Effects of speaking task on intelligibility in Parkinson's disease

KRIS TJADEN[1] and **GREG WILDING**[2]

[1]Department of Communicative Disorders and Sciences, University at Buffalo, Buffalo, NY, USA

[2]Department of Biostatistics, University at Buffalo, Buffalo, NY, USA

Abstract

Intelligibility tests for dysarthria typically provide an estimate of overall severity for speech materials elicited through imitation or read from a printed script. The extent to which these types of tasks and procedures reflect intelligibility for extemporaneous speech is not well understood. The purpose of this study was to compare intelligibility estimates obtained for a reading passage and an extemporaneous monologue produced by 12 speakers with Parkinson's disease (PD). The relationship between structural characteristics of utterances and scaled intelligibility was explored within speakers. Speakers were audio-recorded while reading a paragraph and producing a monologue. Speech samples were separated into individual utterances for presentation to 70 listeners who judged intelligibility using orthographic transcription and direct magnitude estimation (DME). Results suggest that scaled estimates of intelligibility for reading show potential for indexing intelligibility of an extemporaneous monologue. Within-speaker variation in scaled intelligibility also was related to the number of words per speech run for extemporaneous speech.

Keywords

intelligibility; dysarthria; speaking task

Introduction

Intelligibility, defined as the extent to which a speaker's acoustic signal is understood by a listener, is an important construct in the study and management of dysarthria. For example, quantitative measures of intelligibility may be used to establish a baseline prior to treatment, to document treatment progress; to provide an objective estimate of severity for medical, legal or research purposes; or to follow patients over time to document changes in speech severity related to disease progression as well as medical or surgical intervention (Duffy, 2005).

Most published intelligibility tests for dysarthria provide an estimate of speech severity at the sentence- or word level (Enderby & Palmer, 2008; Kent, Weismer, Kent, & Rosenbek,

1989; McHenry and Parle, 2006; Yorkston & Beukelman, 1981; Yorkston, Beukelman and Hakel 1996). The Frenchay Dysarthria Assessment (FDA) is an exception in providing a scaled estimate of intelligibility for conversation (Enderby & Palmer, 2008). Reliance on interval scaling techniques limits the value of these data, however (Schiavetti, 1992). The Assessment of Intelligibility of Dysarthric Speech (Yorkston & Beukelman, 1981) or the computerised version of this test, the Speech Intelligibility Test (SIT) (Yorkston & Beukelman, 1996), is arguably the most widely used clinical tool for quantifying intelligibility in adults with dysarthria (Duffy, 2005). The SIT includes a word task as well as a connected speech task comprising sentences ranging in length from 5 to 15 words. Speakers are provided with a printed list of stimuli and are instructed to read or repeat each item on the list in response to an auditory model provided by the examiner. This procedure is intended to minimise reading errors. Reading a printed script or repeating stimuli in response to an auditory model provides speakers with an external cue to guide performance. Studies from the motor control literature suggest that self-initiated or internally cued movements may be more impaired in Parkinson's disease (PD) than externally cued movements (e.g. Georgiou, Bradshaw, Iansek, Phillips, Mattingley, & Bradshaw, 1994; Cunnington, Iansek, & Bradshaw, 1999). Performance on externally cued speech tasks, such as reading from a printed script or producing sentences or words in response to an auditory model, therefore might be expected to be enhanced for individuals with PD relative to speech tasks which require an individual to speak extemporaneously. Indeed, as reviewed in the following section, several studies suggest that intelligibility for speakers with PD may be reduced for extemporaneous speech tasks compared with tasks for which an external cue or printed script is available.

Kempler and Van Lancker (2002) reported intelligibility data for a variety of speech tasks produced by a single speaker with PD. Results indicated a large difference in intelligibility for conversational, spontaneous speech (29%) compared with reading aloud a printed transcript of the conversational speech sample (78%). Similar observations were made by Canter and Van Lancker (1985) in a case study of PD as well as by Frearson (1985) for an unspecified number of speakers with PD. More recently, Bunton and Keintz (2008) measured intelligibility for four individuals with PD producing a variety of speech tasks. Participants were audio-recorded in the laboratory as they read words and sentences. A monologue also was recorded during which participants described a recent vacation. Finally, covert audio recordings were obtained as participants conversed with an investigator. Orthographic transcription indicated similar intelligibility for words ($M = 91\%$), sentences ($M = 90\%$) and monologues ($M = 88\%$) when participants were only talking and were not required to perform a concurrent manual motor task. Intelligibility for conversational speech was significantly reduced ($M = 74\%$), however, relative to other speech tasks. Finally, the literature contains anecdotal evidence of reduced intelligibility for structured speech tasks in PD compared with spontaneous speech (Sarno, 1968; Weismer, 1984).

