AJSLP

Research Article

Evaluation of Speech Amplification Devices in Parkinson's Disease

Monika D. Andreetta,[a] Scott G. Adams,[a] Allyson D. Dykstra,[a] and Mandar Jog[a]

Purpose: The purpose of this study was to evaluate the efficacy of selected speech amplification devices in individuals with hypophonia and idiopathic Parkinson's disease (PD).
Method: This study compared the effectiveness of seven devices (ADDvox, BoomVox, ChatterVox, Oticon Amigo, SoniVox, Spokeman, and Voicette) to unamplified speech for 11 participants with PD during conversation in 65-dB SPL multitalker noise, using experience ratings collected from participant questionnaires and speech performance measures (i.e., speech-to-noise ratio [SNR], speech intensity, and intelligibility) obtained from audio recordings.
Results: Compared with unamplified speech, device use increased SNR by 1.07–4.73 dB SPL and speech intensity by 1.1–5.1 dB SPL, and it significantly increased transcribed intelligibility from 13.8% to 58.9%. In addition, the type of device used significantly affected speech performance measures (e.g., BoomVox was significantly higher than most of the other devices for SNR, speech intensity, and intelligibility). However, experience ratings did not always correspond to performance measures.
Conclusions: This study found preliminary evidence of improved speech performance with device use for individuals with PD. A tentative hierarchy is suggested for device recommendations. Future research is needed to determine which measures will predict long-term device acceptance in PD.

Idiopathic Parkinson's disease (PD) is the second most common neurodegenerative disease, with an estimated prevalence of between 1 and 3 per 100 people ages 65 years and older (Wirdefeldt, Adami, Cole, Trichopoulos, & Mandel, 2011). There is no known cure for PD, and individuals can anticipate living with the disease and its effects for approximately 15 years postdiagnosis (Duffy, 2013). Therefore, symptom management is of primary concern.

Approximately 70% of individuals with PD will develop a speech impairment, which might not be alleviated with medication, and the majority of these speech symptoms are diagnosed as hypokinetic dysarthria (Adams & Dykstra, 2009; Darley, Aronson, & Brown, 1975; Spencer, Morgan, & Blond, 2009). Hypokinetic dysarthria has been described as encompassing various distinctive speech disturbances, including reduced loudness (hypophonia), short utterances, repeated phonemes, reduced stress, monopitch, monoloudness, inappropriate silences, short rushes of speech, variable rate, increased segment rate, rapid rate overall, and imprecise consonant articulation (Darley et al., 1975; Duffy, 2013; Logemann, Fisher, Boshes, & Blonsky, 1978).

Hypophonia is a significant impairment for 40%–50% of individuals with hypokinetic dysarthria and can manifest even in the early stages of PD (Adams & Dykstra, 2009; Kempler & Van Lancker, 2002). *Hypophonia* refers to a diminished speech intensity, which is generally 2–5 dB SPL lower than that of healthy geriatric individuals and corresponds to a 40% reduction in perceived speech volume (Adams, Winnell, & Jog, 2010; Fox & Ramig, 1997). The underlying pathophysiological mechanism may be diminished adductory force of the vocal folds, given that intensity depends in part on adequate subglottal pressure accumulation (Duffy, 2013). However, there is evidence to suggest that hypophonia may be more accurately attributed to a sensorimotor deficit in the self-perceived loudness of the individual's speech (Adams et al., 2010; Clark, Adams, Dykstra, Moodie, & Jog, 2014; Ho, Bradshaw, & Iansek, 2000; Ho, Bradshaw, Iansek, & Alfredson, 1999).

Speech Intensity and Speech Intelligibility

The relationship between speech intensity and speech intelligibility can be investigated by measuring signal-to-noise ratio (hereinafter referred to as speech-to-noise ratio or SNR; Kryter, 1994). When the noise intensity is subtracted from the speech intensity, the resultant value of the isolated speech signal is the SNR (DeBonis & Donohue, 2008). Conditions in which the speech intensity is less than or equal to

[a]Western University, London, ON, Canada
Correspondence to Monika D. Andreetta: mschel@uwo.ca

Editor: Krista Wilkinson
Associate Editor: Dawna Lewis
Received January 23, 2015
Revision received June 11, 2015
Accepted July 31, 2015
DOI: 10.1044/2015_AJSLP-15-0008

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

the noise (SNR ≤ 0 dB SPL) may impede the processing of speech sounds, which is essential for understanding speech.

Speakers have a natural compensatory response, known as the Lombard effect, that involves increasing their speech intensity in response to increases in background noise (Lane & Tranel, 1971). The Lombard effect helps to maintain a positive SNR and optimize the effectiveness of communication (Pick, Siegel, Fox, Garber, & Kearney, 1989).

