.68	0.334	66.80	86.00	High

^aM = male; F = female. ^bPP = primary progressive; SP = secondary progressive; RR = relapsing remitting. ^cOrdered by relative prominence. ^dOrdered by relative prominence: a = slow rate; b = strain-strangled voice; c = harsh voice; d = breathy voice; e = imprecise consonants; f = irregular articulatory breakdown; g = monopitch; h = monoloudness; i = short phrases; j = excess and equal stress. ^eSentence Intelligibility Test (SIT) scores of the Assessment of Intelligibility of Dysarthric Speech (Yorkston et al., 2007) from Sussman and Tjaden (2012). ^fScaled Speech Severity for Grandfather Passage (from Sussman & Tjaden, 2012; naïve listeners: 0 = *no impairment* and 1.0 = *severe impairment*; for comparison: unaffected neurotypical speakers [$N = 32$]; Scaled Speech Severity, $M = 0.148$, $SD = 0.07$). ^gAs reported in Stipanovic et al. (2016).

Hybridization, Speech Resynthesis, and Stimulus Preparation

The hybridization technique was a residual-excited LPC waveform resynthesis method (similar to Taylor et al., 1998). Each habitually produced sentence was used as a base sentence while selectively imposing or donating the energy envelope, F0 envelope, segment durations, or short-term spectra from the same sentence produced at slow rate by the same speaker. These acoustic parameters were selected based on previously published hybridization studies demonstrating the importance of these acoustic cues in relation to intelligibility gains for clear speech (Kain et al., 2008; Tjaden, Kain, & Lam, 2014). Although additional hybrids might be constructed, the selection was limited to the most promising hybrids to keep experimental length and listener subject numbers within acceptable boundaries.

A schematic illustrating the procedure for creating hybridized sentences is provided in Figure 1. A comprehensive description of the hybridization process may be found in Kain et al. (2008) and Tjaden, Kain, and Lam (2014) and can be summarized in six steps: (a) In an alignment step, segment boundaries and individual glottal pulses were manually identified using Praat (Boersma & Weenink, 2020) and aligned across sentence types to compensate for possible differences in phonemic content. (b) To parallelize the two sentence variants, the waveform of the habitual rate sentences was modified by implementing phoneme deletions or insertions relative to the slow rate waveforms using a time domain cross-fade technique to avoid discontinuities. As a result, the phoneme sequence of the resulting hybridized waveform was identical to those of the waveform of

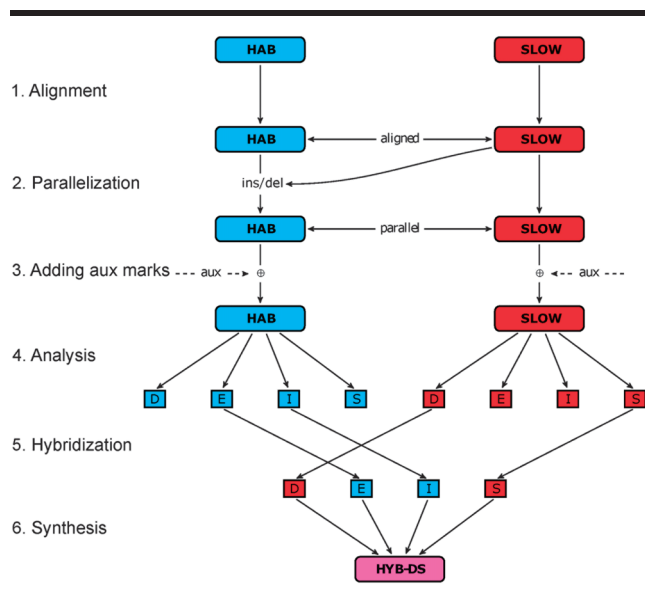
the slow rate sentences. (c) An automated acoustic analysis was carried out on the parallelized waveforms to determine the location of auxiliary marks used for subsequent prosodic modification. (d) In the analysis step, segment durations, short-term spectra, energy trajectories, and F0 trajectories were extracted. (e) During the hybridization step, the acoustic characteristics of interest donated by the slow rate speech samples were combined with complementary acoustic characteristics of habitual speech to form hybrid acoustic parameters. (f) During the final synthesis step, six hybrid speech waveforms were created: intonation (I), energy (E), duration (D), prosody (defined as the combination of intonation, energy, and duration [IED]), short-term spectrum (S), and the combination of duration and short-term spectrum (DS). The 7 speakers \times 10 sentences \times 8 sentence variants (habitual rate, slow rate, and the six hybrid versions) yielded a total of 560 stimuli that were orthographically transcribed by listeners, as described below.

Speakers had mild dysarthria based on transcription scores obtained using the SIT (Yorkston et al., 1996), as reported in Table 1. Thus, to prevent ceiling effects (i.e., transcription scores approaching 100% regardless of sentence variant) in the perceptual task described below, sentence variants were mixed with multitalker babble. Pilot testing of signal-to-noise ratios (SNRs) indicated that an SNR of 0 dB minimized floor and/or ceiling effects. Sentence variants were first intensity-normalized via an A-weighted root-mean-square measure and then mixed with 10-talker babble at an SNR of 0 dB.

Acoustic Measures

Acoustic measures were completed to document the nature of acoustic differences between the habitual and slow rate conditions that may explain the changes in intelligibility during the slow rate. Measures were selected to reflect segmental and suprasegmental acoustic changes usually associated with slow rate and those suggested in previous studies to potentially explain rate-related variation in intelligibility (Kuo & Tjaden, 2016; Tjaden & Wilding, 2004, 2011). In addition, these acoustic measures map on to the presumed acoustic changes imposed by the six different hybrid variants. Specifically, acoustic measures of interest included sentence durations to reflected durational changes primarily imposed by the Hybrid-D variant. Measures of articulatory vowel space areas (VSAs), vowel distinctiveness as measured by F2 interquartile range (IQR) and vocal quality assessed by smoothed cepstral peak prominence (CPPs), were included to assess spectral changes imposed by the Hybrid-S and Hybrid-DS variants. Measures of sound pressure level (SPL) were included to examine adjustments to intensity contours by the Hybrid-E and Hybrid-IED variants, and F0 measures were included to examine adjustments to F0 contours by the Hybrid-I and Hybrid-IED variants. Sentence durations were also obtained to evaluate temporal changes across rate conditions to verify that both speaker groups were slowing rate when instructed to do so. Sentence onsets and offsets were identified using

Figure 1. Block diagram summarizing the hybridization process. aux = auxiliary; D = duration hybrid; del = deletion; E = energy hybrid; HAB = habitually produced speech; HYB-DS = duration and short-term spectra hybrid; I = intonation hybrid; ins = insertion; S = short-term spectrum hybrid; SLOW = slow rate speech.



standard acoustic criteria. VSAs of tense and lax vowels were obtained following procedures detailed in Tjaden et al. (2013) and can be summarized as follows. Formant trajectories of the first formant (F1) and F2 were derived by LPC using TF32 (Milenkovic, 2005). Formant traces were hand-corrected in case of computer-generated tracking errors. Vowel onsets and offsets were identified using conventional acoustic criteria, and F1 and F2 frequencies were extracted at 50% of the vowel duration. All 25 Harvard sentences were used to obtain F1/F2 values for three to five occurrences of each of the four tense and four lax vowels to be able to include a minimum number of eligible tokens for computing VSAs. The VSAs of tense and lax vowels were calculated from the midpoint F1 and F2 frequencies. SPL (mean and standard deviation was obtained using TF32 (Milenkovic, 2005). For each speaker recording session, a calibration tone was recorded and used to account for differences in recording level or gain when calculating SPL from the acoustic signal. To measure SPL mean and standard deviation, Harvard sentences were segmented into phrases or “runs” (i.e., speech bounded by silent period of at least 200 ms). Subsequently, mean and standard deviation values were averaged across runs. Furthermore, a number of acoustic measures were quasi-automatically obtained by means of custom scripts using Praat. These measures included F0 (mean, standard deviation, 90th–10th percentile range) and CPPs, obtained at sentence length. F2 IQR (F2 third quartile–first quartile) was also obtained to quantify segmental characteristics for vocalic segments, following Kuo et al. (2014) and Yunusova et al. (2005). F2 trajectories were measured for concatenated vowel, liquid, and glide segments within each sentence.

