

Research Article

Visual Analog Scale Ratings and Orthographic Transcription Measures of Sentence Intelligibility in Parkinson's Disease With Variable Listener Exposure

Defne Abur,^a Nicole M. Enos,^{b,c} and Cara E. Stepp^{a,c,d}

Purpose: While orthographic transcription (OT) is the gold standard for measures of intelligibility, it is relatively inaccessible to clinicians. This study investigates the relationship between visual analog scale (VAS) ratings and OT measures of intelligibility for speakers with Parkinson's disease (PD), with the eventual goal of developing more clinically feasible assessments of intelligibility.

Method: Twenty speakers with PD and 5 controls read 11 sentences. First, 33 listeners completed an OT task using 1 sentence from each speaker. An additional 33 listeners rated the intelligibility of 1 sentence from each speaker using a VAS, reflecting a minimized exposure VAS (MEV) task.

Lastly, 14 additional listeners each rated the intelligibility of all 11 sentences produced by all speakers using a VAS, reflecting an extended exposure VAS (EEV) task. Smaller listener groups were simulated from each VAS task for comparison to scores from the OT task.

Results: There was a strong relationship between OT and both MEV and EEV. This relationship remained strong ($R^2 \geq .82$) even when only 1 listener in MEV and 2 listeners in EEV were simulated per sentence.

Conclusions: VAS ratings may be a suitable alternative to OT measures of sentence intelligibility for PD using listeners with both minimal and extended exposure to the stimuli.

*I*ntelligibility can be defined as “the degree to which the speaker’s intended message is recovered by the listener” (Kent, Weismer, Kent, & Rosenbek, 1989). *Intelligibility*, thus, defines a speaker’s ability to be understood by the people with whom he or she is speaking. Loss of intelligibility can severely impact daily activities that require communication with family, friends, or colleagues and can be detrimental in an emergency in which communication with medical professionals is necessary. Therefore, intelligibility is an important measurement of the severity of an individual’s speech disorder and its progression and

can be used as a benchmark to assess the effectiveness of treatment (Miller, 2013).

The measurement of intelligibility throughout speech therapy is particularly relevant for individuals with degenerative, progressive conditions that affect speech, such as individuals with Parkinson’s disease (PD). PD is one of the most prevalent neurodegenerative conditions, with a lifetime risk of about 2% (Fahn, 2003; Pringsheim, Jette, Frolkis, & Steeves, 2014). Over 90% of individuals with PD develop hypokinetic dysarthria, a motor speech disorder characterized by reduced speech loudness, diminished prosody, and disruptions in articulation (Duffy, 2013; Logemann, Fisher, Boshes, & Blonsky, 1978), which can substantially reduce speech intelligibility (Miller et al., 2007) and quality of life (Miller, Noble, Jones, Allcock, & Burn, 2008). The current work focuses on the intelligibility of speakers diagnosed with PD.

Despite the clear need for measures of intelligibility, current “gold standard” practices for measuring intelligibility (through orthographic transcription [OT]; Duffy, 2013; Miller, 2013; Stipancic, Tjaden, & Wilding, 2016) are largely inaccessible to speech-language pathologists (SLPs) due to their previous exposure with the patient’s dysarthric speech

^aDepartment of Speech, Language, and Hearing Sciences, Boston University, MA

^bDepartment of Computer Engineering, Boston University, MA

^cDepartment of Biomedical Engineering, Boston University, MA

^dDepartment of Otolaryngology–Head and Neck Surgery, Boston University School of Medicine, MA

Correspondence to Cara E. Stepp: cstepp@bu.edu

D. Abur and N. E. Enos contributed equally to this work.

Editor-in-Chief: Julie Barkmeier-Kraemer

Editor: Nancy Solomon

Received November 26, 2018

Revision received January 30, 2019

Accepted March 28, 2019

https://doi.org/10.1044/2019_AJSLP-18-0275

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

(Beukelman & Yorkston, 1980), time constraints in the clinic, lack of access to resources, and cost (Gurevich & Scamihorn, 2017). OT involves playing a recording of a specific speaker's production of a known stimulus to a naïve listener, who transcribes to the best of their ability what they heard. Several listeners are necessary to mitigate the variable responses across listeners (McHenry, 2011), and listeners should be naïve to the stimuli and the speakers themselves: Listeners who are exposed to a speaker or that speaker's dysarthria are biased toward better performance compared to unfamiliar listeners (Beukelman & Yorkston, 1980; DePaul & Kent, 2000; D'Innocenzo, Tjaden, & Greenman, 2006; Hustad & Cahill, 2003; H. Kim, 2015; H. Kim & Nanney, 2014; Liss, Spitzer, Caviness, & Adler, 2002; Tjaden & Liss, 1995; Utianski, Lansford, Liss, & Azuma, 2011). Intelligibility is calculated based on the number of words from the target stimulus correctly identified in each listener's transcription, averaged across listeners. Despite relying on listeners' perceptions, this technique is considered "objective," since listeners are not asked to make judgments about what they hear and since the resulting scores are quantitative. However, the transcription process is time-consuming and laborious for listeners, as they must listen carefully to accurately transcribe what they heard, as opposed to making a general assessment of the speech. Also, if listeners have prior knowledge of the speaker's dysarthria (such as SLPs) or if naïve listeners complete longer tasks, exposure to the rated stimuli can affect intelligibility measurements (Beukelman & Yorkston, 1980; Hoover, Reichle, Van Tasell, & Cole, 1987; Venkatagiri, 1994).

Given that exposure to a given speaker's dysarthria can affect measures of intelligibility, previous research studies have investigated intelligibility using task designs in which listeners have limited exposure to a speaker's stimuli (Garcia & Dagenais, 1998; Huttunen & Sorri, 2004; Stipancic et al., 2016). However, this is not easily translatable to the clinical setting where clinicians have experience with their client's dysarthric speech, it is difficult to access large groups of naïve listeners who lack experience with specific speech stimuli, and there is limited time. Even with access to large groups of listeners, the speech stimuli required for intelligibility measures may also be expensive to obtain, further impeding the use of this assessment by SLPs. In fact, a recent study found that 35% of practicing SLPs lacked access to standardized intelligibility measures, with 66% of them reporting that cost was the inhibiting factor (Gurevich & Scamihorn, 2017). Moreover, OT requires important decisions when interpreting results; many studies have used a key word scoring paradigm, which only scores the correctness of key words such as subjects and verbs (Hustad, 2006; Stipancic et al., 2016), while others score every word in each sentence (Garcia & Dagenais, 1998; Hustad, 2006; Hustad, Jones, & Dailey, 2003). Additionally, although OT shows good inter- and intrarater reliability in research settings, it is infrequently used in clinics (Miller, 2013).