In summary, structured speech tasks that require an individual to read from a printed script or repeat sentences in response to an auditory model are widely used by both clinicians and researchers for the purpose of measuring intelligibility in dysarthria. Despite the greater face validity of extemporaneous speech tasks, structured speech tasks are typically employed because the consistent content facilitates efficient administration and scoring as well as

comparison of measures within and across speakers. Some studies have reported similar intelligibility for structured and extemporaneous speech produced by individuals with PD, but other studies suggest that structured speech tasks may not well represent intelligibility for extemporaneous speech. Clearly, additional studies are required to help determine the need for a formal intelligibility test for dysarthria that employs extemporaneous speech materials. Thus, the primary purpose of this study was to compare intelligibility estimates for a paragraph reading task and an extemporaneous monologue task produced by speakers with PD. The association between structural characteristics of speech runs and scaled intelligibility also was explored within speakers.

Methods

Speakers

A total of 12 individuals (6 men, 6 women) with PD participated. These speakers are part of an ongoing project investigating acoustic-perceptual characteristics of dysarthria (e.g. Tjaden & Wilding, 2004; 2005; in press). All speakers scored at least 25/30 on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975) and had pure tone audiometric thresholds of 40 dB or better in at least one ear at 1, 2 and 4 kHz. All participants were native speakers of American English, reported no history of neurological disease prior to the diagnosis of PD, had not received surgical treatment for PD and had not participated in the Lee Silverman Voice Treatment® programme. Additional speaker characteristics are summarised in Table I. Male and female speakers are identified with the letters PDM and PDF, respectively. Dysarthria diagnoses, severity estimates and prominent-deviant perceptual characteristics reflect the consensus judgment of three speech–language pathologists (SLPs) based on audio recordings of vowel prolongation, diadochokinesis, the Grandfather Passage (Duffy, 2005) and a 2-minute monologue. As described in the following section, the monologue was one of the speech tasks of interest in the current study.

Procedures

Participants were audio-recorded directly to computer hard disk in a sound-treated room while producing a variety of speech materials. The John Passage (Tjaden & Wilding, 2004) and the first 90 seconds of a 2-minute monologue task were of interest to the current study. These tasks and associated analyses are described more fully in the following sections. Speech recordings took place approximately 1 hour after ingestion of anti-Parkinsonian medication.

Reading passage

The John Passage is a 192-word reading passage designed to include a variety of phonemes. Using TF32 (Milenkovic, 2002), reading passages were segmented into speech runs for presentation to listeners. A speech run was operationally defined as a stretch of speech bounded by a silent period or pause between words of at least 200 ms. Conventional acoustic criteria were used to identify onsets and offsets of runs (see Tjaden & Wilding, 2004). Structural characteristics of runs, in the form of word and syllable counts were obtained to further describe speech runs and also to explore their relationship to intelligibility (e.g.

Yunusova, Weismer, Kent, & Rusche, 2005). The printed script of the reading passage was used to determine word and syllable counts.

Monologue

For the monologue task, speakers were instructed to talk about a topic of interest such as their family or a hobby. Procedures for speech run segmentation were similar to those described previously for the reading passage. An orthographic transcript or gloss of each monologue was created to facilitate identification of speech runs and to assist in obtaining syllable and word counts. A trained research assistant generated an initial transcript, which was reviewed and edited by a second research assistant. A final review of each transcript was performed by the two research assistants and the primary investigator as a group. Repeated listening and discussion were used to reach a consensus concerning the content of transcripts as well as word and syllable counts. For nine speech runs (i.e. approximately 2% of the entire corpus of runs for the monologue task), consensus could not be reached concerning content. Word and syllable counts were not obtained for these speech runs, but the runs were presented to listeners for scaling of intelligibility. Both silent and filled pauses were used to determine boundaries of runs for the monologue task. A filled pause was operationally defined as a sound of hesitation or hesitation device (Goldman-Eisler, 1961). Examples of filled pauses include the sounds or syllables 'ah', 'um' and 'er'. Partial words were assigned a value of 0.5 when determining word counts.

Listeners and listening task

A total of 70 listeners judged intelligibility. Listeners were recruited from postings at the University at Buffalo and were paid an hourly rate for participation. Listeners were native talkers of American English, denied any hearing loss, reported minimal exposure to motor speech disorders, had not taken a course in motor speech disorders and reported no history of speech–language disorders. Individual listeners completed the perceptual task in a double-walled audiometric booth. Stimuli were presented via headphones and presentation was controlled using custom software. The perceptual task was self-paced, with listeners pressing a button on a response box to present the next trial. Listeners manually recorded their intelligibility judgments to a response sheet. Prior to presentation to listeners, speech runs were normalised to a peak intensity of 75 dB using Cool Edit Pro (Syntrillium Corporation Phoenix, AZ, USA) (see also Tjaden and Wilding, 2004; Hustad, 2008). This procedure controlled for the effects of audibility on judgments of intelligibility and also allowed for more straightforward comparison of findings to those reported by Bunton and Keintz (2008), as sentence and monologue tasks produced by speakers with PD in this earlier study were characterised by similar average intensities.

