Hybridizing Conversational and Clear Speech to Investigate the Source of Increased Intelligibility in Speakers With Parkinson's Disease

Kris Tjaden[a], **Alexander Kain**[b], and **Jennifer Lam**[a]

[a]University at Buffalo, NY [b]Center for Spoken Language Understanding, Oregon Health and Science University, Portland

Abstract

Purpose—A speech analysis-resynthesis paradigm was used to investigate segmental and suprasegmental acoustic variables explaining intelligibility variation for 2 speakers with Parkinson's disease (PD).

Method—Sentences were read in conversational and clear styles. Acoustic characteristics from clear sentences were extracted and applied to conversational sentences, yielding 6 hybridized versions of sentences in which segment durations, short-term spectrum, energy characteristics, or fundamental frequency characteristics for clear productions were applied individually or in combination to conversational productions. Listeners ($N = 20$) judged intelligibility in transcription and scaling tasks.

Results—Intelligibility increases above conversation were more robust for transcription, but the pattern of intelligibility improvement was similar across tasks. For 1 speaker, hybridization involving only clear energy characteristics yielded an 8.7% improvement in transcription intelligibility above conversation. For the other speaker, hybridization involving clear spectrum yielded an 18% intelligibility improvement, whereas hybridization involving both clear spectrum and duration yielded a 13.4% improvement.

Conclusions—Not all production changes accompanying clear speech explain its improved intelligibility. Suprasegmental adjustments contributed to intelligibility improvements when segmental adjustments, as inferred from vowel space area, were not robust. Hybridization can be used to identify acoustic variables explaining intelligibility variation in mild dysarthria secondary to PD.

Keywords

intelligibility; acoustics; dysarthria; clear speech

Intelligibility, defined as the extent to which a speaker's utterance is understood by a listener, has long been of interest in dysarthria. As discussed in a variety of sources,

---

Correspondence to Kris Tjaden: tjaden@buffalo.edu.

linguistic variables as well as speaker and listener-related factors are known to impact the extent to which a listener recovers a speaker's intended message (Hustad, 2007, 2008; Liss, 2007; Weismer, 2008; Yorkston, Beukelman, Strand, & Hakel, 2010). In the current study, we focused on the contribution of speaker-related factors to intelligibility, as reflected in acoustic measures of segmental articulatory behavior, respiratory-phonatory behavior, and prosody.

For some time, it has been recognized that numerical indices of intelligibility for speakers with dysarthria, either in the form of percent correct scores or scaled estimates of intelligibility, can reflect varied production deficits. Thus, two speakers might score 85% on the Sentence Intelligibility Test (Yorkston, Beukelman, & Tice, 1996), but the reduced intelligibility for Speaker A is explained by impaired segmental articulation, whereas the reduced intelligibility for Speaker B is attributable to impaired respiratory-phonatory behavior and prosody. To advance theoretical understanding of intelligibility, investigators therefore have moved away from solely quantifying the degree to which intelligibility is reduced in dysarthria to trying to determine the specific components or factors contributing to intelligibility. This line of inquiry also is of clinical importance, as maximizing intelligibility is a common goal of dysarthria treatment. That is, to the extent that the most efficient treatments would focus on speech production variables causing or explaining reduced intelligibility, studies investigating the source of intelligibility variation in dysarthria ultimately may assist in guiding and optimizing management.

A common approach has been to obtain perceptual judgments of intelligibility for one or more speaker groups as well as speech acoustic measures likely to be sensitive to impairments in segmental articulation, voice, prosody, and so forth (e.g., H. Kim, Hasegawa-Johnson, Perlman, 2011; Y. Kim, Kent, & Weismer, 2011; Liu, Tsao, & Kuhl, 2005; Weismer, Jeng, Laures, Kent, & Kent, 2001). The strength of the association between acoustic measures and intelligibility is then quantified. Other researchers aimed to induce variation in speech production characteristics as well as intelligibility by having talkers voluntarily modify speech style (e.g., Hustad & Lee, 2008; Neel, 2009; Tjaden & Wilding, 2004; Turner, Tjaden, & Weismer, 1995). Regardless of the specific details of the approach, the data in these studies have been pooled across speakers before quantifying the linkage between acoustic measures and intelligibility. Speakers in these studies generally span a range of overall dysarthria severity or speech involvement, thereby complicating interpretation of acoustic measures as explanatory variables for intelligibility. That is, relationships between acoustic measures and intelligibility in studies including speakers with a wide range of dysarthria severity may emerge by virtue of both measures being related to a third, mediating variable—overall dysarthria severity (see discussion in Weismer et al., 2001; Yunusova, Weismer, Kent, & Rusche, 2005).

Yunusova et al. (2005) addressed this challenge by studying the relationship between speech acoustic measures and intelligibility within speakers with dysarthria. Because a given talker's overall dysarthria severity can be assumed to be constant during a single experimental recording session, studies employing a within-speaker design minimize the possibility that dysarthria severity is mediating the relationship between intelligibility and speech acoustic measures. Other studies have used an across-speaker design, but have

somewhat constrained the range of dysarthria severity (e.g., Yunusova et al., 2012). Regardless of whether an across-or within-speaker design is employed, however, these types of in vivo speech studies do not allow for strong statements concerning the independent contribution of acoustic variables to intelligibility.

Studies employing an analysis-resynthesis or digital signal–processing approach also have been used to investigate the acoustic basis of intelligibility. Analysis-resynthesis allows for precise control over one or more acoustic parameters, such as fundamental frequency (F0) or duration, while holding other parameters constant. In this manner, the independent contribution of particular acoustic variables or combinations of acoustic variables to intelligibility may be investigated. Such an approach has been widely applied to neurologically normal speech (e.g., Binns & Culling, 2007; Kain, Amano-Kusumoto, & Hosom, 2008; Krause & Braida, 2009; Laures & Weismer, 1999; Liu & Zeng, 2006; Miller, Schlauch, & Watson, 2010; Spitzer, Liss, & Mattys, 2007; Uchanski, Choi, Braida, Reed, & Durlach, 1996; Watson & Schlauch, 2008). As discussed in the following paragraph, several dysarthria studies also have used this approach (e.g., Bunton, Kent, Kent, & Duffy, 2001; Bunton, 2006; Hammen, Yorkston, & Minifie, 1994; Kain et al., 2007).

Bunton et al., (2001) used speech resynthesis to reduce sentence-level F0 range by 25%, 50%, or 100% for speakers with dysarthria secondary to Parkinson's disease (PD) or stroke, as well as a group of healthy talkers. The impact of a reduced sentence-level F0 range on intelligibility was assessed using transcription and scaling tasks. Consistent with a study by Laures and Weismer (1999) in which intelligibility for neurologically normal talkers decreased when sentence-level F0 was flattened using speech resynthesis, the overall trend was for intelligibility to decline as F0 range was reduced, with the strongest effect for the PD group. The authors concluded that sentence-level F0 variation is an important contributor to intelligibility but that the nature of the contribution may vary across dysarthrias or neurological diagnoses. Relatedly, Kain et al. (2007) used speech analysis-resynthesis to enhance vowel intelligibility for a speaker with dysarthria secondary to Friedrich's ataxia. Vowels in CVC words were analyzed, transformed, and resynthesized to more closely approximate vowels produced by a neurologically normal talker. Vowel duration as well as steady-state frequencies of the first (F1), second (F2), and third formants (F3) were the acoustic features that yielded the optimal dysarthria-to-neurologically normal transformation.

Finally, it has long been recognized that a clear speech style enhances intelligibility for neurotypical speakers, and clear speech is recommended as a therapeutic technique for improving intelligibility in dysarthria (Hustad & Weismer, 2007; Smiljani & Bradlow, 2009). Indeed, the term *clear speech benefit* is used to refer to the improved intelligibility of clear speech. Numerous studies, mostly focused on neurotypical speech, also have documented acoustic adjustments associated with a clear speech style. For example, relative to typical or conversational speech, clear speech tends to be characterized by enhanced segmental contrasts, lengthened speech durations, an increase in mean F0, as well as increased F0 variation, and increased intensity envelope modulation (see reviews in Krause & Braida, 2009; Smiljani & Bradlow, 2009). The mere presence of these types of speech acoustic changes cannot be taken as evidence that they are causally related to the improved

intelligibility of clear speech, however. In fact, the acoustic basis of the clear speech benefit has yet to be established, although this has been identified as a major topic of importance for speech research (Smiljani & Bradlow, 2009). In the case of dysarthria, such knowledge also would strengthen the scientific evidence base for therapeutic use of clear speech and may ultimately help to optimize clear speech training programs. Kain et al. (2008) developed a speech analysis-resynthesis technique termed *hybridization* as a method for determining whether individual acoustic parameters (i.e., sentence-level F0, sentence-level energy, segment durations, short-term spectrum) or combinations of these parameters explained the improved intelligibility of clear speech. To date, however, hybridization has only been applied to neurotypical speech.