Listeners and Perceptual Methodology

A total of 885 adults (522 women, 351 men, 12 not specified), with ages of 18–80 years ($M = 35.3$, $SD = 10.7$), judged intelligibility using orthographic transcription. All listeners self-reported to be native speakers of American English; living in the United States; and without a history of speech, language, or hearing problems. Extended demographic information for listeners is reported in Table 2. Participants were recruited using the crowdsourcing website Amazon Mechanical Turk (MTurk; mturk.com). Previous studies have demonstrated the feasibility of using MTurk-sourced transcription scores to quantify intelligibility in dysarthria (Lansford et al., 2016; Nightingale et al., 2020; Yoho & Borrie, 2018). Participants were allowed to participate after they fulfilled a number of prerequisites, including an approval rate of greater than or equal to 99%, more than 50 approved tasks on MTurk, and confirmed status of U.S. residence.

This study was reviewed and approved by the institutional review board of the University at Buffalo. After consenting to participate using the institutional review board–approved consent form, listeners were instructed to transcribe a series of stimuli while using headphones and working in a quiet environment. Sentences were presented one at a time. Following each sentence presentation, listeners

Table 2. Demographic details of listener participants.

Characteristic	<i>n</i>	%
Gender		
Male	351	39.6
Female	522	59.0
Other/n.s.	12	1.4
Age (years)		
≤ 29	299	33.8
30–39	315	35.6
40–49	152	17.2
50–59	71	8.0
≥ 60	28	3.2
Not specified	20	2.2
Education completed		
Doctoral degree	16	1.8
Master’s degree	99	11.2
Bachelor’s degree	309	34.9
Associate degree	165	18.6
High school	283	32.0
Other/n.s.	13	1.5
Region		
Midwest	189	21.4
Northeast	194	21.9
Southeast	256	28.9
Southwest	80	9.2
Rocky mountains	24	2.7
Pacific	134	15.1
Other/n.s.	8	0.9

Note. n.s. = not specified.

were asked to transcribe the sentence as accurately as possible and to guess any words that they did not recognize. Stimulus replay was allowed (cf. Eijk et al., 2020; Fletcher, McAuliffe, Lansford, Sinex, & Liss, 2017). Stimuli were presented in a blocked and randomized fashion, ensuring that no identical sentence text was presented twice, irrespective of speakers or sentence variant. After completing the sentence transcription, participants were asked to complete a demographic questionnaire. The experiment took between 5 and 9 min with associated remunerations between \$0.80 and \$1.20. Listeners were allowed to participate only once. For quality control purposes, transcription data from listeners failing to respond to more than 20% of the presented stimuli, listeners failing to correctly transcribe at least four out of five key words in a control sentence without noise, or listeners who scored 2 *SDs* below the overall mean transcription score were excluded. For each of the 560 stimuli, a minimum of 20 valid transcription scores were obtained (cf. Lansford et al., 2016; Yoho & Borrie, 2018).

Data Analysis

The mean percentage of correctly transcribed key words was calculated for each stimulus sentence, with each sentence containing five key words. The sentence transcripts were manually analyzed for correct words using a rule-book of scoring criteria customized for the stimuli under investigation. As the percentage scores were not normally distributed, the raw transcription scores were subjected to a rationalized arcsine transformation to be expressed as

rationalized arcsine units (rau; Studebaker, 1985). In addition, this transformation to a linear scale ensured valid statistical comparisons (Chiu & Neel, 2020). The transcription scores expressed as rau were the primary dependent variable for comparing groups, namely, speakers with lower intelligibility in the slow versus habitual condition (“Low”) and speakers with higher intelligibility in the slow versus habitual condition (“High”).

Statistical analyses were carried out using R software (R Core Team, 2019). The transcription scores were analyzed separately for the Low and High groups using linear mixed-model analyses. The models contained sentence variant as fixed factor as well as speaker and Harvard sentence as independent random factors. For completeness, the relationship between laboratory-sourced transcription scores reported in Stipancic et al. (2016) for habitual and slow sentence productions and the raw crowdsourced transcription scores for the nonhybridized habitual and slow productions was also examined with Pearson product-moment correlations. In addition, each acoustic measure was analyzed using a linear mixed model with rate (habitual and slow) and group (High and Low) as fixed factors as well as Harvard sentence and speaker as random factors, with the exception of VSAs, which were computed using measures pooled across sentences. As the High and Low speaker groups included different proportions of males and females, speaker gender was evaluated as covariate. Gender significantly contributed to the model analyzing F0 mean. For all other acoustic measures, the effect of gender was absent and subsequently removed as covariate. Significant main effects and interactions were further examined with post hoc pairwise comparisons of estimated marginal means, using Tukey’s method to correct for multiple comparisons. Standardized effect sizes, expressed as Cohen’s *d*, were derived from the estimated marginal means and population standard deviations. In addition to primary analyses for group trends (i.e., High vs. Low), individual speaker data were examined for completeness, consistent with the two prior hybridization studies focusing on individual speaker findings (Kain et al., 2008; Tjaden, Kain, & Lam, 2014). A significance level of .05 was used for all hypothesis testing.

Results

Intelligibility of Sentence Variants

The intelligibility scores expressed as rationalized arcsine transformed transcription scores (means and standard deviations) for the High group are displayed in Figure 2, and transcription scores for the Low group are displayed in Figure 3. Transcription scores of the six hybridized variants are displayed on the left, with scores for nonhybridized habitual and slow rate stimuli on the far right.

The linear mixed-model analysis for the High group indicated a significant main effect of sentence variant, $F(7, 8487) = 108.2, p < .0001$. Post hoc pairwise testing comparing the habitual and slow rates indicated higher transcription scores for the slow versus habitual rate: mean

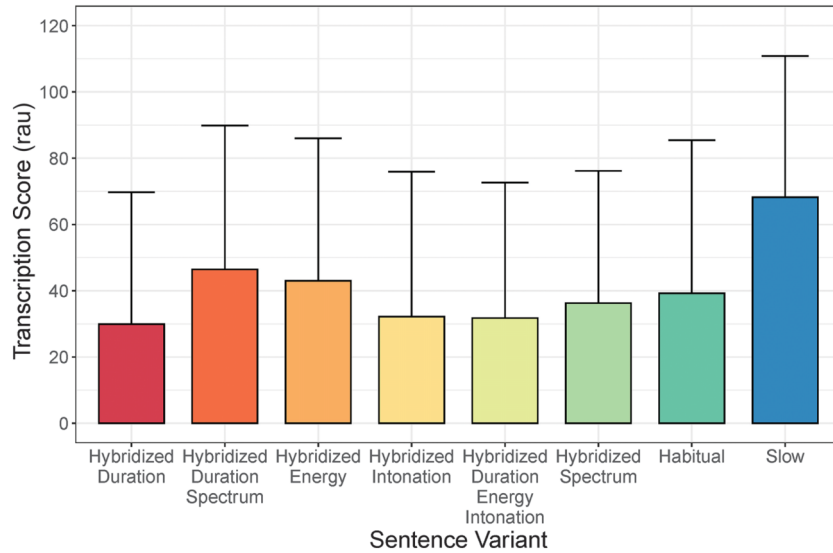
difference = 27.3 rau, $z = 16.2, p < .0001, d = 0.71$. This effect was present for each speaker in the High group: MSF04 mean difference = 22.9 rau, $z = 7.65, p < .0001, d = 0.63$; MSF20 mean difference = 29.3 rau, $z = 8.37, p < .0001, d = 0.80$; MSM05 mean difference = 29.8 rau, $z = 9.27, p < .0001, d = 0.82$; MSM07 mean difference = 27.3, $z = 8.55, p < .0001, d = 0.75$.