Sixty listeners (16 males, 44 females; M age = 32 years; SD = 13 years) judged intelligibility for the reading passage using orthographic transcription and modulus-free direct magnitude estimation (DME). Two different intelligibility measures were obtained for the reading task to determine whether results were dependent on the nature of the method for quantifying intelligibility. This approach also allowed for statistical comparison of scaled estimates of intelligibility for the reading passage and monologue tasks. To minimise

familiarity effects, each listener orthographically transcribed a random ordering of reading passage runs produced by one speaker and scaled intelligibility of a random ordering of reading passage runs produced by a second speaker using DME. Thus, a given listener only heard the content of the reading passage twice. Ten per cent of speech runs were presented twice for use in determining reliability. The order in which listeners performed the transcription and scaling tasks was counterbalanced for a given speaker's speech materials. Order effects were further minimised by creating four random orderings of a speaker's reading passage runs for presentation. In all, five listeners orthographically transcribed speech runs for a given speaker's reading passage and five different listeners provided magnitude estimates of intelligibility for the reading passage produced by the same speaker. Each listener completed the experiment in a single session, which lasted approximately 1 hour.

For orthographic transcription of reading passages, listeners were instructed to write down what the speaker said. A word was scored as correct if the transcript exactly matched the printed script of the reading passage. A per cent correct score was obtained for each speech run by tallying the number of words correctly transcribed by the five listeners, dividing by the total number of possible words and multiplying by 100. An overall per cent correct score for each listener also was obtained by tallying the total number of words correctly transcribed, dividing by the total number of possible words and multiplying by 100. For DME, intelligibility was operationally defined as 'the ease with which speech is understood' (Tjaden & Wilding, 2004; Yunusova et al., 2005). Scaled estimates of intelligibility were converted to a common scale (Engen, 1971). An average scaled estimate of intelligibility was obtained for each speech run by calculating the geometric mean of scale values provided by the five listeners. An overall mean for each speaker was determined by calculating the mean of scale values for all speech runs.

An additional 10 listeners (5 males, 5 females; M age = 25 years; SD = 5 years) scaled intelligibility of speech runs for the monologue task using DME. Speech runs for all speakers were pooled and two random orderings of the stimuli were generated for presentation to listeners. Ten per cent of speech runs were presented twice for use in determining reliability. Each listener completed the experiment in a single session lasting approximately 2 hours.

Scaled estimates of intelligibility were converted to a common scale (Engen, 1971). Using procedures similar to those described for the reading passage, an average intelligibility estimate was calculated for each speech run. An overall scaled estimate of intelligibility also was calculated for each speaker.

Listener reliability

Intrajudge reliability for transcription of reading passage runs was determined by comparing the number of words correctly transcribed for original and reliability trials of speech runs. On average, listeners' transcriptions for original and reliability trials differed by an average of 0.29 correct words (SD = 0.69 words). In approximately 79% of occurrences, the original and reliability trials contained the same number of correctly transcribed words. In an additional 16% of cases, original and reliability trials differed by only one correct word, and

in an additional 3% of cases the first and second trials differed by two correct words. There was a single instance in which one listener's transcription of an original and reliability trial differed by six correct words.

To assess interjudge agreement for transcription of reading passages, per cent correct scores for the five listeners who transcribed speech runs for each talker were compared. Across the 12 speakers with PD, intelligibility scores for the best and worst listener assigned to each speaker differed by an average of 15% ($SD = 12\%$). This finding is consistent with studies showing that non-expert listeners vary in the ability to orthographically transcribe dysarthric speech (e.g. Hustad, Jones, & Dailey, 2003; Hustad, 2006).

Kendall's tau-b was used to ascertain reliability for DME. For the reading passage, intrajudge reliability yielded a matrix of concordant pairs with significant coefficients ($p < 0.05$) ranging from 0.44 to 0.68 ($M = 0.53$; $SD = 0.13$). Interjudge reliability yielded a matrix of concordant pairs with significant coefficients ranging from 0.27 to 0.42 ($M = 0.33$; $SD = 0.04$). For the monologue task, intrajudge reliability yielded a matrix of concordant pairs with significant coefficients ranging from 0.23 to 0.67 ($M = 0.44$; $SD = 0.17$). Interjudge reliability yielded a matrix of concordant pairs with significant coefficients ranging from 0.21 to 0.58 ($M = 0.42$; $SD = 0.10$). Although not overwhelmingly strong, these reliability estimates are at least broadly consistent with those reported in other studies employing modulus-free magnitude estimation to scale intelligibility in dysarthria (Tjaden & Wilding, 2004; Yunusova et al., 2005). The topic of listener reliability is considered further in the Section 'Discussion'.

Data analysis

Speech run characteristics and intelligibility scores were summarised using standard descriptive statistics. The sign test was used to evaluate differences in scaled estimates of intelligibility for the paragraph reading and monologue tasks. Spearman rank order correlations were used to evaluate the strength of association between the various intelligibility estimates. Correlation analysis also was used to explore the relationship between structural characteristics of speech runs and scaled intelligibility within speakers. Tests were two-tailed and a nominal significance level of 0.05 was used.