Another method for assessing intelligibility that has recently shown promise as an alternative to OT is visual analog scale (VAS) rating (Adams, Dykstra, Jenkins, & Jog, 2008; Stipancic et al., 2016). VAS rating involves a listener hearing a speech stimulus from a speaker and proceeding to rate the perceived intelligibility of the speech on a 100-mm or 150-mm scale (Tjaden, Sussman, & Wilding, 2014). While OT tasks can capture specific phonetic information, VAS ratings provide an overall estimate of speech intelligibility for each speech sample. VAS is also faster for the listener to complete (no written transcription is required) since the intelligibility score is simply the value selected on the scale by the listener. VAS ratings could be a cost-effective and efficient alternative to OT for clinical and research settings. Specifically, VAS may be useful if costs and access to the requisite number of listeners make OT-based measures infeasible. Several recent studies have demonstrated a strong correlation between VAS scores and OT-based measures of intelligibility when listeners are naïve to dysarthric speech and also have minimal exposure to task stimuli (Adams et al., 2008; Huttunen & Sorri, 2004; Stipancic et al., 2016; Tjaden, Kain, & Lam, 2014). While recruiting naïve listeners for short tasks is desirable to reduce listener bias, naïve listener recruitment remains an obstacle in acquiring both VAS and OT-based measures of intelligibility in a clinical setting. If the relationship between VAS scores and OT-based measures of intelligibility still holds when listeners have exposure to a speaker's dysarthria, this could make listener recruitment easier in a clinical setting. Thus, to determine the feasibility of implementing VAS ratings in a clinical setting, it is vital to first examine the relationship between VAS ratings and the OT while varying number of listeners and task length to quantify the minimum number of listeners and degree of speech sample exposure required to obtain reliable intelligibility estimates with VAS ratings.

Previous findings have shown that OT measures of intelligibility of dysarthric speech during a reading passage were comparable between SLPs and initially naïve listeners who were given extended exposure to speech samples (Beukelman & Yorkston, 1980). Therefore, it is important to characterize whether the relationship between OT and VAS measures of intelligibility may change for listeners with more experience with the speech stimuli in the task. The goal of the current study is to build on existing knowledge about the relationship between VAS ratings and OT-based intelligibility scores through an evaluation of the impact of the number of listeners and duration of exposure to the dysarthric speech stimuli.

Three experiments were conducted in the current work: (a) 33 naïve listeners completed an OT task, (b) 33 naïve listeners completed a minimized exposure VAS (MEV) task, and (c) 14 listeners completed an extended exposure VAS (EEV) task. We hypothesized the following: (a) Under an ideal, minimized exposure condition, there would be a strong relationship between VAS ratings of intelligibility and OT-based measures of intelligibility. (b) Under an ideal, minimized exposure condition, the strength of the

relationship between VAS ratings of intelligibility and OT-based measures of intelligibility would increase with increasing numbers of listeners. (c) Under a nonideal, EEV task, the strength of the relationship between VAS ratings of intelligibility and OT-based measures of intelligibility would remain strong. (d) Under a nonideal, EEV task, the strength of the relationship between VAS ratings of intelligibility and OT-based measures of intelligibility would increase with increasing numbers of listeners. Listener characteristics, task information, as well as intra- and interlistener reliability are summarized in Figure 1.

Method

Speakers

All speakers were selected from a speech database of 126 older adults with idiopathic PD and 167 older adults without idiopathic PD, as diagnosed by a neurologist. All participants were native speakers of Standard American English. Speakers had no history of speech, language, or hearing impairments other than those associated with PD. Participants were excluded from the selection process if (a) speakers had non-Standard American English accents ($n = 26$); (b) speakers had less than 11 unique sentences recorded ($n = 25$); (c) speakers did not have disease severity ratings using the Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS; Goetz et al., 2008) at the time of recording ($n = 23$); and (d) speakers had concurrent symptoms that affected vocal quality unrelated to PD ($n = 12$). For the remaining 40 speakers, speech recordings were rated by three researchers in the speech, language, and hearing sciences department on a scale of 0 (*no speech symptoms*) to 4 (*severe speech symptoms*). Additionally, the scores from the MDS-UPDRS (Goetz et al., 2008) Part III, administered by a physical therapist or speech researcher with an MDS-UPDRS training certificate, were used to determine PD motor symptom severity. The MDS-UPDRS Part III assessment consists of various motor tasks designed to characterize motor function on both sides of the body. Using the speech ratings and MDS-UPDRS Part III scores, 20 speakers were selected to

represent a wide range of PD symptom severity and degree of speech intelligibility for both male and female speakers, resulting in a group of 12 male speakers with PD and eight female speakers with PD. For speakers without idiopathic PD, five speakers were selected with the same proportion of male to female speakers and as closely age-matched as possible to the speakers with PD. This resulted in three male speakers without PD and two female speakers without PD. Speaker characteristics and MDS-UPDRS Part III scores are listed in Table 1. The MDS-UPDRS Part III scores in the PD group included mild (32 points or less), moderate (58 points or less), and severe impairment (59 points or more) in motor function (Martínez-Martín et al., 2015). All participants completed informed consent in compliance with either the University of Washington Institutional Review Board or the Boston University Institutional Review Board.

Speech Recordings

All speakers produced a novel set of 11 sentences generated from the Sentence Intelligibility Test (Yorkston, Beukelman, & Tice, 1996), which contains a database of 1,100 sentences designed to give low linguistic context to the listener, ranging from five to 15 words per sentence. This resulted in a total of 275 unique sentences.

Sound Presentation

For all auditory perceptual tasks, stimuli were presented in a sound-treated booth. The acoustic signal for each sentence was normalized by the average root-mean-square, such that each stimulus had the same average sound pressure level during presentation. Stimuli were presented in multispeaker babble using overear Sennheiser 280 Pro headphones (Sennheiser Electronic Corporation). The headphones were calibrated such that the majority of speech energy was presented between 60 and 70 dB SPL for each stimulus to mirror conversational speech loudness (Olsen, 1998). The multispeaker babble consisted of four healthy male speakers and four healthy female speakers who were not included in the speaker data set. The multispeaker babble was incorporated into the recordings at a signal-to-noise ratio (SNR)

Figure 1. Number of listeners, number of sentences heard per listener, task completion time, and intrarater and interrater reliability are listed for each of the three study tasks with a schematic of the experiment setup.