In sum, although a fair amount of research has been devoted to the topic, the acoustic basis of intelligibility in dysarthria is not firmly established. Despite the fact that clear speech is used in the treatment of dysarthria, acoustic changes responsible for its improved intelligibility also remain poorly understood. Thus, the overarching aim of the current study was to determine acoustic variables explaining sentence-intelligibility variation associated with a clear speech style for two speakers with PD using the speech analysis-resynthesis approach developed by Kain et al. (2008). In this manner, we sought not only to advance scientific understanding of intelligibility in dysarthria but also to address a major goal of clear speech research.

Method

Speakers and Speech Characteristics

Two men diagnosed with idiopathic PD provided the speech materials. Similar speaker numbers have been used in previously published studies using speech resynthesis to investigate the acoustic basis of intelligibility (Kain et al., 2007, 2008; Laures & Weismer, 1999; Spitzer et al., 2007; Watson & Schlauch, 2008). Talkers 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 neurologic disease, had not received neurosurgical or dysarthria treatment for PD, and scored at least 26 on the Standardized Mini-Mental State Examination (Molloy, 1999). PDM01 was 76 years of age and 12 years postdiagnosis. PDM06 was 66 years of age and 3 years postdiagnosis.

Percent correct intelligibility from the Sentence Intelligibility Test (SIT; Yorkston et al., 1996), as judged by 10 inexperienced listeners, was used to document overall speech mechanism involvement. SIT scores were previously reported for a larger group of speakers, and we refer readers to this earlier study for methodological details concerning the SIT (Sussman & Tjaden, 2012). SIT scores for PDM01 and PDM06 were 87% and 91%, respectively. In addition to having mildly reduced sentence intelligibility, speakers were noted to have perceptual characteristics consistent with hypokinetic dysarthria in the form of reduced segmental precision, as well as a breathy, monotonous voice.

Experimental Speech Materials and Procedure

Speakers read 25 Harvard sentences (IEEE, 1969) in a sound-treated room. Sentences were selected from the larger corpus of Harvard psychoacoustic sentences to include multiple occurrences of various phonemes, including three to five occurrences of four peripheral vowels (i.e.,/i/,/ɑ/,/æ/,/u/) and four nonperipheral vowels (i.e.,/I/,/ɛ/,/ʊ/,/ʌ/) for use in calculating vowel space areas. Harvard sentences included imperatives and declaratives that were syntactically and semantically normal. Each sentence ranged in length from seven to nine words and contained five key words (i.e., *Paste* can *cleanse* the *most dirty brass*). The acoustic signal was transduced using an AKG C410 head-mounted microphone positioned 10 cm and 45 to 50 degrees from the left oral angle. The signal was preamplified, low-pass filtered at 9.8 kHz, and digitized to computer hard disk at a sampling rate of 22 kHz using TF32 (Milenkovic, 2005).

The sentences were first produced in a conversational (i.e., habitual or typical speech) speech style followed by several other speech styles, including *clear*. For the clear style, speakers were provided with the instruction "how you might talk to someone in a noisy environment or with a person who has a hearing loss. Exaggerate the movements of your mouth. You also may be louder and slower than usual. If your regular speech corresponds to a clearness of 100, you should aim for a clearness twice as good or a clearness of 200." These instructions were modeled after other clear speech studies (Smiljani & Bradlow, 2009). For each speaker, a unique ordering of the 25 sentences was recorded in each speech style. For the purposes of the current study, a subset of the same 10 sentences produced in both the conversational and clear styles was of interest for each speaker. Stimuli numbers were selected to be consistent with previous studies using resynthesis to investigate the acoustic basis of sentence intelligibility (Bunton et al., 2001; Laures & Weismer, 1999). To minimize the influence of stimulus or linguistic familiarity on judgments of intelligibility, seven of 10 sentences differed for the two speakers. The remaining three sentences were identical to allow for cross-speaker comparisons for sentences with identical linguistic content. For each speaker, a 1000 Hz calibration tone also was recorded for the purpose of obtaining offline measures of vocal intensity from the acoustic signal.

Measures of vowel space area, articulation rate, F0, and vocal intensity were obtained for speakers with PD, as well as a group of eight healthy males who were part of a previous acoustic study (Tjaden, Lam, & Wilding, 2013). Acoustic measures further documented the nature of the speech impairment for the talkers with PD as well as the nature of their conversational-to-clear speech acoustic adjustments, with attention to segmental and suprasegmental adjustments hypothesized in other studies to explain the improved intelligibility of clear speech. Measures were not intended to comprehensively document conversational-to-clear production adjustments but were selected to be representative of the kinds of production adjustments hypothesized to underlie the clear speech benefit. That is, enhanced segmental contrast, as evidenced by an expanded vowel space area, greater F0 range, and greater intensity envelope modulation have been hypothesized to contribute to the clear speech benefit. Lengthened speech durations also are commonly reported in clear speech, and the combination of duration and short-term spectral changes appear to be important to the clear speech benefit (Kain et al., 2008).

Procedures for obtaining acoustic measures of vowel space area, articulation rate, and SPL are described in Tjaden et al. (2013). Briefly, using TF32 (Milenkovic, 2005), articulation rate was determined by segmenting sentences into runs, defined as a stretch of speech bounded by silent periods between words of at least 200 ms (Turner et al., 1995). Conventional acoustic criteria were used to identify run onsets and offsets from the combined waveform and wideband (300–400 Hz) spectrographic displays. Articulatory rate, in syllables per second, was determined for each run by counting the number of syllables produced and dividing by run duration. Mean sound pressure level (SPL) of each speech run as well as SPL *SD* were obtained from the root-mean-square (RMS) intensity trace of TF32 (Milenkovic, 2005). Mean RMS voltages were exported to Excel and converted to dB SPL with reference to each talker's calibration tone.

Linear-predictive-coding-derived formant trajectories were computed for F1 and F2 across the entire duration of vowels. Vowel onsets and offsets were identified using conventional acoustic criteria. Formant traces generated by TF32 (Milenkovic, 2005) were hand-corrected for computer-generated tracking errors. Formant frequency values for F1 and F2 were extracted at 50% of vowel duration. For each speaker, condition, and vowel, formant frequency values were averaged across tokens for use in calculating a peripheral and non-peripheral vowel space area. Vowel space areas were calculated from midpoint F1 and F2 values for vowel tokens occurring in the larger corpus of 25 Harvard sentences because sentence subsets used for hybridization did not contain sufficient tokens of all vowels for vowel space area computation.

Finally, sentence-level F0 was obtained using Praat (Boersma & Weenink, 2013). Individual glottal pulses were identified by Praat and any computer-generated errors were manually corrected (e.g., doubling or halving of pitch period, failure to identify periods of voicing). For each sentence, F0 mean, minimum, and maximum were extracted and averaged across sentences. F0 range was calculated as the difference between the average F0 minimum and average F0 maximum. As described in the following section, "pitch pulses" or glottal closure intervals also were used in the hybridization.

Hybridization Algorithm and Speech Resynthesis

Hybridization, as implemented in the present study, may be conceptualized as waveform resynthesis of a conversational sentence produced by a speaker while optionally imposing the energy envelope, F0 envelope, segment durations, or short-term spectra from the same sentence produced in a clear speech style by the same speaker. Thus, in contrast to resynthesis studies in which an acoustic parameter such as sentence-level F0 envelope is adjusted by a numerical constant for conversational productions, acoustic manipulations in hybridization represent those voluntarily produced by a speaker. The entire process is summarized in Figure 1. Readers are referred to Kain et al. (2008) for a comprehensive description of the hybridization process. First, for each sentence, the clear phonetic sequence was aligned with the conversational phonetic sequence because the two original waveforms may not be matched exactly in terms of phonetic content (Step 1). In parallelization, the hybrid phoneme sequence was operationally set to be the same as the clear phoneme sequence, and the conversational waveform was modified accordingly by implementing

phoneme deletions or insertions from the clear waveform using a time-domain cross-fade technique to avoid discontinuities (Step 2). Next, a speech analysis was carried out on the parallelized waveforms, consisting of first computing auxiliary marks used for prosodic modification (Step 3) and then extracting short-term spectra, energy trajectories, F0 trajectories, and segment durations (Step 4). Acoustic characteristics of clear speech were subsequently combined with complementary acoustic characteristics of conversational speech to form hybrids (Step 5), which were then used to synthesize six hybrid speech waveforms (Step 6). As indicated in Figure 1, for example, hybrid duration/spectrum specifies that that hybrid was composed of segment duration and short-term spectral characteristics for clear speech, whereas the energy and F0 time histories were taken from the conversational production. It should further be noted that hybridization first required identification of phonetic segments and individual glottal closure instances (see the previous section, where procedures for measuring F0 are described). Using Praat (Boersma & Weenink, 2013), segment boundaries were identified from the waveform and wideband (300–400 Hz) spectrographic displays using conventional acoustic criteria. Pauses were identified and labeled but were not modified during hybridization.

The six hybrids studied were 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). These hybrids were selected on the basis of previous clear speech studies, which suggests the importance of these acoustic cues to the clear speech benefit. Hybrids also allowed for conclusions concerning the importance of prosodic cues, either individually or in combination, versus segmental cues to intelligibility. The DS hybrid was included given Kain et al.'s (2008) study, which found that the combination of spectrum and duration was the primary source of improved intelligibility for clear speech produced by a neurotypical male. Although additional hybrids are possible, the number of hybrids was constrained to minimize the influence of stimulus familiarity on judgments of intelligibility.