Results

Table II summarises stimuli characteristics. Inspection of this table suggests that structural characteristics of runs were broadly similar for the paragraph reading and monologue tasks. Average per cent correct scores and standard deviations for orthographic transcription of the paragraph reading task are reported in Figure 1. Speakers are arranged from least (PDM5 = 66%) to most (PDF2 = 97%) intelligible ($M = 84\%$; $SD = 12\%$). Note that per cent correct scores for the five listeners assigned to a given speaker were most variable for speakers on the left half of the abscissa, corresponding to speakers with the lowest overall intelligibility.

Figure 2 reports overall scaled estimates of intelligibility for the paragraph reading and monologue tasks. Speakers are arranged in the same order as in Figure 1. Scaled estimates of intelligibility for paragraph reading in Figure 2 range from 1.817 to 2.387 ($M = 2.065$; $SD =$

0.165). Similar data for the Monologue task range from 1.699 to 2.210 ($M = 2.056$; $SD = 0.154$). By way of comparison, the nine speech runs from the monologue task for which consensus concerning the orthographic transcript could not be determined had an average scale value of 1.57 ($SD = 0.50$).

Statistical analysis further indicated no difference in scaled intelligibility for the paragraph reading and monologue tasks. Inspection of individual speaker data in Figure 2 generally supports this group finding, although PDM4 and PDF4 exhibit a relatively greater difference in scaled intelligibility for the paragraph reading and monologue tasks compared with other speakers. In contrast to the finding of similar scaled intelligibility for the paragraph reading and monologue tasks, neither intelligibility measure for the paragraph reading task was significantly correlated with scaled estimates of intelligibility for the monologue task (paragraph transcription vs. monologue DME $\rho = 0.51$, $p = 0.09$; paragraph DME vs. monologue DME $\rho = 0.36$, $p = 0.26$). The two intelligibility measures for the paragraph reading task were moderately correlated ($\rho = 0.59$, $p = 0.04$). When PDF4 and PDMF4's data were excluded and correlations were repeated, the strength of association for the two reading passage measures increased to 0.80 ($p = 0.003$), but the relationship between paragraph transcription and scaled intelligibility for the monologue was essentially unchanged ($\rho = 0.54$; $p = 0.10$). The correlation between scaled estimates of intelligibility for the reading passage and monologue tasks increased from 0.36 to 0.61 and approached significance ($p = 0.054$).

Table III reports the results of the within-speaker analysis exploring the relationship between structural characteristics of speech runs and scaled estimates of intelligibility. Word and syllable counts for speech runs were highly correlated (i.e. data pooled across speakers: monologue task $\rho = 0.95$; paragraph reading task $\rho = 0.97$). Thus, word counts were used for this analysis as in other dysarthria studies examining intraspeaker variations in intelligibility (Yunusova *et al.*, 2005). Speakers in Table III are arranged in the same order as for Figures 1 and 2, from least to most intelligible based on per cent correct scores for orthographic transcription of the reading passage. Data in Table III suggest a statistically significant association between word counts and scaled intelligibility for a subset of speakers, most notably for the monologue task.

An additional qualitative analysis was undertaken to further explore the relationship between structural characteristics of speech runs and scaled estimates of intelligibility for the monologue task. For each speaker, the five least intelligible speech runs and five most intelligible speech runs were identified for comparison of word counts. Data are reported in Figure 3. Each symbol in Figure 3 corresponds to a speech run in the monologue task. The upper panel reports scaled estimates of intelligibility and the lower panel reports word counts associated with each speech run. The non-continuous nature of the data in the lower panel results in a certain amount of symbol overlap. Inspection of the upper panel provides an indication of the range of scaled intelligibility for each speaker's monologue task. The lower panel further suggests that the least intelligible runs (black circles) tended to be characterised by relatively fewer words. In fact, 77% of the least intelligible speech runs in Figure 3 had fewer than four words (black circles falling below the horizontal line), whereas only 13% of the most intelligible runs had fewer than four words (grey squares falling below

the horizontal line). Relatedly, only 23% of the least intelligible runs in Figure 3 had four or more words (black circles falling above the horizontal line), whereas 87% of the most intelligible runs had four or more words (grey squares falling above the horizontal line).