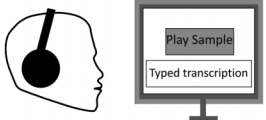
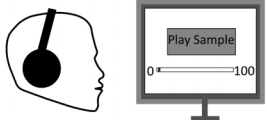
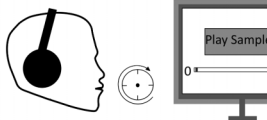
	Orthographic Transcription (OT)	Minimized Exposure VAS (MEV)	Extended Exposure VAS (EEV)
Experimental Task			
Listeners	33 listeners (17 female, 16 male)	33 listeners (16 female, 17 male)	14 listeners (7 female, 7 male)
Number of Sentences	28 sentences	28 sentences	308 sentences
Task Completion Time	6–15 min ($M = 9$ min, $SD = 3$ min)	4–10 min ($M = 5$ min, $SD = 1$ min)	40–83 min ($M = 60$ min, $SD = 15$)
Intrarater Reliability	.76–1 ($M = .93$, $SD = .07$)	.59–1 ($M = .85$, $SD = .11$)	.79–0.96 ($M = .86$, $SD = .05$)
Interrater Reliability	.42–1 ($M = .86$, $SD = .12$)	.48–1 ($M = .83$, $SD = .14$)	.62–0.92 ($M = .76$, $SD = .06$)

Table 1. Speaker characteristics for speakers for all speakers with Parkinson's disease (PD) and control speakers included in the study.

Subject	Age	Sex	MDS-UPDRS Part III (Motor)	Group
PD01	69	M	9	PD
PD02	70	F	9	PD
PD03	70	M	16	PD
PD04	70	M	26	PD
PD05	69	M	32	PD
PD06	68	F	40.5	PD
PD07	60	F	41	PD
PD08	56	M	42	PD
PD09	68	F	42.5	PD
PD10	58	M	43	PD
PD11	58	M	44	PD
PD12	59	M	47	PD
PD13	69	F	48	PD
PD14	71	M	49	PD
PD15	74	M	49	PD
PD16	68	F	51	PD
PD17	74	M	51.5	PD
PD18	82	F	55	PD
PD19	59	M	58	PD
PD20	77	F	85	PD
C01	61	M	—	Control
C02	65	M	—	Control
C03	70	M	—	Control
C04	69	F	—	Control
C05	65	F	—	Control

Note. Speakers with Parkinson's disease are numbered in order of increasing motor severity as measured by Part III of the Movement Disorder Society–sponsored revision of the Unified Parkinson's disease Rating Scale (MDS-UPDRS). M = male; F = female.

of 0 dB, meaning that the level of speech energy was equal to the level of multispeaker babble energy in the stimuli. This SNR was determined during pilot testing with both VAS and OT procedures to reduce ceiling effects on intelligibility scores.

Listeners

A total of 80 listeners were recruited from Boston University and were native speakers of Standard American English. All listeners denied a history of speech, language, or hearing disorders. Prior to the experiment, each listener passed a bilateral hearing screening using a threshold of 25 dB HL at 125, 250, 500, 1000, 2000, 4000, and 8000 Hz (protocol based on the American Speech-Language-Hearing Association, 2005). All listeners completed informed consent in compliance with the Boston University Institutional Review Board.

OT Task

A total of 17 women (age = 18–24 years, $M = 20.3$, $SD = 1.8$) and 16 men (age = 18–28 years, $M = 21.1$, $SD = 3.2$) completed the OT task. For the OT task, listeners were instructed to listen to 28 stimuli once and to transcribe them to the best of their ability without using any punctuation. These 28 stimuli included one randomly

selected stimulus from each speaker (20 stimuli from speakers with PD and five stimuli from control speakers) and three randomly selected sentences that were repeated to assess intralistener reliability (12% of all stimuli). Thus, each listener transcribed at least one sentence stimulus from each speaker, and each of the 11 sentences from each speaker's stimuli was transcribed by three independent listeners (Lippmann, 1997; Miyamoto et al., 1997). A custom MATLAB script (MathWorks, 2016) displayed a user interface that allowed the participant to listen to the stimulus mixed with multispeaker babble and type their transcription into a box on the screen before submitting their transcription. Listeners completed one listening session that had an average duration of 9 min. Intrarater reliability was calculated as the absolute value of the difference in the number of words matching phonemically between the first transcription and the second transcription divided by the total number of words in the sentence. The intrarater reliability for the repeated sentences was averaged for each listener. Since not all listeners rated the same sentences, interrater reliability was also calculated at a sentence-level first (across the three unique listeners) and averaged for all sentences. Thus, the interrater reliability was calculated by determining the standard deviation of the mean of the three listener scores for each sentence, averaging across all 275 sentences, and subtracting the final score from perfect agreement (a score of 1).

MEV Task

A total of 16 women (age = 18–22 years, $M = 20.0$, $SD = 1.2$) and 17 men (age = 18–27 years, $M = 21$, $SD = 2.8$) completed the MEV task. For the MEV task, listeners were instructed to listen to 28 stimuli once and rate the perceived intelligibility of each stimulus on a sliding scale from 0% to 100% intelligible. These 28 stimuli included one randomly selected stimulus from each speaker (20 stimuli from speakers with PD and five stimuli from control speakers) and three randomly selected sentences that were repeated to assess intralistener reliability (12% of all stimuli). Thus, each listener rated at least one sentence stimulus from each speaker using VAS, and each of the 11 sentences from each speaker's stimuli were rated by three independent listeners. Intelligibility was described as the “degree to which speech is understood” (Kent et al., 1989), and participants were told that a score of 0% indicated that they could not understand any of the stimulus at all, whereas a score of 100% indicated that they understood the stimulus perfectly. After listening to each stimulus mixed with multispeaker babble, listeners selected a point on a 100-mm computerized VAS as their intelligibility score using a custom MATLAB script. Listeners were allowed to listen to each stimulus only once and made a rating on the scale. They were not allowed to change previous submissions after moving on to the next stimulus. Listeners completed one listening session that had an average duration of 5 min.

In order to calculate intrarater reliability in a way that would be comparable to the OT task, first the difference

between the first and second VAS rating was found. The absolute value of this difference was subtracted from 100 (the full range of the scale), and the result was divided by 100. The resulting unitless values could range from 0 (*poor reliability*, with the two scores maximally differing) to 1 (*excellent reliability*, with the two scores identical). The intrarater reliability for the repeated stimuli was averaged for each listener. Similar to the OT task, interrater reliability was calculated at a sentence-level first (across the three unique listeners) and averaged for all sentences. The mean of the three listener scores for each sentence were first normalized to be values from 0 to 1 (using the same method as for intrarater reliability). The interrater reliability was calculated by determining the standard deviation of the mean of the three listener scores for each sentence, averaging across all 275 sentences, and subtracting the final score from perfect agreement (a score of 1).