Hybrid versions that may explain acoustically driven intelligibility differences between the habitual and slow rates were of particular interest. Based on the finding of significantly higher intelligibility for slow rate versus habitual rate, a series of post hoc analyses were performed to identify hybrids with higher intelligibility versus habitual, thus becoming more slow-like with respect to intelligibility. Post hoc results comparing the habitual condition to each hybrid variant indicated a significant increase in intelligibility for the Hybrid-DS variant: mean difference = 6.85 rau, $z = 4.02, p = .0015, d = 0.18$. Analysis of individual speaker data revealed that this difference was statistically significant only for MSF20 (mean difference = 16.8 rau, $z = 3.53, p < .0001, d = 0.46$). No statistically significant differences in intelligibility were found for MSF04 (mean difference = 6.06 rau, $z = 1.95, p = .52, d = 0.16$), MSM05 (mean difference = 4.41 rau, $z = 1.45, p = .84, d = 0.12$), or MSM07 (mean difference = 2.89 rau, $z = 0.86, p = .98, d = 0.08$). Post hoc analyses further indicated a trend toward higher intelligibility for the Hybrid-E variant versus habitual: mean difference = 4.75 rau, $z = 2.89, p = .075, d = 0.12$. Individual speaker data indicated that this trend could be attributed to speaker MSF04 (mean difference = 7.11 rau, $z = 2.31, p = .029, d = 0.20$). No significant differences were found for MSF20 (mean difference = 9.25 rau, $z = 2.79, p = .097, d = 0.25$), MSM05 (mean difference = 2.85 rau, $z = 0.99, p = .98, d = 0.08$), or MSM07 (mean difference = 3.76 rau, $z = 1.13, p = .95, d = 0.10$). Thus, the group effects suggesting the contribution of combined spectral and durational properties and energy characteristics to increased intelligibility for slow rate were mainly driven by MSM20 and, to a lesser extent, by MSF04.

The statistical analysis for the Low group indicated a significant main effect of sentence variant, $F(7, 6156) = 82.0, p < .0001$. Post hoc comparisons indicated higher transcription scores for habitual versus the slow rate: mean difference = 17.6 rau, $z = 9.75, p < .0001, d = 0.52$. Each speaker in the Low group also had significantly higher transcription scores for the habitual rate versus the slow rate: MSF02 mean difference = 17.4 rau, $z = 4.98, p < .0001, d = 0.48$; MSF03 mean difference = 16.5 rau, $z = 4.98, p < .0001, d = 0.45$; MSF17 mean difference = 17.6 rau, $z = 5.36, p < .0001, d = 0.48$.

As the acoustic properties of the slow rate were donated to the habitual sentences, the hybrids yielding a significant increase in intelligibility relative to slow (i.e., becoming more habitual-like) were considered to have a preserving effect on intelligibility. All other hybrids may have contributed to the decrease in intelligibility for slow rate. Post hoc testing identified the Hybrid-E variant to have significantly higher intelligibility scores compared to slow: mean difference =

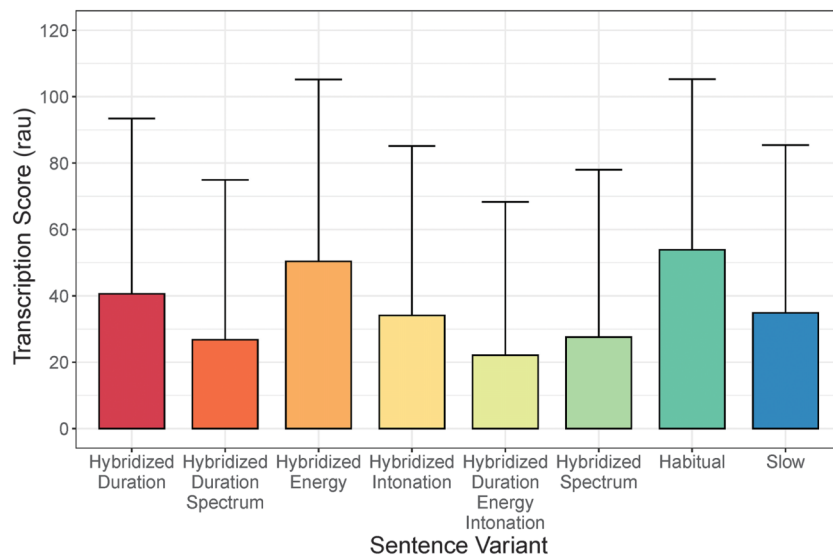
Figure 2. Mean transcription scores and standard deviations of the speakers in the High group for the hybridized sentence variants as well as the habitual and slow rate conditions. rau = rationalized arcsine units.



11.4 rau, $z = 6.42$, $p < .0001$, $d = 0.33$. Analysis of individual speaker data showed that this effect was driven by two speakers. While MSF02 did not show a significant change when slowing rate (mean difference = -2.1 rau, $z = -0.61$, $p = .99$, $d = -0.05$), significant increases were found for MSF03 (mean difference = 14.9 rau, $z = 4.70$, $p < .0001$, $d = 0.41$) and MSF17 (mean difference = 18.5 rau, $z = 5.62$, $p < .0001$, $d = 0.51$). Thus, for two speakers in the Low group, acoustic characteristics other than energy contributed to decreased intelligibility for slow rate.

Comparison of laboratory-sourced transcription scores (cf. Stipanovic et al., 2016) and raw crowdsourced transcription scores expressed as percent correct scores indicated that, for speakers in the High group, average crowdsourced transcription scores were 42.0% (habitual) and 63.9% (slow) compared to laboratory-sourced data of 60.9% (habitual) and 75.9% (slow). For the speakers in the Low group, crowdsourced transcription scores were 52.2% (habitual) and 39.1% (slow), compared to laboratory-sourced scores of 77.1% (habitual) and 54.8% (slow). There was a strong

Figure 3. Mean transcription scores and standard deviations of the speakers in the Low group for the hybridized sentence variants as well as the habitual and slow rate conditions. rau = rationalized arcsine units.



significant correlation between laboratory-sourced and crowdsourced transcription scores of the seven speakers at habitual and slow rates, $r(14) = .92, p < .0001$.

Acoustic Measures

The descriptive results for acoustic measures are summarized in Table 3 for both groups. Means and standard deviations for each measure are displayed for both the habitual and slow rates.

Sentence Duration

The statistical results indicated no significant effect of group, $F(1, 5.00) = 2.63, p = .166$. However, the effect of rate was significant, $F(1, 109.9) = 253.5, p < .0001$, where across groups, sentence durations were longer in the slow rate compared to the habitual condition. A significant Group \times Rate interaction effect was also present, $F(1, 109.9) = 31.0, p < .0001$. Post hoc analyses indicated that both speaker groups decreased rate from habitual to slow (High group: mean difference = 1.23 s, $z = 7.91, p < .0001, d = 1.77$; Low group: mean difference = 2.55 s, $z = 14.8, p < .0001, d = 3.67$). Rate was significantly reduced for all speakers of the High group (MSF04 mean difference = 1.98 s, $z = 7.62, p < .0001, d = 3.41$; MSF20 mean difference = 1.27 s, $z = 4.89, p < .0001, d = 2.19$; MSM05 mean difference = 0.80 s, $z = 3.09, p = .002, d = 1.38$; and MSM07 mean difference = 0.86 s, $z = 3.29, p = .001, d = 1.47$), as well as for speakers in the Low group (MSF02 mean difference = 3.24 s, $z = 12.5, p < .0001, d = 5.57$; MSF03 mean difference = 1.23 s, $z = 4.73, p < .0001, d = 2.12$; and MSF17 mean difference = 3.17 s, $z = 12.2, p < .0001, d = 5.46$). When comparing groups within each rate condition, results indicated no significant difference between the High and Low groups for the habitually produced sentences (mean difference = 1.46 s, $z = 1.11, p = .266, d = 2.1$). There was a significant difference for the slow rate sentences (mean difference = 2.78 s, $z = 2.12, p = .034, d = 4.0$), with longer sentence durations produced by the Low group, indicating that the Low group slowed rate to a greater extent compared to the High group.

VSA

For both the High and Low speaker groups, tense and lax VSAs did not significantly contract or expand when slowing rate. However, the main effect of group approached significance for both VSAs: lax VSA, $F(1, 5) = 5.48, p = .066$, and tense VSA, $F(1, 5) = 6.30, p = .054$, with larger VSAs for the Low group compared to the High group. There were no significant differences in VSAs for habitual and slow rates for individual speakers.

F2 IQR

Statistical analysis of F2 IQR indicated no effect of group or Group \times Rate interaction. The main effect of rate approached significance, $F(1, 107) = 3.77, p = .055$,

indicating higher F2 IQR values for the slow rate. Visual inspection of the data indicated that this trend was attributable to the High group, who exhibited increased F2 IQR values for the slow rate. Further analysis of individual speaker data did not reveal any notable trends.

SPL

Analysis of mean SPL outcomes indicated no significant effects of group and rate. The Group \times Rate interaction was also not significant. Acoustic measures of SPL variation (SPL standard deviation) showed a significant effect of rate, $F(1, 109) = 8.98, p = .0034$, with larger SPL standard deviation for the slow rate. The main effect of group and the Group \times Rate interaction effect were nonsignificant.