Discussion

Scaled estimates of intelligibility for paragraph reading and monologue tasks

A major finding of this study was that scaled estimates of intelligibility for the paragraph reading and monologue tasks were not significantly different (Figure 2). This group result appears to hold for individual speakers as well, although for reasons that are unclear, PDM4 and PDF4 exhibited a relatively greater difference in scaled intelligibility for structured and extemporaneous speech tasks compared with other speakers, albeit in opposite directions. The finding of similar scaled estimates of intelligibility for the paragraph reading and monologue tasks agrees with results reported by Bunton and Keintz (2008) for four speakers with PD, but differs from anecdotal reports and case studies indicating poorer intelligibility for spontaneous speech in PD compared with structured speech materials (e.g. Weismer, 1984; Kempler & Van Lackner, 2002). The finding of similar measures of intelligibility for structured and extemporaneous speech tasks also differs from studies reporting that externally cued movements for individuals with PD are superior to self-initiated movements (e.g. Georgiou et al., 1994; Cunnington et al., 1999). Results from this study therefore suggest caution in generalising from single-speaker reports of PD as well as in assuming that findings from the general motor control literature extend to speech in PD. The current findings as well as those reported by Bunton and Keintz (2008) further indicate that intelligibility measures obtained for structured speech tasks show potential for indexing intelligibility of extemporaneous speech in persons with mostly mild to moderate dysarthria secondary to PD, at least when connected speech materials are overtly elicited in the clinic or laboratory and audibility is similar for both types of speech materials or tasks. The extent to which speech produced by persons with dysarthria in the laboratory or clinic setting is representative of functional communication abilities outside these settings is an important topic for future studies. Dual-task paradigms that require speakers to talk and perform a concurrent manual motor task as well as covert recordings of conversation obtained in a laboratory or clinic setting show promise in this regard as reported by Bunton and Keintz (2008), but need to be evaluated in studies employing greater speaker numbers. Listener judgments of comprehension, as described by Hustad (2008) also may provide insight regarding functional communication abilities. Finally, the novel procedure for determining conversational intelligibility in dysarthria developed by Adams, Dykstra, Jenkins, and Jog (2008), wherein speakers are audio-recorded in the presence of background noise, simulates many real-world communication settings and situations.

In addition to the finding of similar scaled estimates of intelligibility for the paragraph reading and monologue tasks, structural characteristics of speech runs for both tasks were comparable (Table II). This finding is of clinical relevance because structural characteristics of utterances are used to evaluate and monitor breath-grouping characteristics in dysarthria. For example, a speaker who uses the same number of words in each breath group or utterance may be a candidate for therapy aimed at increasing flexibility or variability of

breath-group duration (Duffy, 2005). Therapy of this type may initially involve reading sentences or paragraphs for which breath-groups have been marked, with the end goal of speaking extemporaneously without the use of a printed script to guide performance. The current results suggest that evaluation of breath-grouping characteristics using a printed script provides a reasonable indication of breath-grouping characteristics for extemporaneous speech. By extension, treatment programmes employing reading passages to train breath-group flexibility may be expected to facilitate carry-over to extemporaneous speech.

Although scaled estimates of intelligibility did not differ for the paragraph reading and monologue tasks, results of the correlation analysis indicated that these intelligibility measures were not strongly correlated. This apparent discrepancy can be resolved by considering the nature of the information provided by the two analyses. The group comparison of intelligibility measures reviewed in the preceding section tests whether distributions of intelligibility scores for the paragraph reading and monologue tasks were identical. The correlation analysis indicates the extent to which a given speaker's intelligibility scores are ranked similarly with respect to the entire group. For example, PDM1's scaled intelligibility for the monologue task of 2.064 ranked him seventh most intelligible among the 12 speakers, but his average scaled intelligibility of 2.224 for the paragraph reading task ranked him second most intelligible among the 12 speakers. This example illustrates how a fairly small absolute difference in scaled intelligibility can have a large impact on the relative ranking of an individual with respect to the larger group, which in turn explains why the correlation between scaled estimates of intelligibility for the paragraph reading and monologue tasks was not significant. When data for PDM4 and PDF4 were excluded, the correlation between scaled estimates of intelligibility for the reading passage and monologue tasks increased from 0.36 to 0.61 and approached significance. Nonetheless, until additional studies employing greater speaker numbers are available, it appears that a given speaker's relative intelligibility ranking for a reading passage cannot be taken to be representative of her relative intelligibility ranking for extemporaneous speech.

Comparison of per cent correct scores and scaled estimates of intelligibility

Judging how easily speech is understood is a different perceptual task than writing down each word a speaker says. In fact, writing down each word a speaker says is probably not very representative of perceptual processes used to understand connected speech (see Hustad, 2008; Weismer, 2008). The fact that the correlation between per cent correct scores and scaled estimates of intelligibility for the reading passage was significant, however, hints at the possibility that both perceptual tasks were indexing overall speech severity in a related manner. In contrast, per cent correct scores for the reading passage and scaled estimates of intelligibility for the monologue task were not correlated. It seems likely that differences in the perceptual task of writing down each word that is understood versus estimating how easily speech is understood would be magnified when different types of speech tasks or materials are under consideration. The finding of no relationship between per cent correct scores for the reading passage and scaled estimates of intelligibility for the monologue task further suggests that published intelligibility tests employing orthographic transcription may not well represent of how easily extemporaneous speech is understood.

Structural characteristics of speech runs and relationship to intelligibility

The present study is the first to conduct within-speaker analyses examining the relationship between structural characteristics of speech runs and intelligibility for an extemporaneous speech task produced by speakers with dysarthria. Resultsof this analysis indicated a modest relationship between word counts and scaled intelligibility for a subset of speakers, especially for the monologue task (Table III). The direction of the relationship was such that speech runs characterised by relatively greater numbers of words were judged to be more intelligible. This finding agrees with other studies reporting that contextual cues can facilitate intelligibility of connected speech in dysarthria (e.g. Yorkston & Beukelman, 1978; Hustad, 2007) as well as a dysarthria study by Yunusova et al. (2005) reporting that speech runs in a reading passage characterised by greater numbers of words were associated with higher scaled estimates of intelligibility, at least for some talkers. The robustness of the correlations for certain speakers in Table III further suggests that numbers of words per speech run likely helps to explain variation in intelligibility for these individuals. The qualitative analysis in Figure 3 also indicates the importance of structural characteristics to within-speaker variations in intelligibility for the current speakers. The least intelligible speech runs for the monologue task were much more likely to contain fewer than four words compared with speech runs that were judged to be most intelligible. Number of words per speech run therefore may be reasonable therapy targets for speakers in the current study. Additional research is needed to determine other variables contributing to the within-speaker variations in intelligibility noted for speakers in the current study.