EEV Task

A total of seven women (age = 18–27 years, $M = 22.1$, $SD = 3.5$) and seven men (age = 18–21 years, $M = 19.0$, $SD = 1.0$) who did not participate in the OT or MEV tasks completed the EEV task. For the EEV task, listeners were instructed to listen to all 11 sentences from each speaker (220 sentences from speakers with PD and 55 sentences from control speakers) and 33 randomly selected sentences that were repeated to assess intralister reliability (12% of all sentences), for a total of 308 sentences to reflect an EEV task. The methods were identical to the VAS rating for the MEV task, except for the number of listeners and the number of ratings per listener. Listeners completed one listening session that had an average duration of 60 min, with breaks as necessary. Intrarater reliability was calculated using the same methods as in the MEV task. Interrater reliability was calculated to match the methods used in the OT and MEV tasks, but across all 14 listeners (since they all rated the same 275 sentence set). The mean of the 14 listener scores for each sentence was first normalized to be values from 0 to 1 (using the same method done for intrarater reliability). The interrater reliability was calculated by determining the standard deviation of the mean of the 14 listener scores for each sentence, averaging across all 275 sentences, and subtracting the final score from perfect agreement (a score of 1).

Analysis

For the OT task, intelligibility was calculated as the percentage of words matching phonemically between the listener transcription and the sentences, averaged over three independent listeners. The percent intelligibility for each sentence was initially calculated using a custom MATLAB script, then each participant's transcriptions were hand-checked by the first and second author to include *homonyms* (defined as words that sound the same but are spelled differently, i.e., “their” and “there”) and *misspellings* (defined as words that contained an extra or missing letter that did

not change the tense of the word). The analysis time took a total of approximately 15 min for each participant. These final intelligibility scores were averaged by speaker and task to yield a single OT-based intelligibility score for each speaker. For the MEV and EEV tasks, intelligibility was identified as the raw number on the scale (from 0 to 100) selected by the user. These intelligibility scores were averaged by speaker and task to yield two VAS-based average intelligibility scores for each speaker (one from the MEV task and one from the EEV task). Stimuli that were repeated for reliability analysis were not included in this average and were only used for determining reliability.

To address Hypothesis 1, that OT and MEV intelligibility scores will be strongly related, the strength of the relationship between OT-based measures of intelligibility from the OT task and the MEV ratings were assessed. A least squares regression was used to identify the strength and fit of a linear relationship between transcription-based and MEV-based scores of intelligibility for the speakers.

To address Hypothesis 2, that fewer listeners in the MEV task would reduce the strength of the relationship between the OT task and the MEV task intelligibility scores, a custom MATLAB algorithm was written to simulate all possible smaller groups of listeners in the MEV task for each sentence. The MEV experiment involved each sentence being rated by three listeners (Lippmann, 1997; Miyamoto et al., 1997), so simulated smaller groups of one or two listeners per sentence were used. For example, for the one listener condition, Listener 1, 2, or 3 might be used; for two listeners, Listeners 1 and 2, Listeners 2 and 3, or Listeners 1 and 3 might be used. A total of 11 sentences were used for each speaker, which resulted in a total of 3^{11} or 177,147 possible combinations each for the one or two listeners per sentence conditions. For each simulated listener set, speaker-averaged VAS-based scores were computed. Least squares regressions were performed to evaluate the strength of the relationship between transcription-based and MEV-based scores of intelligibility for the simulated smaller listener groups. An α of .05 was set to be statistically significant. The R^2 values for each smaller simulated group were analyzed to determine a mean and standard deviation of R^2 values.

For Hypothesis 3, that EEV intelligibility scores would also show a strong relationship with OT intelligibility scores, the OT task scores were compared to the EEV task scores. Least squares regression was used to identify the strength and fit of a linear relationship between transcription-based and EEV-based intelligibility scores.

To test Hypothesis 4, that a reduced number of listeners for the EEV task would reduce the strength of the relationship between the EEV task and the OT task, the customized MATLAB algorithm used for Hypothesis 2 was adjusted to simulate all possible smaller groups of listeners (groups of size $n = 1$ to $n = 14$) for the EEV task. For each simulated listener set, speaker-averaged VAS-based intelligibility scores were computed. Least squares regressions were performed to evaluate the strength of the relationship between VAS-based scores from the EEV and

transcription-based scores from the OT task for the simulated smaller listener groups. An α of .05 was set to be statistically significant. The R^2 values for each smaller simulated group were analyzed to determine a mean and standard deviation of R^2 values. In order to compare the results of the EEV task simulation to acceptable margins of error for clinical intelligibility estimates, the percent deviation in intelligibility estimates was calculated for each number of listeners with respect to the minimal possible margin of error (the average for all 14 listeners). Prior work investigating speech intelligibility in a small sample of speakers ($n = 8$) with PD pre- and post-Lee Silverman Voice Therapy (LSVT) treatment reported a significant increase in mean intelligibility of 7.06% for six of the eight speakers (Cannito et al., 2012). This increase in intelligibility (7.06%) was used as a reference to approximate what might be a significant difference in percent intelligibility for the current work. Finally, a regression model was used to determine whether the relationship between the OT task and the MEV task was significantly different than the relationship between the OT task and the EEV task, using the OT task as the input variable and the MEV and EEV tasks as the output variables.

Results

Overall Scores of Intelligibility

For the OT task, speakers with PD had transcription-based intelligibility scores ranging from 32.7% to 94.1% ($M = 68.6\%$, $SD = 16.8\%$), and control speakers had transcription-based intelligibility scores ranging from 75.7% to 91.6% ($M = 81.8\%$, $SD = 5.9\%$). For the MEV task, speakers with PD had VAS-based intelligibility scores ranging from 20.1% to 94.5% ($M = 65.0\%$, $SD = 21.6\%$), and control speakers had intelligibility scores ranging from 73.5% to 94.6% ($M = 82.9\%$, $SD = 8.5\%$). For the EEV task, speakers with PD had VAS-based intelligibility scores ranging from 30.0% to 89.2% ($M = 64.8\%$, $SD = 16.7\%$), and control speakers had intelligibility scores ranging from 72.8% to 85.5% ($M = 78.5\%$, $SD = 5.1\%$). The intelligibility scores for each task are summarized in Table 2.