Figures 2 and 3 show selected amplitude normalized hybrid stimuli (via root-mean-square A-weighted [RMSA]; see Kain et al., 2008), as well as the amplitude normalized conversational and clear productions for PDM01 and PDM06, respectively. The waveform display in Figure 2 corresponds to the conversational production. In each panel of Figure 2, RMS intensity is shown in gray and the F0 time history is shown in black. The aspect ratio of the *x*-axis for the clear production differs from other panels in this figure to facilitate comparison of energy and F0 traces from the clear production to the hybrids as well as the conversation production. The various panels in Figures 2 and 3 display extraction of acoustic features from clear sentences and their application to conversational sentences. Visual inspection of spectrograms in Figure 3 further attests to the excellent quality of the resynthesis.

Perceptual Method and Procedure

Listeners—A total of 20 individuals, including nine males and 11 females, with a mean age of 23 years (*SD* = 6 years), participated. Participants spoke standard American English as their first language and reported no history of speech, language, or hearing pathology or

experience with speech disorders. All listeners passed a bilateral hearing screening at 20 dB HL at octave frequencies between 500 Hz and 8000 Hz (American National Standards Institute, 2004) and were paid a modest fee.

Perceptual task—Speakers with PD had mildly reduced sentence intelligibility, as indexed by the SIT. Thus, to prevent ceiling effects, experimental stimuli were equated for overall amplitude and mixed with multitalker babble, as is typically done in clear speech studies (Smiljani & Bradlow, 2009). Sentences first were equated for average RMSA intensity. To accomplish this, speech waveforms were filtered with an A-weighted filter (International Electrotechnical Commission, 2002), and levels were calculated by averaging frame-by-frame (using frame durations on the order of 10 ms) RMS sample values of nonpausal portions of the speech. Each waveform was then multiplied by an appropriate gain factor so that the resulting waveforms all had the same average RMSA value. Sentences were mixed with multitalker babble sampled at 22 kHz and low-pass filtered at 11 kHz (Bochner, Garrison, Sussman, & Burkard, 2003; Frank & Craig, 1984). Based on pilot testing, a signal-to-noise ratio of −1 dB was applied to each sentence; this ratio minimized floor and ceiling effects in both intelligibility tasks. Sentences were presented via Sony Dynamic Stereo Headphones (MDR-V300) at 75 dB. The dB level of stimuli was calibrated at the beginning of each listening session for five randomly selected experimental sentences using an earphone coupler and Quest Electronics 1700 sound level meter.

Listeners were seated in a sound-treated booth in front of a computer. Following previous studies employing resynthesis to study the acoustic basis of sentence intelligibility, all listeners judged intelligibility in both scaling and orthographic transcription tasks (Bunton et al., 2001; Laures & Weismer, 1999). The two tasks were expected to yield related findings, but transcription intelligibility is thought to largely reflect segmental production characteristics, whereas scaling tasks may be more sensitive to nonsegmental variables (see Weismer, 2008; Weismer & Laures, 2002; Yunusova et al., 2005). Sentences were pooled across speakers, and orders of both tasks and stimuli were randomized and counterbalanced across listeners. Before beginning each task, listeners judged eight practice sentences produced by a male speaker with PD who was not part of the present study. In this manner, the practice task exposed listeners to each of the sentence variants and also familiarized participants with the computer interface.

Scaled estimates of intelligibility were obtained using a vertically oriented, computerized 150 mm vertical Visual Analog Scale, as used in previous studies (Sussman & Tjaden, 2012; Tjaden et al., in press). Endpoints were labeled *as cannot understand anything* and *understand everything*. The scale was continuous and contained no tick marks. Following presentation of each sentence, listeners used a computer mouse to click anywhere along the scale to indicate their judgment. The task was self-paced and listeners were allowed to hear each sentence once. Following completion of the entire experiment, computer software converted responses to numbers ranging from 0 (*cannot understand anything*) to 1 (*understand everything*).

For the transcription task, listeners also heard sentences once and typed their response onto a computer using custom software. Similar to the scaling task, the transcription task was self-

paced, and listeners heard each stimulus only once. Each of the five key words was scored as correct or incorrect. An exact match between the target word and response was required for a key word to be scored as correct. A key word scoring format has been used in other studies and acknowledges that information-bearing words make a more important contribution to utterance meaning than function words such as "a," which typically have little meaning (Bradlow, Toretta & Pisoni, 1996; Hustad, 2006; Kain et al., 2008; Miller et al., 2010).

As previously noted, the eight variants of each Harvard sentence included unmodified conversational (Con) and clear (Clr) productions as well as six hybrids. Thus, for each speaker, judgments of intelligibility were obtained for 80 sentences (10 Con + 10 Clr + 60 hybrids = 80). Each listener judged all conversational and clear sentences produced by both speakers. To reduce the number of exposures to a given sentence, we divided hybrid sentences into two subsets of 60. Hybrid subsets were carefully constructed to include five occurrences of each of the six hybrids for both speakers. Hybrid subsets were then pooled with the conversational and clear productions yielding two sentence sets composed of 100 sentences (Set A, Set B). Each listener judged one sentence set in the transcription task and the other sentence set in the scaling task. In this manner, all clear and conversational sentences were judged by 20 listeners; each hybrid sentence was judged by 10 listeners. In addition, listeners were presented with a random selection of 10 sentences twice for the purpose of determining intrajudge reliability.

Data Analysis

Acoustic measures were characterized descriptively. Intelligibility data for PDM01 and PDM06 were analyzed separately. For each listener, a percent correct score was computed for each of the eight sentence variants (i.e., Con, Clr, I, E, D, IED, DS, S) by pooling sentences, tallying the total number of correctly transcribed key words, dividing by the total number of key words, and multiplying by 100. Similarly, a mean scale value for each of the eight sentence variants was obtained for each listener by averaging scale values across sentences. Percent correct scores were arcsine transformed for statistical analysis, but raw scores are reported in figures and tables. Intelligibility measures for all sentence variants were characterized using standard descriptive statistics, such as means and standard deviations. Because acoustic cues contributing to conversational-to-clear improvements in intelligibility were of interest, statistical analyses consisted of planned comparisons in the form of paired *t* tests wherein intelligibility for conversational stimuli was compared with variants for which intelligibility improved above conversational levels. A significance level of .05 was used for all hypothesis testing.

Results

Acoustic Measures

Descriptive statistics for acoustic measures are reported in Table 1. Relative to the conversational style for control males, conversational speech for PDM06 was characterized by notable compression of both vowel space areas, a reduction in mean SPL, and reduced sentence-level F0 range. Mean SPL and F0 range for PDM01's conversational style also were reduced in comparison with control males, and his articulation rate was slightly

accelerated. These acoustic differences for the speakers with PD and control males are consistent with classic perceptual descriptions of "prosodic insufficiency" in the dysarthria of PD.

Table 1 further indicates that similar to the control males, both PD speakers slowed articulation rate, increased mean SPL and mean F0, and enhanced sentence-level SPL and F0 modulation for clear sentences. These conversational-to-clear adjustments are consistent with other studies of PD and studies of neurotypical speech (Dromey, 2000; Goberman & Elmer, 2004; Smiljanić & Bradlow, 2009). PDM06 and control males also expanded vowel space areas in the clear condition, whereas vowel space areas for PDM01 were slightly reduced for clear versus conversational. Although studies of normal speech consistently report an expanded vowel space area for clear speech (see Smiljanić & Bradlow, 2009), these types of segmental adjustments appear less consistently in the clear speech of PD speakers (Goberman & Elmer, 2004).

Listener Reliability

For the scaling task, Pearson product correlation coefficients for the first and second presentation of sentences ranged from .57 to .99 across the 20 listeners, with a mean of .80 (*SD* = .13). For transcription, Pearson product correlations relating the percentage of correctly transcribed words in the first and second presentation of sentences ranged from −.12 to 1.00 across the 20 listeners, with a mean of .76 (*SD* = .24). Removal of the outlier with poor intrajudge reliability yielded correlations for the 19 remaining listeners ranging from .58 to 1.0, with a mean of .80 (*SD* = .13). Transcription results were identical regardless of whether data for the listener with poor reliability were included or excluded. However, to be conservative, this listener's transcription data were excluded from the study. Findings for transcription reported in the remainder of the article are based on responses from 19 listeners.

Interjudge reliability was assessed using the intra-class correlation coefficient (ICC; see also Y. Kim & Kuo, 2012; Neel, 2009; Van Nuffelen, De Bodt, Wuyts, & Van de Heyning, 2009). Interjudge reliability was assessed separately for Set A and Set B stimuli, as sentence sets were composed of different hybrid variants. For the scaling task, the single measure ICC was .52 and average ICC was .92 for both Set A and Set B sentences. For Set A transcription, the single measure ICC was .54 and the average ICC was .92. For Set B transcription, the single measure ICC was .53 and the average ICC was .91. Thus, for both intelligibility tasks, listeners demonstrated good agreement on a per-item basis and excellent agreement on average (Cichetti, 1994; Fleiss, 1986). All ICCs were also statistically significant (*p* < .001). For both scaling and transcription, interjudge and intrajudge reliability metrics compare favorably with previously published intelligibility studies of dysarthria (e.g., Bunton et al., 2001; Y. Kim & Kuo, 2012; Neel, 2009; Van Nuffelen et al., 2009; Yunusova et al., 2005). Listener reliability also is considered further in the discussion.