Finally, the issue of listener reliability for the scaling task deserves comment. An advantage of free-modulus magnitude estimation over magnitude estimation employing a modulus is that the former procedure allows for cross-study comparisons. Anecdotal evidence suggests that listeners find the task of numerically scaling intelligibility, in the absence of an anchor or modulus, to be unusual (Tjaden & Wilding, 2004). It is perhaps not entirely unexpected then, that intrajudge and interjudge reliability for the scaling task was not particularly robust, and it is important to keep the issue of listener reliability in mind when drawing conclusions from the present data set.

In conclusion, keeping in mind the issue of listener reliability, findings from this study may be interpreted to suggest that scaled estimates of intelligibility for a paragraph reading task produced by individuals with dysarthria secondary to PD show potential for indexing intelligibility of an extemporaneous monologue. However, per cent correct scores for orthographic transcription of sentence-level material, as used in published intelligibility tests, did not provide a good indication of how easily extemporaneous speech is understood. Results further indicated that number of words per speech run for the extemporaneous speech task likely contributed to within-speaker variations in scaled intelligibility. One implication is that speakers in this study may benefit from therapies aimed at manipulating number of words per speech run. These results also suggest the feasibility of using within-speaker variation in intelligibility as a mechanism for identifying explanatory variables underlying intelligibility in dysarthria.

Acknowledgments

References

Adams SG, Dykstra A, Jenkins M, Jog M. Speech-to-noise levels and conversational intelligibility in hypophonia and Parkinson's disease. Journal of Medical Speech-Language Pathology. 2008; 16:165–172.

Bunton K, Keintz C. The use of a dual-task paradigm for assessing speech intelligibility for clients with Parkinson disease. Journal of Medical Speech-Language Pathology. 2008; 16:141–155. [PubMed: 21637738]

Canter GJ, Van Lancker D. Disturbances of the temporal organization of speech following bilateral thalamic surgery in a patient with Parkinson's disease. Journal of Communication Disorders. 1985; 18:329–349. [PubMed: 4056082]

Cunnington R, Iansek R, Bradshaw JL. Movement-related potentials in Parkinson's disease: external cues and attentional strategies. Movement Disorders. 1999; 14:63–68. [PubMed: 9918346]

Duffy, JR. Motor speech disorders: substrates, differential diagnosis, and management. St. Louis, MO: Elsevier Mosby; 2005.

Enderby, P., Palmer, R. Frenchay dysarthria assessment. 2. Austin, TX: Pro-Ed; 2008.

Engen, T. Psychophysics: II. Scaling methods. In: Kling, JW., Riggs, LA., editors. Woodworth & Schlosberg's experimental psychology. 3. New York: Holt, Rinehart and Winston Inc; 1971. p. 47-86.

Folstein MF, Folstein SE, McHugh PR. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research. 1975; 12:189–198. [PubMed: 1202204]

Frearson B. A comparison of the AIDS sentence list and spontaneous speech intelligibility scores for dysarthric speech. Australian Journal of Human Communication Disorders. 1985; 13:5–21.

Georgiou N, Bradshaw JL, Iansek R, Phillips JG, Mattingley JB, Bradshaw JA. Reduction in external cues and movement sequencing in Parkinson's disease. Journal of Neurology, Neurosurgery, and Psychiatry. 1994; 57:368–370.

Goldman-Eisler F. A comparative study of two Hesitation phenomena. Language and Speech. 1961; 4:18–26.

Hustad K. Estimating the intelligibility of speakers with dysarthria. Folia Phoniatrica. 2006; 58:217–228.

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

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

Hustad K, Jones T, Dailey S. Implementing speech supplementation strategies: effects on intelligibility and speech rate of individuals with chronic severe dysarthria. Journal of Speech, Language and Hearing Research. 2003; 46:462–474.

Kempler D, Van Lancker D. Effect of speech task on intelligibility in dysarthria: a case study of Parkinson's disease. Brain and Language. 2002; 80:449–464. [PubMed: 11896652]

Kent RD, Weismer G, Kent JF, Rosenbek JC. Toward phonetic intelligibility testing in dysarthria. Journal of Speech and Hearing Research. 1989; 54:482–499.

Milenkovic, P. TF32 computer program. Madison, WI: University of Wisconsin-Madison; 2002.

McHenry M, Parle A. Construction of a set of unpredictable sentences for intelligibility testing. Journal of Medical Speech-Language Pathology. 2006; 14:269–271.