F0

The analysis of mean F0 measures showed a significant effect of gender, $F(1, 4.01) = 15.3, p = .017$. No significant main or interaction effects were found; mean F0 did not significantly change for both speaker groups when slowing rate. The results of F0 variation (F0 standard deviation) analysis indicated a main effect of rate, $F(1, 131) = 4.84, p = .030$, with a higher F0 standard deviation for slow rate. Effects of group and Group \times Rate were nonsignificant. Results of the analysis of F0 range were not significant.

CPPs

Analysis of the CPPs at the sentence level showed a significant effect of rate, $F(1, 131) = 12.5, p = .0006$, with higher CPPs values for slow rate sentences. The effect of group was not significant, but the Rate \times Group effect was significant, $F(1, 131) = 8.80, p = .0035$. The post hoc analysis showed that the increase in CPPs when slowing rate was solely driven by the Low group (mean difference = 1.03 dB, $z = 4.31, p < .0001, d = 1.11$), while the rate effect was not significant for the High group. Analysis of individual speaker data for the Low group indicated an increase in CPPs when slowing rate for two speakers: MSF02 (mean difference = 1.24 dB, $z = 3.37, p = .0008, d = 1.51$) and MSF17 (mean difference = 2.40 dB, $z = 6.54, p < .0001, d = 2.92$). An increase in CPPs also was found for speaker MSF04 of the High group: mean difference = 0.94 dB, $z = 2.55, p = .011, d = 1.14$.

Discussion

Maximizing intelligibility is an important treatment goal for speakers with dysarthria, including dysarthria of MS. Studies evaluating the effect of slowed rate on intelligibility have reported variable outcomes. Improved knowledge of the source of these variable outcomes is important for optimizing the clinical use of rate reduction techniques in the treatment of dysarthria and also may advance conceptual understanding of intelligibility. The current study used an analysis–hybridization–resynthesis technique to create sentence variants that incorporated habitual-to-slow

Table 3. Acoustic measures (mean and standard deviation) for habitual and slow rate sentence productions.

Group	VSA (Hz ²)		Dur (s)	SPL (dB)		F0 (Hz)			F2 IQR (Hz)	CPPs (dB)
	Tense	Lax		M	SD	M	SD	Range		
High										
Habitual	249631 (160193)	57406 (36577)	2.47 (0.49)	70.1 (3.0)	8.0 (1.0)	136.0 (29.5)	46.7 (17.5)	70.5 (50.5)	639 (194)	6.49 (1.07)
Slow rate	298161 (161893)	72622 (53146)	3.70 (1.07)	69.8 (5.0)	8.5 (1.6)	140.1 (33.4)	49.6 (18.4)	66.6 (50.5)	704 (235)	6.58 (1.59)
Low										
Habitual	591284 (226230)	123735 (27057)	3.90 (1.89)	69.7 (2.4)	8.5 (1.0)	177.0 (19.3)	40.3 (17.7)	86.0 (57.6)	735 (190)	7.04 (1.84)
Slow rate	628224 (168800)	135986 (13721)	6.44 (2.68)	71.2 (6.2)	9.0 (1.0)	176.1 (18.6)	48.2 (16.2)	108.1 (55.4)	784 (216)	8.07 (2.23)

Note. VSA = vowel space area; Dur = duration; SPL = sound pressure level; F0 = fundamental frequency; F2 IQR = second formant interquartile range; CPPs = smoothed cepstral peak prominence.

acoustic changes produced by speakers with dysarthria secondary to MS.

Before considering hybridization results, it is worthwhile to consider how crowdsourced transcription intelligibility results obtained in the current study for habitual and slow sentence productions compared to transcription results obtained in the laboratory, as reported in Stipancic et al. (2016). Analysis of the crowdsourced transcription scores indicated that each of the speakers in the High group indeed displayed higher intelligibility during slow rate compared to habitual rate and each of the speakers in the Low group was significantly more intelligible at habitual compared to slow rate. The crowdsourced transcription scores were, on average, about 18% lower compared to laboratory-sourced transcription scores, even in more favorable speech-to-noise listening conditions (0 dB vs. -3 dB). Differences in laboratory and online transcription scores were slightly larger in the habitual conditions (approximately 20% difference) compared to the slow conditions (approximately 15% difference). One possibility is that the slow condition allowed more time for the online recruited listeners to correctly transcribe sentences, possibly leveling the listening circumstances across groups to a larger extent compared to the habitual condition. The finding of lower crowdsourced transcription scores may be attributed to a higher quality listening environment for laboratory-sourced judgments and a reduced control over the online experimental setup and hardware. In addition, despite the use of quality checks and the efforts to signal and discard data suggestive of inattentive or uncooperative rater behavior, overall online listener engagement may have been lower compared to traditional methods of collecting ratings (Byun et al., 2015). Nevertheless, while the overall lower absolute performance of online listeners compared to laboratory listeners caution against directly comparing absolute transcription scores across data collection paradigms, the “relative differences” across speakers and speaking styles were comparable for both data collection methods, as evidenced by the strong significant positive correlation, providing further support for employing crowdsourced perceptual ratings of disordered speech.

A number of further observations can be derived from the perceptual results for hybridized stimuli. First, transcription scores for the High group showed that sentence duration alone did not play an observable role in the increased intelligibility associated with slowed rate. The associated Hybrid-D variant, in which the durational characteristics of sentences produced at a slow rate were donated to habitually produced sentences, did not result in an increase in intelligibility relative to the habitually produced sentence. The current findings therefore suggest that a presumed listener benefit of slowed rate, that is, allowing the listener more processing time to decode the signal (cf. Liss, 2007), did not explain the High group’s increased intelligibility for the slow rate. This is in line with findings by Hammen et al. (1994), who lengthened segment durations in sentences produced by speakers with dysarthria and reported no improvement in intelligibility compared to habitually produced sentences. These findings are corroborated by

the intelligibility results of the Low group who, despite slowing down to a larger degree compared to the High group, showed decreased intelligibility during slow rate compared to habitual rate.

While, for the High group, duration alone did not cause improved intelligibility for slow rate, the combination of spectral and durational hybridization (DS) adjustments and, to a lesser extent, energy (E) adjustments did contribute to the increase in intelligibility for slow rate sentences. Further analysis of individual speaker indicated that these findings were largely carried by select speakers in the High group, indicating substantial within-group variation with respect to the acoustic characteristics that may explain increased intelligibility for slow rate. This result, in turn, may be evidence of the diverse underlying strategies speakers employ to increase intelligibility when slowing rate (Yorkston et al., 2010).

While the Hybrid-DS variant was the most prominent contributor to increased intelligibility for sentences produced at a slow rate, these findings were not unequivocally captured in the acoustic measures. With respect to VSAs, no statistically significant differences were observed across groups and rate conditions. This result contrasts with studies reporting that VSA reductions are associated with reduced intelligibility in speakers with dysarthria (e.g., Fletcher, McAuliffe, Lansford, & Liss, 2017; Lansford & Liss, 2014), but see the work of Weismer et al. (2000), who did not find a relationship between VSAs and perceptual impressions of intelligibility as well as individual speaker results reported in Turner et al. (1995). The crowdsourced intelligibility findings in the current study would also seem to suggest larger VSAs for the High group when slowing rate. While not statistically significant, possibly due to large within-group differences, both speaker groups increased tense and lax VSAs. For tense VSAs, the High group increased VSAs about 20%, while the Low group showed an increase of about 6%. For lax VSAs, the High group increased VSAs by 26% and the Low group by 10%. These results suggest that the increased intelligibility of the High group at slow rate may have been at least partly driven by the larger VSA expansions realized by these speakers.

Results for F2 IQR were similarly challenging to align with the perceptual findings. Measures of F2 IQR are thought to be reflective of articulatory movement and overall working space, and F2 IQR has been found to be reduced in speakers with dysarthria and a predictor of intelligibility (Yunusova et al., 2005), although other studies have reported complex relationships between intelligibility and F2 IQR (Feenaughty et al., 2014). The current intelligibility results would seem to suggest larger F2 IQR values for the High group at slow rate, but this was not borne out in the statistical analysis. However, the High group did show a slightly larger increase in F2 IQR compared to the Low group (10% vs. 6%) when slowing rate. Overall, the results of VSA and F2 IQR measures in relation to the perceptual findings indicate that the spectral cues mediating change in intelligibility were modest at best. One implication

is that spectral cues other than vowel characteristics may play a role in explaining intelligibility variation.