Intelligibility

Figure 4 reports mean percent correct scores and scale values pooled across sentences and listeners. Vertical bars indicate ± 1 *SD*. Results for conversational and clear stimuli are

shown on the extremes of the *x*-axis, and results for hybrids are reported in the middle of the *x*-axes. Mean scale values for conversational and clear stimuli also are reported in Table 2, as are overall percent correct scores. These overall percent correct scores were obtained by pooling responses across listeners, tallying the total number of correctly transcribed key words, dividing by the total number of possible key words, and multiplying by 100. Difference scores for hybrids in Table 2 further capture the change in intelligibility. Negative difference scores indicate reduced intelligibility relative to conversation, whereas positive scores indicate improved intelligibility relative to conversation. Difference scores for transcription were calculated from overall percent correct scores; difference scores for the scaling task were calculated using scale values averaged across listeners.

Several observations can be made from Figure 4 and Table 2. First, although increases in intelligibility above conversation are more robust for transcription, for both speakers, the overall pattern of results with respect to improvements in intelligibility above conversation is similar for the two tasks. For example, clear stimuli were consistently judged to be the most intelligible. In addition, of the six hybrid variants, the E hybrid was associated with the greatest improvement in intelligibility above conversational levels for PDM01 in both tasks, whereas the S hybrid was associated with the greatest improvement in intelligibility in both tasks for PDM06, followed by the DS hybrid. For both intelligibility tasks, the magnitude of the improvements in intelligibility also was greater for PDM01.

Results for paired *t* tests comparing conversation to each hybrid variant associated with improved intelligibility were as follows: For PDM01, there was a significant improvement in intelligibility for Clr versus Con, transcription $t(18) = -6.119$, $p < .001$; scaling $t(19) = 4.468$, $p < .001$. The E-Con comparison for PDM01 was also significant for the transcription task, $t(18) = -2.149$, $p = .046$, but not the scaling task, $t(19) -1.783$, $p = .09$.

To supplement the main analysis for PDM01 reviewed in the preceding paragraph, in which sentences were pooled, we report selected individual sentence results for transcription in Figure 5. Individual sentences are identified with numbers and symbols. Percent correct scores were obtained by pooling responses across listeners, tallying the number of key words correctly transcribed, dividing by the total number of key words, and multiplying by 100. The main conclusion to be drawn from Figure 5 is that the E hybrid was associated with improved intelligibility relative to Con for eight of 10 sentences for PDM01. Moreover, comparison among the six hybrids indicated that the E hybrid was associated with the best intelligibility for all of PDM01's sentence productions.

For PDM06, intelligibility also improved for Clr versus Con stimuli: transcription $t(18) = 7.375$, $p < .001$; scaling $t(19) = -8.383$, $p < .001$; S versus Con: transcription $t(18) = 12.648$, $p < .001$; scaling $t(19) = -7.712$, $p < .001$; and DS versus Con: transcription $t(18) = -4.213$, $p = .001$; scaling $t(19) = -4.402$, $p < .001$. Although Table 2 indicates that the I and E hybrids for PDM06 were more intelligible than Con, paired comparisons involving these hybrids and Con were not significant ($p > .05$). Selected individual sentence results for PDM06 are shown in Figure 6. Percent correct scores were computed in the same manner as those in Figure 5. Symbol color indicates sentences for which the DS hybrid (white) or S-only hybrid (black) was associated with the best intelligibility among all hybrids relative to

the Con style. Figure 6 indicates that the S hybrid was associated with greatest gain in intelligibility above the Con style for six sentences, whereas the DS hybrid was associated with the greatest gain in intelligibility above Con style for three sentences. Results were idiosyncratic for the remaining sentences.

Discussion

Most previous studies using a digital signal–processing approach to investigate acoustic variables explaining intelligibility in dysarthria have focused on manipulation of a single acoustic parameter (e.g., Bunton, 2006; Bunton et al., 2001; Hammen et al., 1994). In contrast, we investigated the contribution of segmental as well as sentence-level, suprasegmental variables to intelligibility variation in dysarthria. Acoustic manipulations in many of these previous studies also involved operationally defined magnitudes of change (i.e., altering all F0 values by a percentage), which may not be representative of acoustic change in naturally occurring speech. In the current study, however, we examined voluntary, conversational-to-clear changes to the speech signal. There are several conclusions to be drawn from this work.

First, despite the fact that clear speech for both speakers was associated with a variety of acoustic changes, only some conversational-to-clear speech acoustic adjustments contributed to the clear speech benefit. Consider first the acoustic differences between the conversational and clear speech styles. As has been widely reported in clear speech studies of neurotypical talkers as well as Goberman and Elmer's (2004) clear speech study of PD, clear sentences were characterized by lengthened segment durations, as inferred from a reduced articulation rate. Although the magnitude of the rate reduction for the speakers with PD was somewhat less than for a group of age-matched control males, both PDM01 and PDM06 reduced articulation rate by approximately 20% when using a clear speech style. Both speakers also enhanced sentence-level F0 modulation and increased mean F0 when using a clear speech style. F0 range for PDM01 increased by a factor of approximately 1.6, whereas F0 range for PDM06 increased by a factor of 1.5. These findings for F0 also are in agreement with previous clear speech studies (e.g., Bradlow, Krause, & Hayes, 2003; Goberman & Elmer, 2004; Picheny, Durlach, & Braida, 1986). For both PDM01 and PDM06, however, neither the D nor I hybrids were associated with a significant improvement in intelligibility relative to Con, nor did the IED hybrid improve intelligibility above conversational levels.

Previous investigators of normal speech suggested that lengthening of segment durations alone is not sufficient to explain the clear speech intelligibility benefit (Krause & Braida, 2002; Picheny, Durlach, & Braida, 1989). Digitally lengthening segment durations in sentences produced by speakers with dysarthria also does not improve intelligibility above habitual levels (Hammen et al., 1994). It therefore is not unexpected that the D hybrid did not enhance intelligibility. However, the finding that the I hybrid also did not improve intelligibility above conversational levels seems at odds with Bunton's (2006) study reporting that an enhanced sentence-level F0 range improved vowel intelligibility in PD as well as studies suggesting a relationship between sentence-level F0 modulation and intelligibility (see reviews in Ramig, 1992; Watson & Schlauch, 2008). One possibility is

that listeners perceived sentences having an exaggerated F0 range to be unnatural, and a reduction in speech naturalness offset any improvements in intelligibility.

Alternatively, Miller et al. (2010) found that intelligibility for five healthy talkers was reduced, on average, by approximately 13% when F0 contours for conversational sentences were increased by a factor of 1.75. As Miller et al. (2010) note, their results should not be interpreted to suggest that exaggerated F0 contours have no beneficial perceptual effect, but that the relationship between an exaggerated F0 contour and intelligibility is complex. For example, although a clear speech style may be accompanied by an overall increase in sentence-level F0 range, F0 contrastivity between syllables—which is known to be important for lexical segmentation—might be reduced in clear speech. Studies are needed to further investigate the contribution of exaggerated sentence-level F0 to intelligibility for a variety of speaker and listener populations because suprasegmental cues may become increasingly important for recovering a speaker's intended message when segmental cues are not robust or are unreliable (see discussion in Liss, 2007).

Second, results indicated that the source of the clear speech benefit for one speaker was attributable to adjustments in short-term spectrum, whereas energy characteristics explained the clear speech benefit for the other speaker. This pattern of results further held for the majority of individual sentences (see Figures 5 and 6). The implication is that even for speakers with the same neurological diagnosis or dysarthria, as well as broadly similar speech involvement as indexed by the SIT (Yorkston et al., 1996), acoustic variables explaining improved intelligibility for clear speech may differ. Determining talker characteristics that help to predict acoustic variables explaining intelligibility variation requires further study. Similar to results for neurologically normal speech reported by Kain et al. (2008), the current results hint at the primacy of segmental articulatory variables to clear speech improvements in intelligibility. That is, although both speakers made a number of suprasegmental adjustments when instructed to use a clear speech style, PDM06 also enhanced segmental contrast, as inferred from measures of vowel space area (see Table 1). PDM06's hybrids involving the clear spectrum also were associated with a significant improvement in intelligibility above conversation. Why a clear speech style was not associated with increased vowel space areas for PDM01 is unknown. This might be related to the fact that PDM01's vowel space areas in Table 1 for conversation resembled those for control males. This could either reflect unimpaired segmental articulation or compensation for a slightly accelerated rate, reduced SPL, or reduced F0 range (for related discussion, see Goberman & Elmer, 2004). Nonetheless, the finding that energy characteristics explained the improved intelligibility of clear speech for PDM01 is in agreement with research suggesting that increased intensity envelope modulation contributes to the improved intelligibility of clear speech (Krause & Braida, 2004, 2009). Thus, suprasegmental adjustments do contribute to improved intelligibility of clear speech, perhaps especially when speakers do not enhance segmental contrasts.