Sarno M. Speech impairment in Parkinson's disease. Archives of Physiological Medicine and Rehabilitation. 1968; 49:269–275.

Schiavetti, N. Scaling procedures for the measurement of speech intelligibility. In: Kent, RD., editor. Intelligibility in speech disorders: Theory, measurement, and management. Amsterdam: John Benjamins; 1992. p. 11-34.

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.

Tjaden K, Wilding GE. Effect of rate reduction and increased loudness on acoustic measures of anticipatory coarticulation in multiple Sclerosis and Parkinson Disease. Journal of Speech, Language, and Hearing Research. 2005; 48:261–277.

Tjaden K, Wilding G. The impact of rate reduction and increased loudness on fundamental frequency characteristics in dysarthria. Folia Phoniatrica. in press.

Weismer, G. Articulatory characteristics of Parkinsonian dysarthria: segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination. In: McNeil, M.Rosenbek, J., Aronson, A., editors. The dysarthrias: psychology, acoustics, perception, management. San Diego, CA: College-Hill Press; 1984. p. 191-130.

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

Yorkston KM, Beukelman DR. A comparison of techniques for measuring intelligibility of dysarthric speech. Journal of Communication Disorders. 1978; 11:499–512. [PubMed: 739065]

Yorkston, KM., Beukelman, DR. Assessment of intelligibility of dysarthric speech. Tigard, OR: CC Publications; 1981.

Yorkston, KM., Beukelman, DR., Hakel, M. The speech intelligibility test. Lincoln, NE: Tice Technology Services; 1996.

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.