The third acoustic measure that may help to explain perceptual results with respect to the Hybrid-DS variant is CPPs. CPPs is a measure of the periodicity of the spectrum of a signal, with higher values associated with perceptual impressions of decreased dysphonia (Lowell et al., 2011; Riesgo & Nöth, 2019). To the extent that a lower degree of dysphonia might be associated with improved intelligibility, the perceptual results would seem to suggest higher CPPs values for the High group when slowing rate. The results of CPPs sentence-level measures, however, indicated an increase in CPPs at slow rate for the Low group only. A possible explanation might be that the greater slowing of rate by the Low group compared to the High group contributed to the improved voice quality for these speakers. The current results are further evidence that measures of cepstral peak prominence may not be adequate predictors of intelligibility gains (Fletcher, McAuliffe, Lansford, Sinex, & Liss, 2017). Overall, these incongruities between perceptual and acoustic findings are difficult to explain but are not unlike those reported in other studies (e.g., Bunton & Weismer, 2001; McRae et al., 2002).

While mean SPL did not vary across groups and rates, SPL variation (SPL standard deviation) increased for slowed rate. This increase in SPL variation at slow rate might be interpreted as improved prosodic modulation, potentially beneficial to the listener and contributing to improved intelligibility in the High group (Duffy, 2019). Intonational information, however, did not seem to influence listener perception, as evidenced by an absence of intelligibility improvements of the Hybrid-I variant for both groups. The acoustic measures of F0 revealed an increase in F0 standard deviation during slowed rate, but the absence of any group-related effects suggests the absence of any differential effect for the High and Low groups. Reduced variation in F0 is usually linked to impressions of monopitch and has been found to be associated with a decrease in intelligibility in dysarthria (Bunton et al., 2001). Qualitatively drawing on the results for F0 range, it can be noted that the speakers in the High group generally maintained F0 range at habitual levels when slowing rate, while the Low group markedly increased F0 range. This trend for an increased F0 range might be evidence of articulatory distortions, excess stress, or poorly controlled pitch, possibly contributing to the decreased intelligibility at slow rate for the Low group (Lowit et al., 2010).

For the Low group, almost all acoustic hybrid parameters potentially explain the intelligibility decline for slowed speech. Of the different hybrids, the E variant was the only variant that contributed to preserving intelligibility when slowing rate, with findings of significantly higher transcription scores for the Hybrid-E variant compared to slow rate for speakers MSF03 and MSF17. Given the beneficial effect of the Hybrid-E variant to intelligibility in speakers of the High group, these findings suggest that energy envelope characteristics can be an important factor in enhancing or maintaining intelligibility. The results of the acoustic analyses further suggest a mediating role of SPL variation, of which

speakers in both groups showed significant increases when slowing rate. Energy contours with increased variation may help listeners parse and identify strong syllable onsets, thus contributing to improved intelligibility (Liss, 2007). Studies evaluating the intelligibility benefit of clear speech strategies have reported similar beneficial effects of energy envelope modulation characteristics (Krause & Braida, 2002, 2009; Tjaden, Kain, & Lam, 2014). Additional studies are needed to determine whether energy envelope characteristics might explain the improved intelligibility associated with increased vocal intensity, as reported in other studies (Cannito et al., 2011; Sapir et al., 2003). Using the hybridization paradigm to systematically evaluate the contribution of energy contour characteristics to the intelligibility gain of loud speech would improve the evidence base of this treatment in the management of dysarthria.

Speakers in the current study varied with respect to MS type, dysarthria diagnosis, and clinical speech characteristics. Given the small number of participants, it is difficult to distinguish patterns of speaker characteristics that may predict or connect to the results of the perceptual and acoustic analyses. Both speaker groups included speakers with spastic-ataxic mixed dysarthria, exhibiting slow rate, imprecise consonants, irregular articulatory breakdowns, and harsh voice as most prominent clinical speech characteristics. Considering the important role of the Hybrid-DS variant in mediating intelligibility gains for the High group, it may be predicted that imprecise consonants and irregular articulatory breakdown would be among the most prominent deviant speech characteristics for this group. However, these are also present in speakers of the Low group. Overall, these results underline the varied speech characteristics and individualized speech profiles present in speakers with MS.

Limitations and Conclusions

Results are constrained in a number of aspects. First, as the crowdsourced transcription scores were not normally distributed, overall effect sizes of the perceptual results were relatively small, limiting the strength of observations that may be derived from the results. Nonnormally distributed transcription scores have also been observed in studies using laboratory-sourced transcriptions (e.g., Chiu & Forrest, 2018; Fletcher et al., 2019; Yunusova et al., 2005). Even after transformation of crowdsourced or laboratory-sourced transcription scores, this might obscure potentially relevant findings for individual speakers where differences in transcription scores for habitual and slow rates may be considered clinically significant but failed to reach statistical significance, a dichotomy signaled elsewhere (Lansford et al., 2019). An example is speaker MSF20's increased transcription score of 9.25 rau of the Hybrid-E variant, which was found to be nonsignificant and to have a small effect size but passed the threshold by some considered as clinically significant change of 5%–8% for sentence-level intelligibility (Stipancic et al., 2016; Yorkston & Beukelman, 1984).

In addition, while the current residual-excited LPC analysis and resynthesis approach is known for producing

high-quality speech stimuli (Kain et al., 2008), it cannot be ruled out that the signal quality of the hybridized stimuli may have influenced perceptual outcomes. This also complicates statements about the detrimental effects of acoustic variables on intelligibility. For example, intelligibility scores for prosody, that is, the Hybrid-IED variant, were lower than habitual productions for the High group and the slow production for the Low group. It is therefore difficult to establish whether hybridizing prosodic characteristics across rate conditions is affecting intelligibility relatively more than its separate constituents of duration, intonation, and loudness or whether an overall reduced signal quality of the Hybrid-IED variants has contributed to the current findings. However, since it was possible to create hybrids for both speaker groups that were more intelligible compared to their respective least intelligible original production, the potential overall reduction in signal quality of the hybrids is likely limited.

Because intelligibility of speakers in this study was largely preserved in quiet listening conditions, the stimuli were mixed with multitalker babble to prevent transcription ceiling effects. Conflicting findings have been reported regarding the addition of noise to intelligibility in dysarthria. Chiu and Forrest (2018) reported a significant intelligibility decrease in noise for speakers with dysarthria and PD compared to neurotypical speakers, while Yoho and Borrie (2018) reported the decrease in intelligibility due to increasing levels of noise to be similar for dysarthric and neurologically healthy speech. As speech produced by neurotypical speakers was not included, it is unclear to what extent adding multitalker babble to disordered speech might have negatively affected intelligibility relative to neurotypical speech. However, the discrepancy between noise-added crowdsourced transcription scores, on one hand, and the higher laboratory-sourced transcription and SIT scores judged in optimal circumstances, on the other hand, indicates that the latter listening conditions are perhaps not entirely representative of the wide range of communication environments encountered by speakers on a regular basis. Although the presently employed noise levels of 0 dB SNR are not entirely representative of real-world SNRs (Weisser & Buchholz, 2019; Wu et al., 2018), the striking intelligibility gap between quiet and noisy listening conditions reported here and elsewhere (e.g., Chiu & Forrest, 2018) suggests that clinical or laboratory-sourced intelligibility judgments might not adequately capture intelligibility of speakers with dysarthria in more real-world communication environments. In addition, there is still much to be learned about potential additional interaction effects of noise with resynthesized and degraded speech. It is possible that added noise to resynthesized speech may have further detrimental effects on intelligibility. Future research is needed to elucidate these potential interaction effects. It should also be noted that evaluating the intelligibility of mild dysarthria with added noise is different from assessing intelligibility of moderate–severe dysarthria. Other studies suggest that transcription intelligibility scores across different levels of dysarthria severity do not accurately reflect listener comprehension, and added noise

might weaken the relationship between intelligibility and comprehension proficiency (Fontan et al., 2015; Hustad, 2008). One speaker (i.e., MSF03) displayed elevated pure-tone thresholds in one ear. While unlikely, hearing status may have influenced rate-related speech changes for this individual. Further research could examine the role of hearing loss in the effective realization of acoustic changes following behavioral treatment. In addition, the wide age range of participating listeners might have impacted on the mean and variability of transcription scores, as elderly listeners have been found to show increased difficulty understanding speech in noise (Helfer & Freyman, 2008). However, the majority (87%) of listeners were younger than 50 years of age, suggesting that listener age played a minimal role in the current study.