The current findings are based on conversational and clear speech styles elicited in a laboratory setting for two speakers with PD. It would be premature to generalize results to clinical practice. This line of inquiry may ultimately prove to have clinical implications, however. Recall that energy characteristics contributed to improved intelligibility for

PDM01, whereas spectral adjustments explained the improved intelligibility of clear speech for PDM06. Whether a therapeutic technique that focuses on these variables might yield even greater improvements in intelligibility is unknown. For example, it might be speculated that a therapeutic technique that focuses on maximizing intensity envelope modulation might be most efficacious for enhancing intelligibility for PDM01, whereas a therapeutic technique that focuses on maximizing segmental or spectral contrast might be most efficacious for enhancing intelligibility for PDM06. Future studies could evaluate this suggestion. Results would have implications for a "one size fits most" approach to dysarthria treatment wherein the same therapeutic technique is recommended almost exclusively for a particular neurological diagnosis or dysarthria.

Finally, results suggest the hybridization process developed by Kain et al. (2008) is a powerful technique for identifying acoustic variables or combinations of variables explaining improvements in sentence intelligibility for speakers with mild dysarthria secondary to PD. Whether hybridization is possible using speech signals produced by talkers with more severe dysarthria for whom segment boundaries are blurred or indistinct and source characteristics are very poor remains to be determined. It seems likely that hybridization could be further exploited to identify speech changes explaining the perceptual consequences of dysarthria therapy techniques other than clear speech such as rate manipulation, contrastive stress, or an increased vocal intensity. Similar to clear speech, each of these global therapeutic techniques for dysarthria tends to be associated with simultaneous adjustments in multiple speech parameters such as duration, energy, F0, and segmental articulation (Yorkston, Hakel, Beukelman, & Fager, 2007). Changes to the speech signal responsible for improved perceptual outcomes associated with rate reduction, an increased vocal intensity, and contrastive stress remain poorly understood, although this knowledge would improve the scientific evidence base for dysarthria treatment and may ultimately help to optimize behavioral treatment or training programs for dysarthria (Liss, 2007; Neel., 2009; Yorkston et al., 2007).

Caveats and Conclusions

Several factors should be kept in mind when interpreting the present results. First, although speaker numbers are consistent with those used in other studies employing resynthesis, results are limited to two speakers with mild dysarthria secondary to PD and should not be generalized to other neurological diagnoses or more severe dysarthria. Perceptual judgments of intelligibility also were obtained in the presence of an adverse perceptual environment, as is commonly done in studies investigating the acoustic basis of intelligibility in neurotypical speech, including the original hybridization study of Kain and colleagues (2008). The importance of investigating speech intelligibility measurement in dysarthria in adverse listening conditions is noted by Yorkston and colleagues (Yorkston et al., 2007, 2010), and the multitalker babble used in the present study is arguably a more ecologically valid perceptual environment than other types of background noise, such as broadband noise or multitalker babble for which the F0 contour has been inverted. How different types of background noise impact intelligibility of neurologically normal speech has been a topic of study for some time (e.g., Binns & Culling, 2007; Festen & Plomp, 1990; Plomp & Mimpen, 1979), but intelligibility in background noise is only beginning to be reported in published

dysarthria studies (Bunton, 2006; Cannito et al., 2012). Preliminary data based on a small number of speakers further suggests that background noise impacts intelligibility in dysarthria in a slightly different way than for neurologically normal talkers and may even differ depending upon the perceptual characteristics or severity of the dysarthria (McAuliffe, Good, O'Beirne, LaPointe, 2008; McAuliffe, Schaefer, O'Beirne & LaPointe, 2009).

None of the hybrids were more intelligible than clear speech, and some hybrids were associated with reduced intelligibility relative to conversational speech. Similar results were reported by Kain et al. (2008). Although the quality of the resynthesis was very high (e.g., compare quality of hybrid spectrograms in Figure 3 to clear and conversational productions), the reduced intelligibility for some hybrids may be a byproduct of signal processing (see also Krause & Braida, 2009). That is, all hybrids involved acoustic changes that naturally occur when talkers adopt a clear speech style. However, in the majority of hybrids, acoustic changes that simultaneously occur in naturally produced clear speech were separated from one another, and listeners may have sensed that these stimuli were unusual. Last, it might be speculated that boundary artifacts associated with the digital signal processing somehow facilitated lexical segmentation and thus intelligibility. If this were the case, however, it is unclear why more hybrids were not associated with improved intelligibility relative to conversation, and perhaps especially why hybrids involving spectrum did not enhance intelligibility for PDM01.

Finally, intelligibility was assessed using both scaling and transcription tasks. In general, the pattern of findings with respect to improvements in intelligibility was similar for both tasks, although paired-comparisons for transcription were more consistently significant. Additional studies are needed, but the present results may suggest that a scaling task provides a more conservative estimate of improvements in intelligibility, at least for mild dysarthria presented in multi-talker babble. Interestingly, listener reliability was similar for both the transcription and scaling tasks. Thus, regardless of the specific task, listeners vary within themselves as well as among one another in their judgments of intelligibility for mild dysarthria presented in background noise. Determining the source of listener variability in judgments of intelligibility for dysarthria is an ongoing research challenge (see Choe, Liss, Azuma, & Mathy, 2012; McHenry, 2011).

In sum, not all production changes for clear speech produced by speakers with PD contribute to its improved intelligibility. The source of the improved intelligibility for one speaker's clear speech was attributable to adjustments in short-term spectrum, whereas energy characteristics explained the improved intelligibility of clear speech produced by another speaker. Thus, even for speakers with the same neurological diagnosis or dysarthria as well as broadly similar speech involvement as indexed by the SIT (Yorkston et al., 1996), acoustic variables explaining intelligibility improvement may differ. Results further suggest hybridization is a powerful technique for identifying acoustic variables explaining intelligibility variation in mild dysarthria.

Acknowledgments

References

American National Standards Institute. Specifications for audiometers. New York, NY: Author; 2004. (ANSI S3.6-2004)

Binns C, Culling JF. The role of fundamental frequency contours in the perception of speech against interfering speech. The Journal of the Acoustical Society of America. 2007; 122:1765–1776. [PubMed: 17927436]

Bochner JH, Garrison WM, Sussman JE, Burkard RF. Development of materials for the clinical assessment of speech recognition: The speech sound pattern discrimination test. Journal of Speech, Language, and Hearing Research. 2003; 46:889–900.

Boersma, P., Weenink, D. Praat: Doing phonetics by computer (Version 5.3.41) [Computer program]. 2013. Retrieved from http://www.praat.org/

Bradlow AR, Kraus N, Hayes E. Speaking clearly for children with learning disabilities: Sentence perception in noise. Journal of Speech, Language, and Hearing Research. 2003; 46:80–97.

Bradlow AR, Torretta GM, Pisoni DB. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. Speech Communication. 1996; 20:255–272. [PubMed: 21461127]

Bunton K. Fundamental frequency as a perceptual cue for vowel identification in speakers with Parkinson's disease. Folia Phoniatrica et Logopaedica. 2006; 58:323–339. [PubMed: 16966834]

Bunton K, Kent RD, Kent JF, Duffy JR. The effects of flattening fundamental frequency contours on sentence intelligibility in speakers with dysarthria. Clinical Linguistics & Phonetics. 2001; 15:181–193.

Cannito MP, Suiter DM, Beverly D, Chorna L, Wolf T, Pfeiffer RM. Sentence intelligibility before and after voice treatment in speakers with idiopathic Parkinson's disease. Journal of Voice. 2012; 26:214–219. [PubMed: 22209057]

Choe, Y-k, Liss, JM., Azuma, T., Mathy, P. Evidence of cue use and performance differences in deciphering dysarthric speech. The Journal of the Acoustical Society of America. 2012; 131:EL112–EL118. DOI: 10.1121/1.3674990 [PubMed: 22352609]

Cichetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment. 1994; 6:284–290.

Dromey C. Articulatory kinematics in patients with Parkinson Disease using different speech treatment approaches. Journal of Medical Speech-Language Pathology. 2000; 8:155–161.

Festen JM, Plomp R. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. The Journal of the Acoustic Society of America. 1990; 88:1725–1736.

Fleiss, JL. The design and analysis of clinical experiments. New York, NY: Wiley; 1986.

Frank T, Craig CH. Comparison of the auditec and rintelmann recordings of the NU-6. Journal of Speech and Hearing Disorders. 1984; 49:267–271. [PubMed: 6748621]

Goberman AM, Elmer LW. Acoustic analysis of clear versus conversational speech in individuals with Parkinson disease. Journal of Communication Disorders. 2004; 38:215–230. [PubMed: 15748725]

Hammen VL, Yorkston KM, Minifie FD. Effects of temporal alterations on speech intelligibility in Parkinsonian dysarthria. Journal of Speech and Hearing Research. 1994; 37:244–253. [PubMed: 8028306]

Hustad KC. A closer look at transcription intelligibility for speakers with dysarthria: Evaluation of scoring paradigms and linguistic errors made by listeners. American Journal of Speech-Language Pathology. 2006; 15:268–277. [PubMed: 16896176]

Hustad K. Effects of speech stimuli and dysarthria severity on intelligibility scores and listener confidence ratings for speakers with cerebral palsy. Folia Phoniatrica et Logopaedica. 2007; 59:306–317. [PubMed: 17965573]

Hustad KC. The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. Journal of Speech, Language, and Hearing Research. 2008; 51:562–573.