OT Versus MEV Intelligibility Scores

The relationship between the scores from the OT task and the MEV intelligibility scores was investigated. There was a strong, positive relationship between scores from the OT task and the MEV task, yielding an R^2 value of .886 ($p < .001$; see Figure 2). Smaller listener groups were simulated from the VAS-based intelligibility scores to identify how the strength of this relationship changed with varying numbers of listeners per sentence for the MEV task. These scores were averaged for each speaker and compared to the speaker-averaged transcription scores from the OT task. Linear regressions were performed with the simulated VAS-based intelligibility scores, revealing that R^2 increased with increasing number of listeners in the MEV task. The average R^2 values ranged from .835 for one listener per sentence to .886 for three listeners per sentence (see Figure 3).

OT Versus EEV Intelligibility Scores

The relationship between the scores from the OT task and the EEV intelligibility scores was investigated. There was a strong, positive relationship between scores from the OT task and the EEV task, yielding an R^2 value of .892 ($p < .001$; see Figure 2). A regression model was used to compare the relationship between the OT and MEV tasks and the relationship between the OT and EEV tasks (regressions shown in Figure 2) with the OT task scores as the input and the MEV and EEV task scores as outputs. The model revealed a statistically significant difference between the MEV and EEV task regression line constants ($p = .028$) and slopes ($p = .018$). The slope for the MEV task shows, overall, that the MEV task ratings tended to underestimate the OT task measures of intelligibility for low intelligibility scores ($< 50\%$ intelligible). However, for higher intelligibility scores ($> 50\%$ intelligible), there was more variability in the MEV rating estimation of the OT task scores. The EEV task ratings of intelligibility were very good estimates of the OT task scores, showing a trend for slight underestimation of the OT task measures of intelligibility across all intelligibility levels.

Smaller listener groups were simulated from the EEV task intelligibility scores to identify how the strength of this relationship changed with varying numbers of listeners per sentence. These scores were averaged for each speaker and compared to the speaker-averaged transcription scores from the OT task. Linear regressions were performed with the simulated VAS-based intelligibility scores, revealing that R^2 increased with increasing number of listeners in the EEV task. The average R^2 values ranged from .748 for one listener per sentence to .894 for 14 listeners per sentence (see Figure 3). The percent deviation in average intelligibility estimates, in relation to the average for all 14 listeners, decreased as a function of the number of listeners, ranging from 8.03% for one listener to 0.41% for 13 listeners (see Figure 4).

Discussion

The purpose of this study was to further investigate the relationship between VAS-based and transcription-based intelligibility scored with varying lengths of listening tasks and varying numbers of listeners. The current work found a strong, linear relationship between OT and VAS-based scores of intelligibility, regardless of whether the task involved MEV ratings or EEV ratings. This finding supports Study Hypotheses 1 and 2 (that the OT scores would show a strong relationship with both MEV and EEV ratings, respectively). Simulations of smaller listener groups revealed that both the MEV task and EEV task yielded speaker-averaged VAS scores that were strongly correlated to the OT task scores. Additionally, the strength of this relationship increased with increasing numbers of listeners used for both the MEV and EEV tasks, supporting Study Hypotheses 2 and 4 (that a larger number of listeners would increase the strength of the relationship between the OT task scores compared to the MEV and EEV ratings,

Table 2. Intelligibility scores for speakers with Parkinson's disease (PD) and control speakers (Control) for each experimental task.

Measure	Intelligibility scores (PD)		Intelligibility scores (Control)	
	Range	<i>M</i> (<i>SD</i>)	Range	<i>M</i> (<i>SD</i>)
Orthographic transcription	32.7%–94.1%	68.6% (16.8%)	75.7%–91.6%	81.8% (5.9%)
Minimized exposure visual analog scale	20.1%–94.5%	65.0% (21.6%)	73.5%–94.6%	82.9% (8.5%)
Extended exposure visual analog scale	30.0%–89.2%	64.8% (16.7%)	72.8%–85.5%	78.5% (5.1%)

Note. *M* = mean across speakers, *SD* = standard deviation across speakers.

respectively). These results confirm works by Adams et al. (2008) and Stipancic et al. (2016), which showed a strong relationship between transcription and VAS-based scores of intelligibility and extended knowledge about this relationship by looking at the effect of decreasing numbers of listeners.

The OT task scores were compared to the MEV ratings and yielded a strong, linear relationship reaching a peak R^2 of .886 for all 25 speakers (20 speakers with PD and five control speakers). The strength of the linear relationship between the OT and MEV task scores is very comparable to the results of Stipancic et al. (2016), who found an R^2 of .89 between OT-based measures and VAS-based estimates of intelligibility for 78 speakers (32 control speakers, 30 speakers with Multiple Sclerosis, and 16 speakers with PD).

The OT task scores were also compared to the EEV task ratings and yielded a strong, linear relationship, yielding an R^2 of .892. Similarly, Adams et al. (2008) compared transcription-based intelligibility scores to VAS-based estimates of intelligibility for 25 speakers with PD and 15 control speakers using nonnaïve listeners (two listeners who were graduate students in speech-language pathology rated

all stimuli). Though the listener group differed from the current work, Adams et al. also found a strong relationship between OT-based measures and VAS-based estimates of intelligibility, reporting an R^2 of .89 for the control speakers and an R^2 of .93 for the speakers with PD.

While both the MEV and EEV tasks showed a strong relationship ($R^2 \geq .886$) with the OT task scores, a regression model revealed a significant difference in the best fit regression lines for the MEV and EEV tasks. The MEV task ratings tended to underestimate the OT task measures of intelligibility, especially for low intelligibility scores (in agreement with Stipancic et al., 2016). Similarly, Hustad (2006) found lower percent intelligibility ratings compared to OT-based measures of intelligibility for listener ratings of four speakers with chronic dysarthria and a medical diagnosis of cerebral palsy. The EEV task ratings showed better overall estimation of the OT scores compared to the estimates from the MEV task ratings. However, the EEV task ratings tended to slightly underestimate the OT-based measures of intelligibility across all intelligibility score levels. As expected, the simulations of smaller groups of listeners revealed that the strength of the relationship between the OT task scores and both MEV and EEV task scores increased with increasing numbers of listeners (see Figure 3). Notably, the average R^2 value for one listener in the MEV task simulation suggest that speaker-averaged VAS scores of intelligibility with a single naïve listener per sentence (11 listeners) can be used to predict OT-based scores of intelligibility ($R^2 = .835$, $SD = .026$).