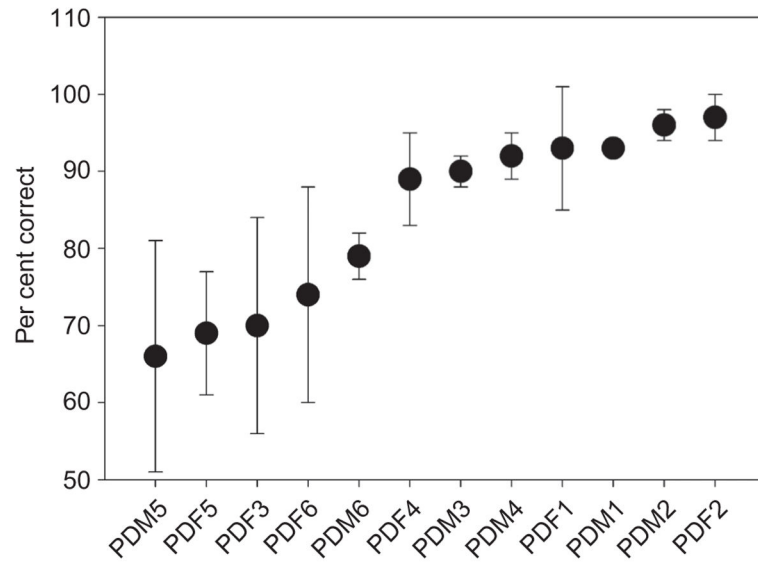**Figure 1.**

Average intelligibility scores for orthographic transcription of the paragraph reading task are reported. Vertical bars indicate one standard deviation. Speakers are arranged from least to most intelligible.
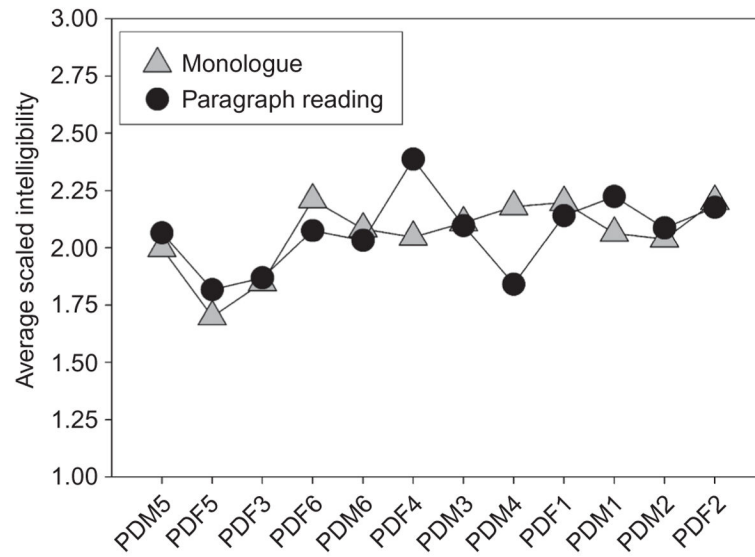
Figure 2.
Average scaled estimates of intelligibility are reported for the paragraph reading and monologue tasks. Speakers are arranged in the order of least to most intelligible, based on per cent correct scores for orthographic transcription of the paragraph reading task (see also Figure 1).
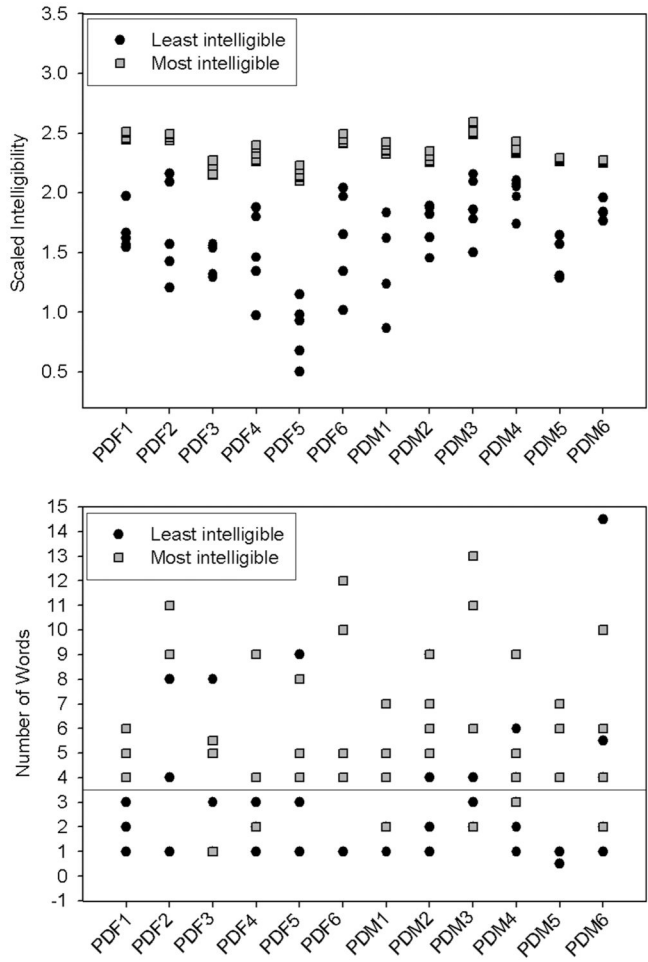
Figure 3.
Results of the within-speaker qualitative analysis comparing structural characteristics of the least and most intelligible speech runs for the monologue task. Each symbol corresponds to a speech run. The upper panel reports scaled estimates of intelligibility for speech runs and the bottom panel reports word counts for each speech run. Note in the bottom panel that the majority of the least intelligible speech runs contained fewer than four words, as indicated by the relative proportion of filled circles falling below the horizontal line.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table I

Speaker characteristics are reviewed.

Subject code	Age	Years post diagnosis	Dysarthria type	Dysarthria severity	Deviant perceptual characteristics
PDF1	42	6	Hypokinetic	Moderate	Monoloud, reduced loudness, variable rate
PDF2	62	3	Hypokinetic	Mild	Imprecise consonant, slow rate
PDF3	50	3	Hypokinetic	Moderate/severe	Hypernasal, imprecise consonant, short rushes
PDF4	72	9	Hypokinetic	Moderate	Reduced loudness, variable rate, short rushes
PDF5	81	3	Hypokinetic	Severe	Repeated phonemes, low pitch, reduced loudness
PDF6	45	13	Hypokinetic	Moderate/severe	Fast rate, breathy voice, mono loud
PDM1	69	12	Hypokinetic	Moderate	Mono pitch, mono loud, reduced stress
PDM2	74	1	Hypokinetic	Mild	Breathy, low pitch, slow rate
PDM3	72	4	Hyperkinetic	Mild	Harsh, forced inspiration/expiration, low pitch
PDM4	64	17	Hypokinetic	Moderate	Mono pitch, mono loud, short rushes
PDM5	60	8	Hypokinetic	Moderate	Breathy, short rushes, repeated phonemes
PDM6	64	8	Hypo/hyperkinetic mild/moderate	Breathy, fast rate, voice stop pages	

Note: Perceptual judgments reflect the consensus of three speech–language pathologists (see text for details).

Table II

Stimuli characteristics are summarised.

	Total number	Mean number runs per speaker (*SD*) (Interquartile range)	Mean number words per run (*SD*) (Interquartile range)	Mean number Syllables per run (*SD*) (Interquartile range)
Paragraph reading	411	34 (14) (27–43)	5.7 (4.0) (3–7)	7.1 (5.2) (4–9)
Monologue	463	39 (8) (31–43)	5.8 (5.2) (2–7.5)	7.4 (6.7) (3–10)

Table III

Spearman rank order correlations relating the number words per speech run and scaled estimates of intelligibility are reported.

	Paragraph reading	**Monologue**
PDM5	0.26	0.**31**
PDF5	0.**31**	0.26
PDF3	0.32	−0.14
PDF6	0.44	0.**49**
PDM6	0.22	0.14
PDF4	−0.03	0.20
PDM3	0.37	0.**41**
PDM4	0.19	0.16
PDF1	0.**37**	0.30
PDM1	0.18	0.**48**
PDM2	0.08	0.**38**
PDF2	0.29	0.**62**

Note: Significant correlations ($p < 0.05$) are indicated in bold. Speakers are arranged from least to most intelligible, according to per cent correct scores for orthographic transcription of the reading passage (see Figure 1).