Hustad KC, Lee J. Changes in speech production associated with alphabet supplementation. Journal of Speech, Language, and Hearing Research. 2008; 51:1438–1450.

Hustad, KC., Weismer, G. Interventions to improve intelligibility and communicative success for speakers with dysarthria. In: Weismer, G., editor. Motor speech disorders. San Diego, CA: Plural; 2007. p. 261-303.

IEEE. IEEE recommended practice for speech quality measurements. IEEE Transactions on Audio and Electroacoustics. 1969; 17:225–246.

International Electrotechnical Commission. Electroacoustics—Sound level meters—Part 1: Specifications. Geneva, Switzerland: Author; 2002. (Paper No. 61672)

Kain A, Amano-Kusumoto A, Hosom JP. Hybridizing conversational and clear speech to determine the degree of contribution of acoustic features to intelligibility. The Journal of the Acoustical Society of America. 2008; 124:2308–2319. [PubMed: 19062869]

Kain AB, Hosom JP, Niu X, van Santen JPH, Fried-Oken M, Staehely J. Improving the intelligibility of dysarthric speech. Speech Communication. 2007; 49:743–759.

Kim H, Hasegawa-Johnson M, Perlman A. Vowel contrast and speech intelligibility in dysarthria. Folia Phoniatrica et Logopaedica. 2011; 63:187–194. [PubMed: 20938200]

Kim Y, Kent RD, Weismer G. An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. Journal of Speech, Language, and Hearing Research. 2011; 54:417–429.

Kim Y, Kuo C. Effect level of presentation to listeners on scaled speech intelligibility of speakers with dysarthria. Folia Phoniatrica et Logopaedica. 2012; 64:26–33. [PubMed: 21952405]

Krause JC, Braida LD. Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. The Journal of the Acoustical Society of America. 2002; 112:2165–2172. [PubMed: 12430828]

Krause JC, Braida LD. Acoustic properties of naturally produced clear speech at normal speaking rates. The Journal of the Acoustical Society of America. 2004; 115:362–378. [PubMed: 14759028]

Krause JC, Braida LD. Evaluating the role of spectral and envelope characteristics in the intelligibility advantage of clear speech. The Journal of the Acoustical Society of America. 2009; 125:3346–3357. [PubMed: 19425675]

Laures JS, Weismer G. The effects of flattened fundamental frequency on intelligibility at the sentence level. Journal of Speech, Language, and Hearing Research. 1999; 42:1148–1156.

Liss, JM. The role of speech perception in motor speech disorders. In: Weismer, G., editor. Motor speech disorders. San Diego, CA: Plural; 2007. p. 261-303.

Liu HM, Tsao FM, Kuhl PK. The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. The Journal of the Acoustical Society of America. 2005; 117:3879–3889. [PubMed: 16018490]

Liu S, Zeng FG. Temporal properties in clear speech perception. The Journal of the Acoustical Society of America. 2006; 120:424–432. [PubMed: 16875238]

McAuliffe, MJ., Good, PV., O'Beirne, GA., LaPointe, LL. Influence of auditory distraction upon intelligibility ratings in dysarthria; Poster presented at the Conference on Motor Speech; Monterey, CA. 2008. Retrieved from http://hdl.handle.net/10092/3398

McAuliffe, MJ., Schaefer, M., O'Beirne, GA., LaPointe, LL. Effect of noise upon the perception of speech intelligibility in dysarthria; Poster presented at the American Speech-Language and Hearing Association Convention; New Orleans, LA. 2009 Nov. Retrieved from http://hdl.handle.net/10092/3410

McHenry M. An exploration of listener variability in intelligibility judgments. American Journal of Speech-Language Pathology. 2011; 20:119–123. [PubMed: 21317298]

Milenkovic, P. TF32 [Computer software]. Madison: Department of Electrical and Computer Engineering, University of Wisconsin—Madison; 2005.

Miller SE, Schlauch RS, Watson PJ. The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. The Journal of the Acoustical Society of America. 2010; 128:435–443. [PubMed: 20649237]

Molloy, D. Standardized Mini-Mental State Examination. Troy, NY: New Grange Press; 1999.

Neel AT. Effects of loud and amplified speech on sentence and word intelligibility in Parkinson disease. Journal of Speech, Language, and Hearing Research. 2009; 52:1021–1033.

Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. Journal of Speech and Hearing Research. 1986; 29:434–445. [PubMed: 3795886]

Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing. III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. Journal of Speech and Hearing Research. 1989; 32:600–603. [PubMed: 2779204]

Plomp R, Mimpen AM. Speech-reception thresholds for sentences as a function of age and noise level. The Journal of the Acoustical Society of America. 1979; 66:1333–1342. [PubMed: 500971]

Ramig, L. The role of phonation in speech intelligibility: A review and preliminary data from patients with Parkinson's disease. In: Kent, R., editor. Intelligibility in speech disorders: Theory, measurement and management. Amsterdam, the Netherlands: John Benjamins; 1992. p. 119-155.

Smiljani R, Bradlow AR. Speaking and hearing clearly: Talker and listener factors in speaking style changes. Language and Linguistics Compass. 2009; 3:236–264. [PubMed: 20046964]

Spitzer SM, Liss JM, Mattys SL. Acoustic cues to lexical segmentation: A study of resynthesized speech. The Journal of the Acoustical Society of America. 2007; 122:3678–3687. [PubMed: 18247775]

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

Tjaden K, Lam J, Wilding G. Vowel acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions. Journal of Speech, Language, and Hearing Research. 2013; 56:1485–1502.

Tjaden K, Sussman JE, Wilding GE. 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. (in press).

Tjaden K, Wilding GE. Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings. Journal of Speech, Language, and Hearing Research. 2004; 47:766–783.

Turner GS, Tjaden K, Weismer G. The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. Journal of Speech and Hearing Research. 1995; 38:1001–1013. [PubMed: 8558870]

Uchanski RM, Choi SS, Braida LD, Reed CM, Durlach NI. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. Journal of Speech and Hearing Research. 1996; 39:494–509. [PubMed: 8783129]

Van Nuffelen G, De Bodt M, Wuyts F, Van de Heyning P. The effect of rate control on speech rate and intelligibility of dysarthric speech. Folia Phoniatrica et Logopaedica. 2009; 61:69–75. [PubMed: 19287175]

Watson PJ, Schlauch RS. The effect of fundamental frequency on the intelligibility of speech with flattened intonation contours. American Journal of Speech-Language Pathology. 2008; 17:348–355. [PubMed: 18840699]

Weismer, G. Speech intelligibility. In: Ball, MJ.Perkins, MR.Muller, N., Howard, S., editors. The handbook of clinical linguistics. Oxford, England: Blackwell; 2008. p. 568-582.

Weismer G, Jeng JY, Laures JS, Kent RD, Kent JF. Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. Folia Phoniatrica et Logopedica. 2001; 53:1–18.

Weismer G, Laures JS. Direct magnitude estimates of speech intelligibility in dysarthria: Effects of a chosen standard. Journal of Speech, Language, and Hearing Research. 2002; 45:421–433.

Yorkston, K., Beukelman, D., Strand, E., Hakel, M. Management of motor speech disorders in children and adults. Austin, TX: Pro-Ed; 2010.

Yorkston, K., Beukelman, DR., Tice, R. Sentence Intelligibility Test. Lincoln, NE: Tice Technologies; 1996.

Yorkston K, Hakel M, Beukelman DR, Fager S. Evidence for effectiveness of treatment of loudness, rate, or prosody in dysarthria: A systematic review. Journal of Medical Speech-Language Pathology. 2007; 15:11–36.

Yunusova Y, Green JR, Greenwood L, Wang J, Pattee GL, Zinman L. Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis. Folia Phoniatrica et Logopaedica. 2012; 64:94–102. [PubMed: 22555651]