In the EEV task, simulations revealed that even just two listeners demonstrated an average of R^2 of 82.2% ($SD = 5.4\%$) in predicting the OT task scores. Additionally, increasing the number of listeners in the EEV task from just one to two listeners per sentence had a large impact on both the average R^2 ($M = 0.748$, $SD = 0.097$ for one listener, $M = 0.822$, $SD = 0.054$ for two listeners; see Figure 3) and the degree of deviation in intelligibility estimates relative to the average for all 14 listeners (8.03% for one listener and 4.5% for two listeners; see Figure 4). Taken together with the results of Cannito et al. (2012), who report a 7.06% increase in intelligibility for six speakers with PD post-LSVT treatment, the results from the EEV task simulations suggest that even just two listeners could predict OT-based scores of intelligibility with $R^2 = 82.2\%$ ($SD = 5.2$), with a margin of error across speaker ratings able to capture what might be a significant change in intelligibility ($\geq 7.06\%$).

Figure 2. Average transcription-based intelligibility scores compared to average visual analog scale (VAS)-based intelligibility scores for control speakers (open markers) and speakers with Parkinson's disease (solid markers) for the minimized exposure VAS (MEV; magenta triangles) task and extended exposure VAS (EEV; blue circles) task. The perfect agreement between transcription and VAS scores is represented by a gray dotted line.

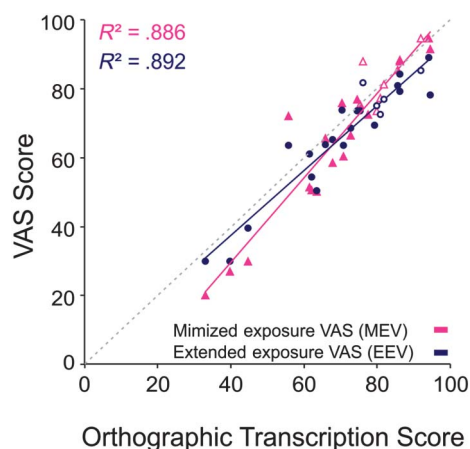
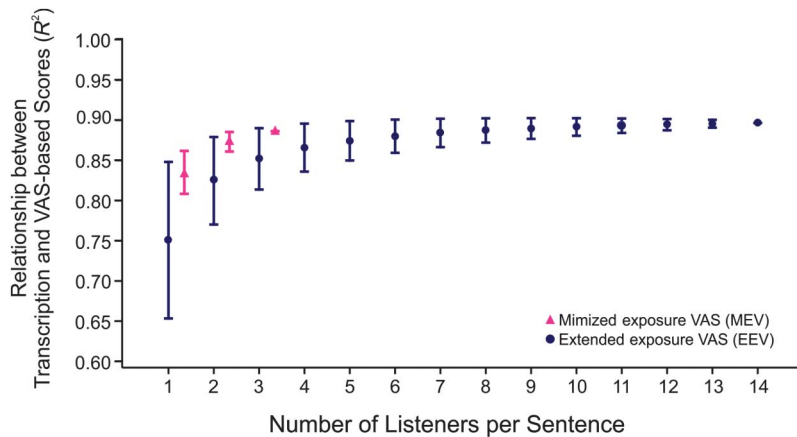


Figure 3. The strength of the relationship between the minimized exposure VAS (MEV; magenta triangles) task and extended exposure VAS (EEV; blue circles) task and the orthographic transcription scores shown as a function of simulated groups for number of listeners per sentence for MEV (varying from one to three listeners) and EEV (varying from one to 14 listeners) tasks. VA = visual analog scale.

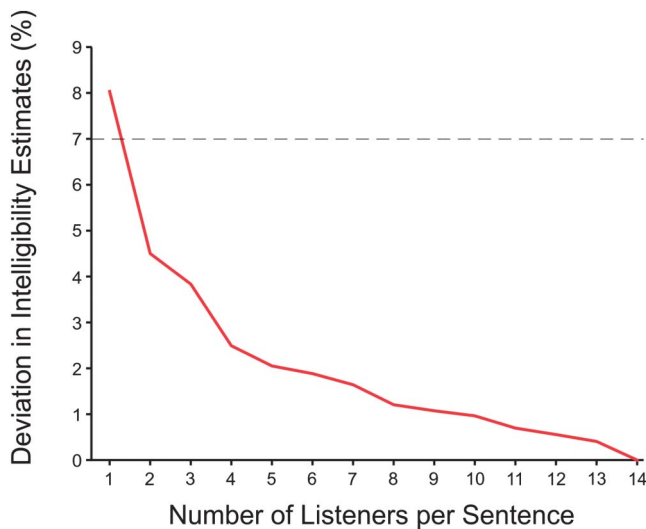


However, it is important to acknowledge the small sample size of the significant changes in intelligibility reported by Cannito et al. Thus, these results may not be representative for a larger sample of patients with PD.

Implications for Assessment of Intelligibility

The current work confirms that there is a strong relationship ($R^2 \geq .886$) between OT and VAS-based scores

Figure 4. The percent deviation in average intelligibility estimates (red solid line) in relation to the average for all 14 listeners for the extended exposure visual analog scale task simulation. The black dotted line denotes the average percent change in intelligibility (7.06%) reported for speakers with a small sample of patients with Parkinson's disease pre- and post-Lee Silverman Voice Therapy therapy in Cannito et al. (2012).



of intelligibility, in agreement with previous findings (Adams et al., 2008; Stipancic et al., 2016), while also investigating how this relationship changes with varying number of listeners. However, when weighing options for reliable intelligibility measures, it is also important to factor in the time necessary to recruit listeners, administer tasks, and analyze the intelligibility measures. A summary of the time required for the OT, MEV, and EEV tasks are listed in Figure 1.

The current work, using simulations of smaller listener groups, suggests that both MEV and EEV tasks can yield average intelligibility estimates that reliably ($R^2 \geq .82$) predict OT measures of intelligibility in less time than is required for OT. These results support that an MEV task could be the most time-effective application of intelligibility assessment for speakers with PD, with simulations of smaller listener groups showing one naïve listener per sentence can reach estimates of intelligibility that are strong predictors ($R^2 = .822$, $SD = .054$) of OT scores. However, accessibility to naïve listeners still remains an issue for the ultimate goal of clinical application of this type of VAS rating task. The EEV task was designed to resemble the experience of a listener familiar with dysarthric speech, since access to listeners with increased exposure may be easier than access to naïve listeners in a clinical setting. Simulations of smaller listener groups suggest that listeners with increased exposure to dysarthric speech could substantially improve the validity of their own single ratings by enlisting another listener with increased exposure to provide a second set of ratings. One listener's VAS ratings in the EEV task predicted 74.8% ($SD = 9.7\%$) of the OT-based measures of intelligibility, whereas two listeners increased the strength of the prediction to 82.2% ($SD = 2.6\%$). Additionally, with two listeners in the EEV task, the margin of error in VAS ratings (4.5%) suggests that intelligibility estimates would be reliable with regard to an approximation

of significant changes in intelligibility, based on reported differences in six speakers with PD pre- and post-LSVT therapy (Cannito et al., 2012).