The High and Low speaker groups included different proportions of males and females. Two out of four speakers in the High group were male, while all speakers in the Low group were female. The effects of gender on acoustics have been a well-covered area of research, with studies describing notable differences in acoustics between male and female speakers. Notably, with respect to segmental articulation, female speakers tend to operate in a significantly larger formant working area compared to male speakers due to differences in vocal tract size and vocal fold mass as well as sociophonetic differences (Neel, 2008). However, except for F0 mean, the analysis of gender as covariate and the post hoc results of individual speaker data did not indicate differences between male and female speakers, indicating that gender-specific trends did not play a role in the current study.

Consonant imprecision was a perceptual feature present in the speech of two speakers. The segmental acoustic measures employed in this study were not suited to quantify this feature, potentially failing to capture certain speaker characteristics that may have played a role in explaining findings with respect to the Hybrid-DS variant. While the sentence-level acoustic measures used in this study are well established and representative of the nature of speech production adjustments hypothesized to be accompanying behavioral treatment techniques including slowed rate, such coarse-grained acoustic measures might not be adequately sensitive. More fine-grained acoustic measures such as degree of coarticulation or formant transition rate might have better captured production changes when slowing rate and could be pursued in future studies.

Furthermore, while the current study identified both segmental and suprasegmental properties as sources of improved intelligibility during slowed speech, individuals within each group varied substantially with respect to the potential acoustic sources that may contribute to the intelligibility gain for slowed rate. These results may indicate that speakers make individualized speech production adjustments when slowing rate. This strengthens the observation that the management of dysarthria, even within neurological diagnoses such as MS, requires an individualized and tailored treatment approach. The current study indicated that hybridization may be a viable technique to systematically manipulate and subsequently identify acoustic variables explaining intelligibility gains and losses during

interventions focusing on adjusting segmental and suprasegmental aspects of dysarthric speech in MS, with a major finding that durational changes alone were not sufficient to explain improved intelligibility for slowed rate.

Acknowledgments

This study was funded by the National Institutes of Health (Grant NIH-NIDCD R01DC004689 awarded to Kris Tjaden). The authors gratefully thank all speaker and listener participants as well as Amanda Allen, Tuan Dinh, Nicole Feeley, Grace Galvin, Rebecca Jaffe, Hannah Koellner, Laura Saitta, Sara Silverman, and Noora Somersalo for their help with stimuli preparation, data collection, and analysis. Preliminary results of this study were presented at the 2019 International Congress of Phonetic Sciences and the 2020 Biennial Conference on Motor Speech.

References

- Binns, C., & Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *The Journal of the Acoustical Society of America*, 122(3), 1765–1776. <https://doi.org/10.1121/1.2751394>
- Blanchet, P. G., & Snyder, G. J. (2009). Speech rate deficits in individuals with Parkinson's disease: A review of the literature. *Journal of Medical Speech-Language Pathology*, 17(1), 1–7.
- Blanchet, P. G., & Snyder, G. J. (2010). Speech rate treatments for individuals with dysarthria: A tutorial. *Perceptual and Motor Skills*, 110(3), 965–982. <https://doi.org/10.2466/pms.110.3.965-982>
- Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (Version 6.1.09) [Computer software]. Phonetic Sciences, University of Amsterdam. <http://www.praat.org>
- Bunton, K., Kent, R. D., Kent, J. F., & Duffy, J. R. (2001). The effects of flattening fundamental frequency contours on sentence intelligibility in speakers with dysarthria. *Clinical Linguistics & Phonetics*, 15(3), 181–193. <https://doi.org/10.1080/02699200010003378>
- Bunton, K., & Weismer, G. (2001). The relationship between perception and acoustics for a high-low vowel contrast produced by speakers with dysarthria. *Journal of Speech, Language, and Hearing Research*, 44(6), 1215–1228. [https://doi.org/10.1044/1092-4388\(2001/095\)](https://doi.org/10.1044/1092-4388(2001/095))
- Byun, T. M., Halpin, P. F., & Szeredi, D. (2015). Online crowd-sourcing for efficient rating of speech: A validation study. *Journal of Communication Disorders*, 53, 70–83. <https://doi.org/10.1016/j.jcomdis.2014.11.003>
- Cannito, M. P., Suiter, D. M., Beverly, D., Chorna, L., Wolf, T., & Pfeiffer, R. M. (2011). Sentence intelligibility before and after voice treatment in speakers with idiopathic Parkinson's disease. *Journal of Voice*, 26(2), 214–219. <https://doi.org/10.1016/j.jvoice.2011.08.014>
- Chiu, Y.-F., & Forrest, K. (2018). The impact of lexical characteristics and noise on intelligibility of Parkinsonian speech. *Journal of Speech, Language, and Hearing Research*, 61(4), 837–846. https://doi.org/10.1044/2017_JSLHR-S-17-0205
- Chiu, Y.-F., & Neel, A. (2020). Predicting intelligibility deficits in Parkinson's disease with perceptual speech ratings. *Journal of Speech, Language, and Hearing Research*, 63(2), 433–443. https://doi.org/10.1044/2019_JSLHR-19-00134
- De Bodt, M., Huici, M., & Van De Heyning, P. (2002). Intelligibility as a linear combination of dimensions in dysarthric speech. *Journal of Communication Disorders*, 35(3), 283–292. [https://doi.org/10.1016/S0021-9924\(02\)00065-5](https://doi.org/10.1016/S0021-9924(02)00065-5)
- Duffy, J. R. (2019). *Motor speech disorders: Substrates, differential diagnosis, and management* (4th ed.). Elsevier.
- Eijk, L., Fletcher, A., McAuliffe, M., & Janse, E. (2020). The effects of word frequency and word probability on speech rhythm in dysarthria. *Journal of Speech, Language, and Hearing Research*, 63(9), 2833–2845. https://doi.org/10.1044/2020_JSLHR-19-00389
- 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(11), 857–878. <https://doi.org/10.3109/02699206.2014.921839>
- 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. https://doi.org/10.1044/2016_JSLHR-S-15-0355
- Fletcher, A. R., McAuliffe, M. J., Lansford, K. L., Sinex, D. G., & Liss, J. M. (2017). Predicting intelligibility gains in individuals with dysarthria from baseline speech features. *Journal of Speech, Language, and Hearing Research*, 60(11), 3043–3057. https://doi.org/10.1044/2016_JSLHR-S-16-0218
- Fletcher, A. R., Risi, R. A., Wisler, A., & McAuliffe, M. J. (2019). Examining listener reaction time in the perceptual assessment of dysarthria. *Folia Phoniatrica et Logopaedica*, 71(5–6), 297–308. <https://doi.org/10.1159/000499752>
- Fontan, L., Tardieu, J., Gaillard, P., Woisard, V., & Ruiz, R. (2015). Relationship between speech intelligibility and speech comprehension in babble noise. *Journal of Speech, Language, and Hearing Research*, 58(3), 977–986. https://doi.org/10.1044/2015_JSLHR-H-13-0335
- Hammen, V. L., Yorkston, K. M., & Minifie, F. D. (1994). Effects of temporal alterations on speech intelligibility in Parkinsonian dysarthria. *Journal of Speech and Hearing Research*, 37(2), 244–253. <https://doi.org/10.1044/jshr.3702.244>
- Hardcastle, W. J., & Tjaden, K. (2008). Coarticulation and speech impairment. In M. J. Ball, M. R. Perkins, N. Muller, & S. Howard (Eds.), *The handbook of clinical linguistics* (pp. 506–524). Blackwell. <https://doi.org/10.1002/9781444301007.ch32>
- Helfer, K. S., & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and Hearing*, 29(1), 87–98. <https://doi.org/10.1097/AUD.0b013e31815d638b>
- Hertrich, I., & Ackermann, H. (1998). Auditory perceptual evaluation of rhythm-manipulated and resynthesized sentence utterances obtained from cerebellar patients and normal speakers: A preliminary report. *Clinical Linguistics & Phonetics*, 12(6), 427–437. <https://doi.org/10.3109/02699209808985236>
- Hustad, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language, and Hearing Research*, 51(3), 562–573. [https://doi.org/10.1044/1092-4388\(2008/040\)](https://doi.org/10.1044/1092-4388(2008/040))
- Institute of Electrical and Electronics Engineers. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, 17(3), 225–246. <https://doi.org/10.1109/TAU.1969.1162058>
- Kain, A., Amano-Kusumoto, A., & Hosom, J.-P. (2008). Hybridizing conversational and clear speech to determine the degree of contribution of acoustic features to intelligibility. *The Journal of the Acoustical Society of America*, 124(4), 2308–2319. <https://doi.org/10.1121/1.2967844>
- Kain, A., Hosom, J.-P., Niu, X., van Santen, J. P. H., Fried-Oken, M., & Staehely, J. (2007). Improving the intelligibility of dysarthric speech. *Speech Communication*, 49(9), 743–759. <https://doi.org/10.1016/j.specom.2007.05.001>

- Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C.** (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54*(4), 482–499. <https://doi.org/10.1044/jshd.5404.482>
- Kim, Y., Kent, R. D., & Weismer, G.** (2011). An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. *Journal of Speech, Language, and Hearing Research, 54*(2), 417–429. [https://doi.org/10.1044/1092-4388\(2010/10-0020\)](https://doi.org/10.1044/1092-4388(2010/10-0020))
- Krause, J. C., & Braida, L. D.** (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *The Journal of the Acoustical Society of America, 112*(5), 2165–2172. <https://doi.org/10.1121/1.1509432>
- Krause, J. C., & Braida, L. D.** (2009). Evaluating the role of spectral and envelope characteristics in the intelligibility advantage of clear speech. *The Journal of the Acoustical Society of America, 125*(5), 3346–3357. <https://doi.org/10.1121/1.3097491>
- Kuo, C., & Tjaden, K.** (2016). Acoustic variation during passage reading for speakers with dysarthria and healthy controls. *Journal of Communication Disorders, 62*, 30–44. <https://doi.org/10.1016/j.jcomdis.2016.05.003>
- Kuo, C., Tjaden, K., & Sussman, J. E.** (2014). Acoustic and perceptual correlates of faster-than-habitual speech produced by speakers with Parkinson's disease and multiple sclerosis. *Journal of Communication Disorders, 52*, 156–169. <https://doi.org/10.1016/j.jcomdis.2014.09.002>
- Lansford, K. L., Borrie, S. A., & Barrett, T. S.** (2019). Regularity matters: Unpredictable speech degradation inhibits adaptation to dysarthric speech. *Journal of Speech, Language, and Hearing Research, 62*(12), 4282–4290. https://doi.org/10.1044/2019_JSLHR-19-00055
- Lansford, K. L., Borrie, S. A., & Bystricky, L.** (2016). Use of crowdsourcing to assess the ecological validity of perceptual-training paradigms in dysarthria. *American Journal of Speech-Language Pathology, 25*(2), 233–239. https://doi.org/10.1044/2015_AJSLP-15-0059
- Lansford, K. L., & Liss, J. M.** (2014). Vowel acoustics in dysarthria: Mapping to perception. *Journal of Speech, Language, and Hearing Research, 57*(1), 68–80. [https://doi.org/10.1044/1092-4388\(2013/12-0263\)](https://doi.org/10.1044/1092-4388(2013/12-0263))
- Lansford, K. L., Liss, J. M., Caviness, J. N., & Utianski, R. L.** (2011). A cognitive-perceptual approach to conceptualizing speech intelligibility deficits and remediation practice in hypokinetic dysarthria. *Parkinson's Disease, 2011*, Article 150962. <https://doi.org/10.4061/2011/150962>
- Lim, S.-J., Shinn-Cunningham, B. G., & Perrachione, T. K.** (2019). Effects of talker continuity and speech rate on auditory working memory. *Attention, Perception, & Psychophysics, 81*(4), 1167–1177. <https://doi.org/10.3758/s13414-019-01684-w>
- Liss, J. M.** (2007). The role of speech perception in motor speech disorders. In G. Weismer (Ed.), *Motor speech disorders: Essays for Ray Kent* (pp. 187–219). Plural.
- Liu, S., & Zeng, F.-G.** (2006). Temporal properties in clear speech perception. *The Journal of the Acoustical Society of America, 120*(1), 424–432. <https://doi.org/10.1121/1.2208427>
- Lowell, S. Y., Colton, R. H., Kelley, R. T., & Hahn, Y. C.** (2011). Spectral- and cepstral-based measures during continuous speech: Capacity to distinguish dysphonia and consistency within a speaker. *Journal of Voice, 25*(5), e223–e232. <https://doi.org/10.1016/j.jvoice.2010.06.007>
- Lowit, A., Kuschmann, A., MacLeod, J. M., Schaeffler, F., & Mennen, I.** (2010). Sentence stress in ataxic dysarthria: A perceptual and acoustic study. *Journal of Medical Speech-Language Pathology, 18*(4), 77–82.
- 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. https://doi.org/10.1044/2016_AJSLP-15-0182
- McAuliffe, M. J., Kerr, S. E., Gibson, E. M. R., Anderson, T., & LaShell, P. J.** (2014). Cognitive-perceptual examination of remediation approaches to hypokinetic dysarthria. *Journal of Speech, Language, and Hearing Research, 57*(4), 1268–1283. https://doi.org/10.1044/2014_JSLHR-S-12-0349
- McHenry, M. A.** (2003). The effect of pacing strategies on the variability of speech movement sequences in dysarthria. *Journal of Speech, Language, and Hearing Research, 46*(3), 702–710. [https://doi.org/10.1044/1092-4388\(2003/055\)](https://doi.org/10.1044/1092-4388(2003/055))
- McRae, P. A., Tjaden, K., & Schoonings, B.** (2002). Acoustic and perceptual consequences of articulatory rate change in Parkinson disease. *Journal of Speech, Language, and Hearing Research, 45*(1), 35–50. [https://doi.org/10.1044/1092-4388\(2002/003\)](https://doi.org/10.1044/1092-4388(2002/003))
- Mefferd, A. S.** (2015). Articulatory-to-acoustic relations in talkers with dysarthria: A first analysis. *Journal of Speech, Language, and Hearing Research, 58*(3), 576–589. https://doi.org/10.1044/2015_JSLHR-S-14-0188
- Mefferd, A. S.** (2019). Effects of speaking rate, loudness, and clarity modifications on kinematic endpoint variability. *Clinical Linguistics & Phonetics, 33*(6), 570–585. <https://doi.org/10.1080/02699206.2019.1566401>
- Milenkovic, P.** (2005). *TF32*. University of Wisconsin–Madison.
- Miller, S. E., Schlauch, R. S., & Watson, P. J.** (2010). The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. *The Journal of the Acoustical Society of America, 128*(1), 435–443. <https://doi.org/10.1121/1.3397384>
- Molloy, D. W.** (1999). *Standardized Mini-Mental State Examination*. New Grange Press.
- Neel, A. T.** (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research, 51*(3), 574–585. [https://doi.org/10.1044/1092-4388\(2008/041\)](https://doi.org/10.1044/1092-4388(2008/041))
- 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*(4), 1021–1033. [https://doi.org/10.1044/1092-4388\(2008/08-0119\)](https://doi.org/10.1044/1092-4388(2008/08-0119))
- Nightingale, C., Swartz, M., Ramig, L. O., & McAllister, T.** (2020). Using crowdsourced listeners' ratings to measure speech changes in hypokinetic dysarthria: A proof-of-concept study. *American Journal of Speech-Language Pathology, 29*(2), 873–882. https://doi.org/10.1044/2019_AJSLP-19-00162
- R Core Team.** (2019). *R: A language and environment for statistical computing* (Version 3.6.2) [Computer software]. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Riesgo, C. A. F., & Nöth, E.** (2019). What makes the cepstral peak prominence different to other acoustic correlates of vocal quality? *Journal of Voice, 34*(5), 806.e1–806.e6. <https://doi.org/10.1016/j.jvoice.2019.01.004>
- Rudzicz, F.** (2013). Adjusting dysarthric speech signals to be more intelligible. *Computer Speech & Language, 27*(6), 1163–1177. <https://doi.org/10.1016/j.csl.2012.11.001>
- 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*(4), 387–399. [https://doi.org/10.1044/1058-0360\(2003/085\)](https://doi.org/10.1044/1058-0360(2003/085))
- Spitzer, S. M., Liss, J. M., & Mattys, S. L.** (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *The*