Yunusova Y, Weismer G, Kent RD, Rusche NM. Breath-group intelligibility in dysarthria: Characteristics and underlying correlates. Journal of Speech, Language, and Hearing Research. 2005; 48:1294–1310.
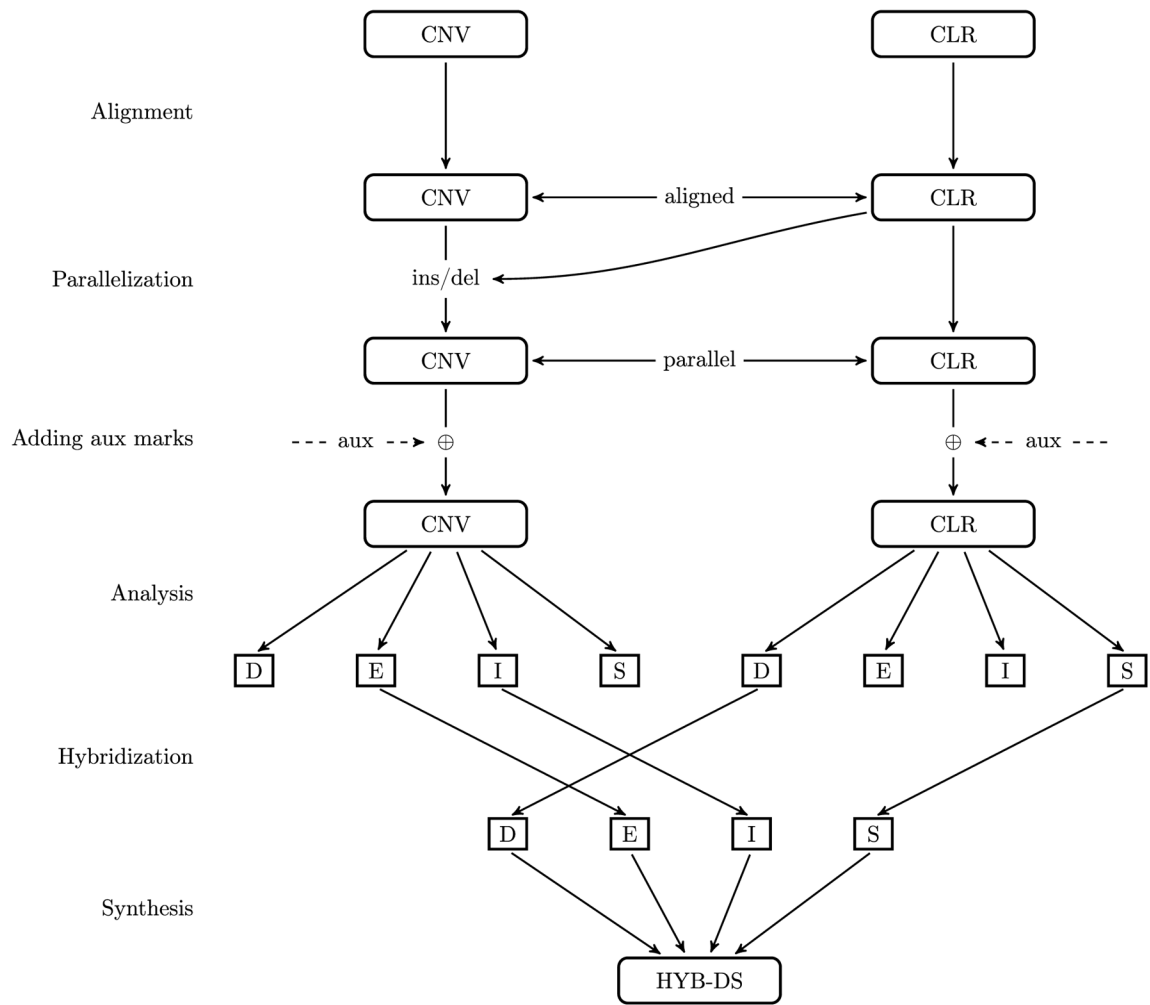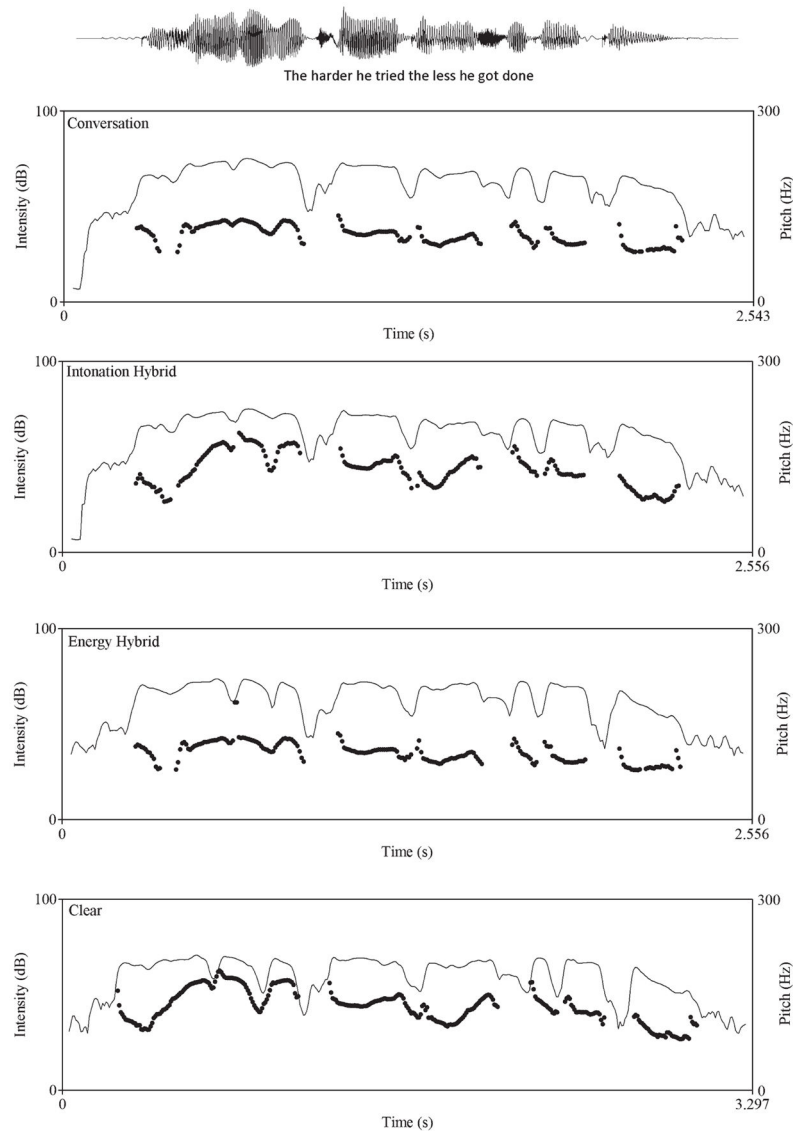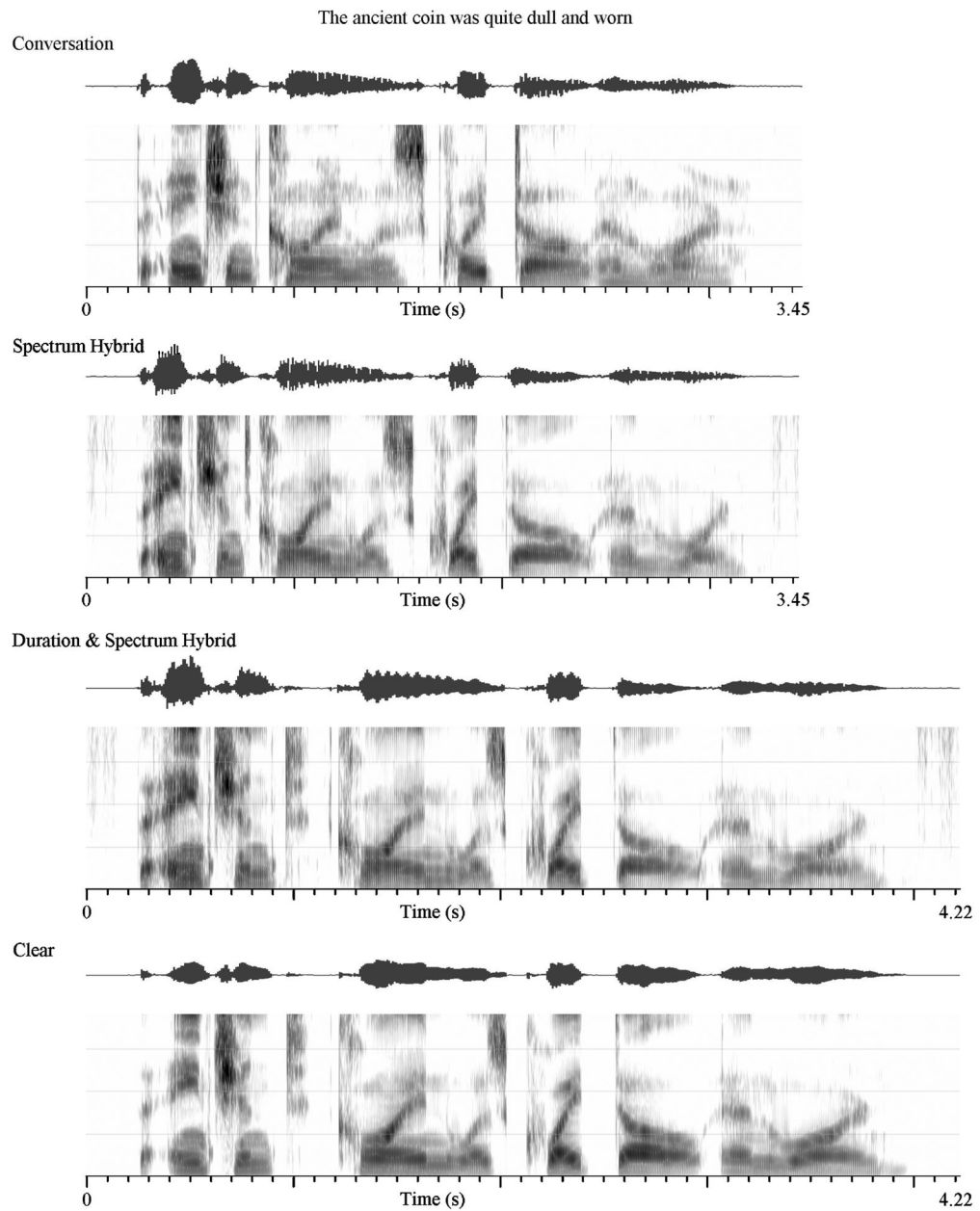
Figure 1.
Block diagram summarizing hybridization process. CNV = conversational; CLR = clear; ins = insert; del = delete; aux = auxiliary; D = duration; E = energy; I = intonation; S = short-term spectrum; HYB-DS = hybrid of duration and short-term spectra.
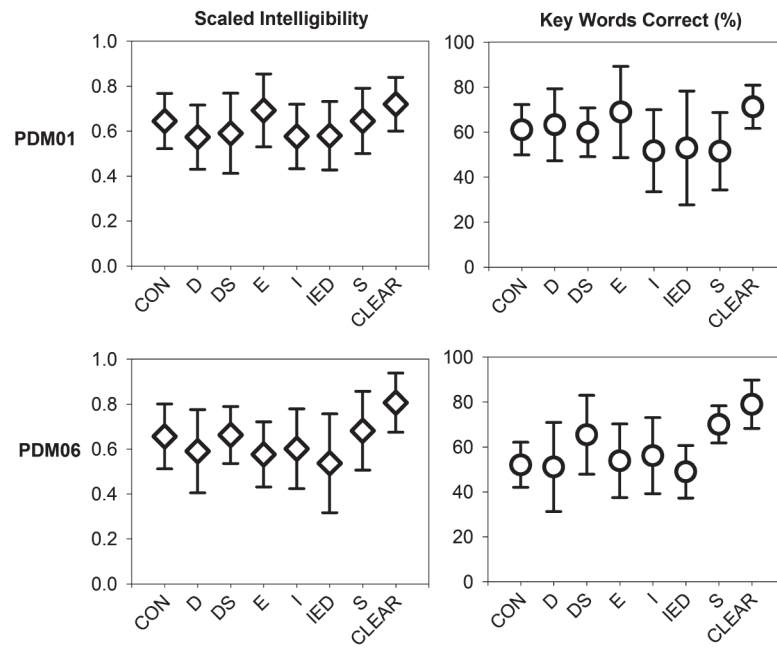
Figure 2.
Selected amplitude normalized hybrid stimuli as well as the amplitude normalized conversational and clear productions for PDM01 are shown. The waveform display corresponds to the conversational production. In each panel, root-mean-square intensity is shown in gray, and the F0 time history is shown in black. The aspect ratio of the *x*-axis for the clear production differs from other panels to facilitate comparison of energy and F0 traces from the clear production to the hybrids as well as the conversation production.
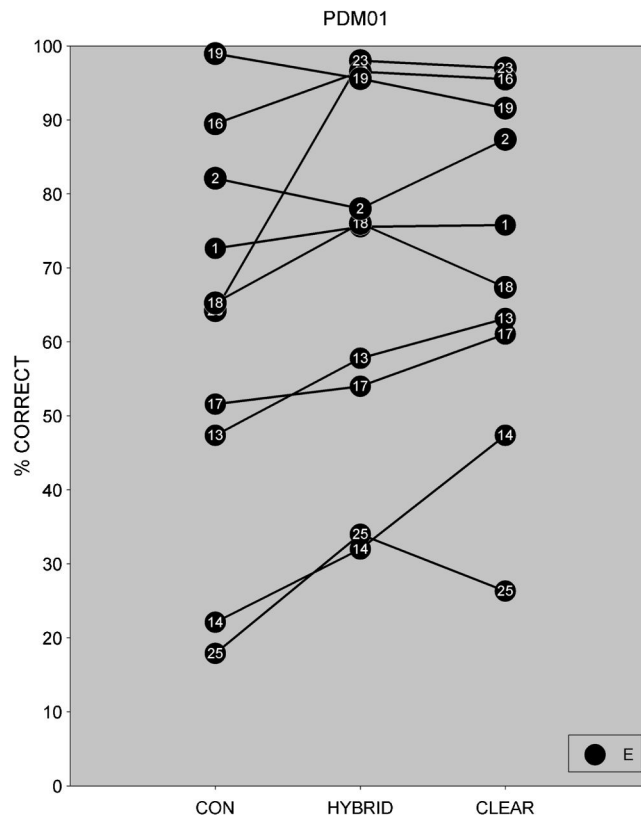
Figure 3.
Selected amplitude normalized hybrid stimuli, as well as the amplitude normalized conversational and clear productions for PDM06 are shown.
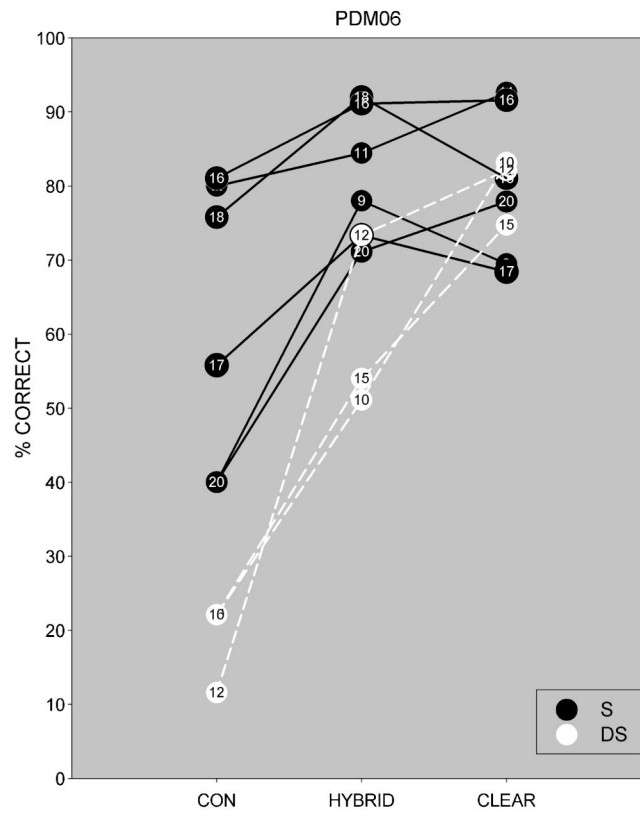
Figure 4.
Mean percentage correct scores and scale values pooled across sentences and listeners are reported. Vertical bars indicate ±1 *SD*. Results for conversational (CON) and clear stimuli are shown on the extremes of the *x*-axis, and results for hybrids—duration (D); the combination of duration and short-term spectrum (DS); energy (E); intonation (I); prosody, defined as the combination of intonation, energy, and duration (IED); and short-term spectrum (S)—are reported in the middle of the *x*-axes.