Limitations and Future Work

While the current work suggests that using VAS-based intelligibility measures may be viable for individuals with PD, there are important limitations in terms of the speech stimuli used that should be addressed in future work. Intelligibility ratings in the current work were performed using a specific set of stimuli that may not generalize to other types of speech. Recordings of the Sentence Intelligibility Test sentences are not naturally elicited speech, and natural speech might have more prosody and contextual clues for listeners. Additionally, the intelligibility of speakers with PD has been reported to differ in conversational speech compared to reading tasks (Kempler & Van Lancker, 2002). Thus, future studies should incorporate naturally generated speech stimuli as well to determine how this affects the relationship between VAS-based and OT-based measures of intelligibility. The current results also only used speech stimuli from healthy older control speakers and speakers with PD; therefore, the strength of the relationship between VAS-based and transcription-based intelligibility with varying numbers of listeners per sentence needs to be explored using stimuli from other speaker populations to interpret the current results in the context of a wider clinical population. Since the majority of the speakers in the current study had intelligibility that was greater than 50% at 0 dB SNR, it is possible that speakers with different types of dysarthria and lower speech intelligibility may yield different trends than the current work. For example, Y. Kim, Kent, and Weismer (2011) found that speakers with PD and multiple system atrophy had higher mean speech intelligibility compared to speakers with stroke and traumatic brain injury, and Stipanovic et al. (2016) reported comparably high mean intelligibility scores for speakers with multiple sclerosis and PD. Thus, whereas speakers with multiple sclerosis and multiple system atrophy may yield similar results to the current work due to higher intelligibility levels, speakers who have suffered a stroke or a traumatic brain injury may show weaker associations between VAS and OT due to lower intelligibility levels.

The presentation of the stimuli chosen in the current work may also have affected intelligibility ratings. Each sentence was presented a single time to prevent speech exposure; however, it is possible that this design choice resulted in a memory burden during the transcription of longer sentences. The speech stimuli were also intensity-normalized across all speakers to prevent recording environment bias since stimuli were taken from a large database consisting of speech collected at various locations using different equipment, which may have affected perceived intelligibility. Speakers with PD often exhibit hypophonia, or decreased speech loudness (Duffy, 2013), contributing to changes in intelligibility, so future work should investigate the specific effects of hypophonia on VAS-based and transcription-

based measures of intelligibility. For the OT-based measures of intelligibility, listeners were instructed to disregard punctuation in order to decrease the possibility of punctuation errors to be interpreted as changes in intelligibility; however, this could have increased the cognitive demand for the task and impacted the current results. It is important to note that the speech stimuli for both VAS and OT tasks were presented in multispeaker babble, which is not reflective of a clinical setting in which patients would be speaking in a quiet room. Unlike prior work (Sussman & Tjaden, 2012), the current work also did not exclude listener data based on intrarater or interrater reliability scores, so the range in reliability (see Figure 1) values may have also affected the results.

Lastly, though the MEV task yielded reliable scores of sentence intelligibility in the shortest amount of time in the current work (5 min, on average; see Figure 1), this does not account for the amount of time needed to recruit naïve listeners. The EEV task, designed to represent increased exposure to disordered speech and to be the most feasible task in terms of clinical recruitment of listeners, revealed that two listeners per sentence could yield a strong relationship ($R^2 = .822$, $SD = .054$) between speaker-averaged VAS scores and transcription-based scores of intelligibility. The current work did not determine the minimum number of sentences needed to yield a strong relationship in an EEV task, so this should be explored in future work. The EEV task results in the current work remain promising for VAS-based scores of intelligibility but suggest that at least two listeners with exposure to a speaker's dysarthria would need to have availability to administer the ratings, which may not be feasible to recruit in a busy clinic with many patients requiring these assessments. Future studies should extend and validate these results by performing a similar study with SLP ratings of intelligibility to determine how VAS ratings of individuals with increased exposure to dysarthric speech compare to ratings by expert SLPs.

Conclusion

This study provides a further investigation of the relationship between VAS-based and transcription-based scores of intelligibility. We investigated the effect of an ideal, MEV task compared to a nonideal, EEV task on the relationship between VAS scores and OT scores of intelligibility. The current work suggests that, under an MEV task, one listener per sentence can yield a strong relationship ($R^2 = .835$, $SD = .026$) between speaker-averaged VAS scores and transcription-based scores of intelligibility. Under an EEV task, designed to reflect increased exposure to an individual's dysarthric speech, simulations of smaller listener groups showed that at least two listeners per sentence were needed to yield a strong relationship ($R^2 = .822$, $SD = .054$) between speaker-averaged VAS scores and transcription-based scores of intelligibility of speakers with PD while also maintaining sensitivity to what might be significant changes in intelligibility ($\geq 7.06\%$ deviation in intelligibility estimates; Cannito et al., 2012).

Thus, VAS ratings administered by listeners with more exposure to a specific speaker's dysarthria may be more clinically accessible to recruit and result in a less time-consuming measure of speech intelligibility. Future work is still needed to extend these findings to other speaker populations, differing types of speech stimuli, and clinical application of VAS-based ratings.

Acknowledgments

This work was supported by Grants DC015570 (awarded to C. E. S.), DC016270 (awarded to C. E. S.), and DC013017 (awarded to C. M) from the National Institute on Deafness and Other Communication Disorders and by the Undergraduate Research Opportunities Program (N. M. E.) from Boston University. The authors thank Sujata Pradhan for assistance with speaker recordings and Talia Mittleman, Ashling Lupiani, Thomas Elliott, Kat Kolin, and Paige Clabby for their help with participant recruitment.