- Journal of the Acoustical Society of America*, 122(6), 3678–3687. <https://doi.org/10.1121/1.2801545>
- Stipancic, K. L., Tjaden, K., & Wilding, G.** (2016). Comparison of intelligibility measures for adults with Parkinson's disease, adults with multiple sclerosis, and healthy controls. *Journal of Speech, Language, and Hearing Research*, 59(2), 230–238. https://doi.org/10.1044/2015_JSLHR-S-15-0271
- Stipancic, K. L., Yunusova, Y., Berry, J. D., & Green, J. R.** (2018). Minimally detectable change and minimal clinically important difference of a decline in sentence intelligibility and speaking rate for individuals with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research*, 61(11), 2757–2771. https://doi.org/10.1044/2018_JSLHR-S-17-0366
- Studebaker, G. A.** (1985). A “rationalized” arcsine transform. *Journal of Speech and Hearing Research*, 28(3), 455–462. <https://doi.org/10.1044/jshr.2803.455>
- 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(4), 1208–1219. [https://doi.org/10.1044/1092-4388\(2011/11-0048\)](https://doi.org/10.1044/1092-4388(2011/11-0048))
- Taylor, P., Black, A. W., & Caley, R.** (1998). The architecture of the festival speech synthesis system. In *The Third ESCA/COCOSDA Workshop (ETRW) on Speech Synthesis*. ISCA Archive.
- Tjaden, K., Kain, A., & Lam, J.** (2014). Hybridizing conversational and clear speech to investigate the source of increased intelligibility in speakers with Parkinson's disease. *Journal of Speech, Language, and Hearing Research*, 57(4), 1191–1205. https://doi.org/10.1044/2014_JSLHR-S-13-0086
- Tjaden, K., Lam, J., & Wilding, G.** (2013). Vowel acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions. *Journal of Speech, Language, and Hearing Research*, 56(5), 1485–1502. [https://doi.org/10.1044/1092-4388\(2013/12-0259\)](https://doi.org/10.1044/1092-4388(2013/12-0259))
- Tjaden, K., Sussman, J. E., & Wilding, G. E.** (2014). Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in Parkinson's disease and multiple sclerosis. *Journal of Speech, Language, and Hearing Research*, 57(3), 779–792. https://doi.org/10.1044/2014_JSLHR-S-12-0372
- Tjaden, K., & Wilding, G.** (2011). The impact of rate reduction and increased loudness on fundamental frequency characteristics in dysarthria. *Folia Phoniatrica et Logopaedica*, 63(4), 178–186. <https://doi.org/10.1159/000316315>
- Tjaden, K., & Wilding, G. E.** (2004). Rate and loudness manipulations in dysarthria. *Journal of Speech, Language, and Hearing Research*, 47(4), 766–783. [https://doi.org/10.1044/1092-4388\(2004/058\)](https://doi.org/10.1044/1092-4388(2004/058))
- Turner, G., Martin, H., & de Jonge, R.** (2008). Influence of amplification on word intelligibility and phonetic contrast errors in dysarthria. *Journal of Medical Speech-Language Pathology*, 16(4), 267–274.
- Turner, G., 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(5), 1001–1013. <https://doi.org/10.1044/jshr.3805.1001>
- Utianski, R. L., Lansford, K. L., Liss, J. M., & Azuma, T.** (2011). The effects of topic knowledge on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *Journal of Medical Speech-Language Pathology*, 19(4), 25–36. <https://doi.org/10.4061/2011/150962>
- Van Nuffelen, G., De Bodt, M., Vanderwegen, J., Van de Heyning, P., & Wuyts, F.** (2010). Effect of rate control on speech production and intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica*, 62(3), 110–119. <https://doi.org/10.1159/000287209>
- Van Nuffelen, G., De Bodt, M., Wuyts, F., & Van de Heyning, P.** (2009). The effect of rate control on speech rate and intelligibility of dysarthric speech. *Folia Phoniatrica et Logopaedica*, 61(2), 69–75. <https://doi.org/10.1159/000208805>
- Walsh, B., & Smith, A.** (2012). Basic parameters of articulatory movements and acoustics in individuals with Parkinson's disease. *Movement Disorders*, 27(7), 843–850. <https://doi.org/10.1002/mds.24888>
- Weismer, G.** (2008). Speech Intelligibility. In M. J. Ball, M. R. Perkins, N. Muller, & S. Howard (Eds.), *The handbook of clinical linguistics* (pp. 568–582). Blackwell. <https://doi.org/10.1002/9781444301007.ch35>
- Weismer, G., Laures, J. S., Jeng, J.-Y., Kent, R. D., & Kent, J. F.** (2000). Effect of speaking rate manipulations on acoustic and perceptual aspects of the dysarthria in amyotrophic lateral sclerosis. *Folia Phoniatrica et Logopaedica*, 52(5), 201–219. <https://doi.org/10.1159/000021536>
- Weisser, A., & Buchholz, J. M.** (2019). Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions. *The Journal of the Acoustical Society of America*, 145(1), 349–360. <https://doi.org/10.1121/1.5087567>
- Wu, Y.-H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., & Oleson, J.** (2018). Characteristics of real-world signal-to-noise ratios and speech listening situations of older adults with mild-to-moderate hearing loss. *Ear and Hearing*, 39(2), 293–304. <https://doi.org/10.1097/AUD.0000000000000486>
- Yoho, S. E., & Borrie, S. A.** (2018). Combining degradations: The effect of background noise on intelligibility of disordered speech. *The Journal of the Acoustical Society of America*, 143(1), 281–286. <https://doi.org/10.1121/1.5021254>
- Yorkston, K. M.** (2007). The degenerative dysarthrias: A window into critical clinical and research issues. *Folia Phoniatrica et Logopaedica*, 59(3), 107–117. <https://doi.org/10.1159/000101769>
- Yorkston, K. M., & Beukelman, D. R.** (1984). *Assessment of intelligibility of dysarthric speech*. Pro-Ed.
- Yorkston, K. M., Beukelman, D. R., Strand, E., & Hakel, M.** (2010). *Management of motor speech disorders in children and adults* (3rd ed.). Pro-Ed.
- Yorkston, K. M., Beukelman, D. R., & Tice, R.** (1996). *Sentence Intelligibility Test*. Tice Technology Services.
- Yorkston, K. M., Hakel, M., Beukelman, D. R., & Fager, S.** (2007). Evidence for effectiveness of treatment of loudness, rate, or prosody in dysarthria: A systematic review. *Journal of Medical Speech-Language Pathology*, 15(2), 11–36.
- Yunusova, Y., Weismer, G., Kent, R. D., & Rusche, N. M.** (2005). Breath-group intelligibility in dysarthria. *Journal of Speech, Language, and Hearing Research*, 48(6), 1294–1310. [https://doi.org/10.1044/1092-4388\(2005/090\)](https://doi.org/10.1044/1092-4388(2005/090))

Appendix

List of Speech Stimuli: Harvard Sentences With Key Words in Bold

1. **Glue** the **sheet** to the **dark blue background**
 2. The **box** was **thrown beside** the **parked truck**
 3. **Four hours** of **steady work** faced us
 4. The **hogs** were **fed chopped corn** and **garbage**
 5. The **soft cushion** broke the **man's fall**
 6. The **girl** at the **booth** sold **fifty bonds**
 7. She **blushed when** he **gave** her a **white orchid**
 8. **Note closely** the **size** of the **gas tank**
 9. The **square wooden crate** was **packed** to be **shipped**
 10. He **sent** the **figs**, but **kept** the **ripe cherries**
 11. A **cup** of **sugar** makes **sweet fudge**
 12. **Place** a **rosebush** near the **porch steps**
 13. A **saw** is a **tool used** for **making boards**
 14. The **dune rose** from the **edge** of the **water**
 15. The **ink stain** dried on the **finished page**
 16. The **harder** he **tried** the **less** he got **done**
 17. **Paste** can **cleanse** the **most dirty brass**
 18. The **ancient coin** was **quite dull** and **worn**
 19. The **tiny girl** took off her **hat**
 20. The **pot boiled**, but the **contents** failed to **jell**
 21. The **sofa cushion** is **red** and of **light weight**
 22. An **abrupt start** does **not win** the **prize**
 23. These **coins** will be **needed** to **pay** his **debt**
 24. **Hoist** the **load** to **your left shoulder**
 25. **Burn peat** after the **logs** give out
-