Figure 5.
Transcription results for individual sentences are reported for PDM01.

Figure 6.

Transcription results for individual sentences are reported for PDM06. White symbols indicate sentences for which the duration/spectrum hybrid was associated with the best intelligibility among the six hybrid variants as well as improved intelligibility above conversational ($n = 3$). Black symbols indicate sentences for which the spectrum hybrid was associated with the best intelligibility among the six hybrids as well as improved intelligibility relative to conversational ($n = 6$).

Table 1

Acoustic measures for both clear and conversational sentence production to characterize the speech patterns of participants.

Speaker	Vowel space area (Hz²)		Sound pressure level (dB)		Fundamental frequency (Hz)		Mean articulation rate (syllables/second)
	Peripheral	Nonperipheral	M	SD	M	Range	
PDM01							
Conversation	244,503	73,101	72.76	7.69	120	65	4.13
Clear	212,831	71,433	76.73	9.97	151	105	3.15
PDM06							
Conversation	91,462	23,238	71.60	8.44	113	35	3.65
Clear	171,423	50,169	77.63	10.27	128	53	2.91
Control males							
Conversation	225,205	56,084	74.80	8.56	114	91	3.76
Clear	350,181	104,116	79.83	9.82	138	122	2.43

Note. For comparison purposes, selected acoustic measures for eight male control talkers reported in Tjaden et al. (2013) are provided.

Table 2

Overall percentage correct scores and mean scale values for clear and conversational stimuli.

| Speaker | Con | Clr | Difference relative to conversation | | | | | |
			D	DS	S	E	I	IED
			Key words transcription					
PDM01	61.5%	71.6%	**1.9**	−1.1	−10.7	**8.7**	−10.7	−9.1
PDM06	51.2%	77.8%	−0.4	**13.4**	**18**	**1.4**	**5**	−2
			Scaled intelligibility[a]					
PDM01	.65	.72	−.07	−.05	0	**.05**	−.07	−.07
PDM06	.66	.81	−.07	**.01**	**.03**	−.08	−.05	−.12

Note. Intelligibility difference scores are reported for hybrid stimuli to capture the change in intelligibility relative to conversation. Negative difference scores indicate reduced intelligibility relative to conversation, whereas positive difference scores indicate improved intelligibility relative to conversation. Difference scores for transcription were calculated from overall percent correct scores, whereas difference scores for the scaling task were calculated using scale values averaged across listeners. Bolded cells indicate hybrids for which intelligibility was improved relative to conversation.

[a]1 = *completely understand,* 0 = *cannot understand.*