References

- Adams, S. G., Dykstra, A., Jenkins, M., & Jog, M. (2008). Speech-to-noise levels and conversational intelligibility in hypophonia and Parkinson's disease. *Journal of Medical Speech-Language Pathology, 16*(4), 165–173.
- American Speech-Language-Hearing Association. (2005). *Guidelines for manual pure-tone threshold audiometry*. Rockville, MD: Author.
- Beukelman, D. R., & Yorkston, K. M. (1980). Influence of passage familiarity on intelligibility estimates of dysarthric speech. *Journal of Communication Disorders, 13*(1), 33–41.
- Cannito, M. P., Suiter, D. M., Beverly, D., Chorna, L., Wolf, T., & Pfeiffer, R. M. (2012). Sentence intelligibility before and after voice treatment in speakers with idiopathic Parkinson's disease. *Journal of Voice, 26*(2), 214–219.
- DePaul, R., & Kent, R. D. (2000). A longitudinal case study of ALS: Effects of listener familiarity and proficiency on intelligibility judgments. *American Journal of Speech-Language Pathology, 9*(3), 230–240.
- D'Innocenzo, J., Tjaden, K., & Greenman, G. (2006). Intelligibility in dysarthria: Effects of listener familiarity and speaking condition. *Clinical Linguistics & Phonetics, 20*(9), 659–675.
- Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management*. St. Louis, MO: Elsevier Health Sciences.
- Fahn, S. (2003). Description of Parkinson's disease as a clinical syndrome. *Annals of the New York Academy of Sciences, 991*(1), 1–14.
- Garcia, J. M., & Dagenais, P. A. (1998). Dysarthric sentence intelligibility: Contribution of iconic gestures and message predictiveness. *Journal of Speech, Language, and Hearing Research, 41*(6), 1282–1293.
- Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martínez-Martín, P., . . . Movement Disorder Society UPDRS Revision Task Force. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders, 23*(15), 2129–2170.
- Gurevich, N., & Scamihorn, S. L. (2017). Speech-language pathologists' use of intelligibility measures in adults with dysarthria. *American Journal of Speech-Language Pathology, 26*(3), 873–892.
- Hoover, J., Reichle, J., Van Tasell, D., & Cole, D. (1987). The intelligibility of synthesized speech: Echo II versus Votrax. *Journal of Speech and Hearing Research, 30*(3), 425–431.
- Hustad, K. C. (2006). 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, 15*(3), 268–277.
- Hustad, K. C., & Cahill, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology, 12*(2), 198–208.
- Hustad, K. C., Jones, T., & Dailey, S. (2003). Implementing speech supplementation strategies: Effects on intelligibility and speech rate of individuals with chronic severe dysarthria. *Journal of Speech, Language, and Hearing Research, 46*(2), 462–474.
- Huttunen, K., & Sorri, M. (2004). Methodological aspects of assessing speech intelligibility among children with impaired hearing. *Acta Oto-Laryngologica, 124*(4), 490–494.
- Kempler, D., & Van Lancker, D. (2002). Effect of speech task on intelligibility in dysarthria: A case study of Parkinson's disease. *Brain and Language, 80*(3), 449–464.
- Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54*(4), 482–499.
- Kim, H. (2015). Familiarization effects on consonant intelligibility in dysarthric speech. *Folia Phoniatrica et Logopaedica, 67*(5), 245–252.
- Kim, H., & Nanney, S. (2014). Familiarization effects on word intelligibility in dysarthric speech. *Folia Phoniatrica et Logopaedica, 66*(6), 258–264.
- Kim, Y., Kent, R. D., & Weismer, G. (2011). An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. *Journal of Speech, Language, and Hearing Research, 54*, 417–429.
- Lippmann, R. P. (1997). Speech recognition by machines and humans. *Speech Communication, 22*(1), 1–15.
- Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America, 112*(6), 3022–3030.
- Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *Journal of Speech and Hearing Disorders, 43*(1), 47–57.
- Martínez-Martín, P., Rodríguez-Blázquez, C., Alvarez, M., Arakaki, T., Arillo, V. C., Chaná, P., . . . Merello, M. (2015). Parkinson's disease severity levels and MDS-Unified Parkinson's Disease Rating Scale. *Parkinsonism & Related Disorders, 21*(1), 50–54.
- MathWorks. (2016). MATLAB 2016 (2016b). Natick, MA: Author.
- McHenry, M. (2011). An exploration of listener variability in intelligibility judgments. *American Journal of Speech-Language Pathology, 20*(2), 119–123.
- Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language & Communication Disorders, 48*(6), 601–612.
- Miller, N., Allcock, L., Jones, D., Noble, E., Hildreth, A. J., & Burn, D. J. (2007). Prevalence and pattern of perceived intelligibility changes in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry, 78*(11), 1188–1190.
- Miller, N., Noble, E., Jones, D., Allcock, L., & Burn, D. J. (2008). How do I sound to me? Perceived changes in communication in Parkinson's disease. *Clinical Rehabilitation, 22*(1), 14–22.
- Miyamoto, R. T., Robbins, A. M., Svirsky, M., Todd, S., Kirk, K. I., & Riley, A. (1997). Speech intelligibility of children with multi-channel cochlear implants. *Annals of Otolaryngology, Rhinology & Laryngology, 106*(5), 35.

-
- Olsen, W. O.** (1998). Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, & Fidell (1977) report. *American Journal of Audiology*, 7(2), 21–25.
- Pringsheim, T., Jette, N., Frolkis, A., & Steeves, T. D.** (2014). The prevalence of Parkinson's disease: A systematic review and meta-analysis. *Movement Disorders*, 29(13), 1583–1590.
- Stipancic, K. L., Tjaden, K., & Wilding, G.** (2016). Comparison of intelligibility measures for adults with Parkinson's disease, adults with multiple sclerosis, and healthy controls. *Journal of Speech, Language, and Hearing Research*, 59(2), 230–238. https://doi.org/10.1044/2015_JSLHR-S-15-0271
- Sussman, J. E., & Tjaden, K.** (2012). Perceptual measures of speech from individuals with Parkinson's disease and multiple sclerosis: Intelligibility and beyond. *Journal of Speech, Language, and Hearing Research*, 55(4), 1208–1219.
- Tjaden, K., Kain, A., & Lam, J.** (2014). Hybridizing conversational and clear speech to investigate the source of increased intelligibility in speakers with Parkinson's disease. *Journal of Speech, Language, and Hearing Research*, 57(4), 1191–1205.
- Tjaden, K., & Liss, J. M.** (1995). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics & Phonetics*, 9(2), 139–154.
- Tjaden, K., Sussman, J. E., & Wilding, G. E.** (2014). Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in Parkinson's disease and multiple sclerosis. *Journal of Speech, Language, and Hearing Research*, 57(3), 779–792.
- Utianski, R. L., Lansford, K. L., Liss, J. M., & Azuma, T.** (2011). The effects of topic knowledge on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *Journal of Medical Speech-Language Pathology*, 19(4), 25–36.
- Venkatagiri, H.** (1994). Effect of sentence length and exposure on the intelligibility of synthesized speech. *Augmentative and Alternative Communication*, 10(2), 96–104.
- Yorkston, K., Beukelman, D., & Tice, R.** (1996). *Sentence Intelligibility Test*. Lincoln, NE: Tice Technologies.