The Lombard effect may be triggered when noise exceeds 50 dB SPL, and the extent to which the effect is exhibited may depend on the nature of the message in terms of the sender's unconscious perception of the importance of preserving intelligibility in order to convey the intended message (Dykstra, 2007). However, given the sensorimotor integration deficits exhibited in PD, it is important to consider the extent to which the Lombard effect is retained in PD and the effect that this has on speech intensity modulation and intelligibility (Adams et al., 2006; Adams, Haralabous, Dykstra, Abrams, & Jog, 2005; Lane & Tranel, 1971).

Kent, Weismer, Kent, and Rosenbek (1989) defined *speech intelligibility* as the accurate perception of the speech signal being transmitted. Because both the perception and production of a speech signal are fundamental to intelligibility, the low SNR from hypophonia may exacerbate reductions in speech intelligibility associated with the articulatory difficulties of hypokinetic dysarthria. Adams, Dykstra, Jenkins, and Jog (2008) reported that when both individuals with PD and control participants produced an SNR below 1.8 dB SPL during a conversation task, speech intelligibility scores fell below 50% (i.e., fewer than half of the words in each utterance could be identified by the listener). For both groups, an SNR of 5–7 dB SPL was associated with intelligibility scores of approximately 80%. In addition, despite some evidence that individuals with PD displayed a typical Lombard effect, a negative relationship was found between SNR and background noise levels and between intelligibility scores and noise levels for both individuals with PD and control participants (Adams et al., 2008; Ho et al., 1999). For example, the control group had an average SNR of approximately 5 dB SPL and an average intelligibility score of 90% in 60 dB SPL of noise, which fell to an average of approximately 3 dB SPL SNR and an average intelligibility score of approximately 70% in 70 dB SPL of noise (Adams et al., 2008). In contrast, the PD group had an average SNR of approximately 3 dB SPL and an average intelligibility score of 65% in 60 dB SPL of noise, which fell to an average of approximately 1.5 dB SPL SNR and an average intelligibility score of approximately 45% in 70 dB SPL of noise (Adams et al., 2008). These findings of a negative correlation between intelligibility and background noise and an overall reduction in conversational speech intelligibility in PD relative to control are consistent with a more recent investigation conducted by Dykstra, Adams, and Jog (2012). This suggests that as background noise increases, the SNR and intelligibility decrease, and this effect is even more pronounced for individuals with PD who have consistently lower speech intensity relative to controls.

Many intelligibility tests developed to assess dysarthria (e.g., reading or repetition tasks) assume uniformity of speech impairments and may underestimate impairments that manifest during the more variable speech tasks of everyday life (Kempler & Van Lancker, 2002). There is evidence that spontaneous speech is significantly less intelligible than reading aloud and repeating utterances (Frearson, 1985; Kempler & Van Lancker, 2002; Tjaden, 2006). In addition, conversation and other cognitively demanding speech tasks can greatly reduce speech intensity, especially for individuals with PD (Ho et al., 1999). Thus, experimental manipulations of both background noise level and speech task type are critical to accurately measure intelligibility for individuals with PD.

Therapeutic Treatment and Limitations

Miller, Noble, and Jones (2006) found that the four most prominent aspects of communication that affect quality of life for individuals with PD were social interaction, extemporaneous speech, intelligibility, and voice quality. Individuals with PD reported that the speech deficits of greatest concern were the perceived reduction in the ability to communicate, altered self-perception, and difficulty in long-term compensation for speech deficits, especially loudness, which often resulted in social disengagement (Miller et al., 2006). Thus, rehabilitation interventions that are focused on these key communication concerns may affect overall patient outcomes most profoundly (National Institute for Clinical Excellence, 2005).

Currently, the three common approaches to treatment of speech symptoms associated with hypokinetic dysarthria, including hypophonia, are behavioral speech therapy, biofeedback therapy, and prosthetic or assistive speech devices (Adams & Dykstra, 2009). There is substantial research that supports successful treatment outcomes for hypophonia from behavioral therapy (e.g., the Lee Silverman Voice Treatment; Ramig, Fox, & Sapir, 2004) and biofeedback therapy, which is the use of computer-based programs to provide online feedback on speech parameters, such as loudness (Scott & Caird, 1983). However, some evidence suggests that these improvements may fail to transfer outside the clinical setting (Adams & Dykstra, 2009; Rubow & Swift, 1985). This issue is a significant concern because some studies suggest that cognitive and sensorimotor impairments in PD may impair learning and that learning may be heavily context dependent for this clinical population (Adams & Dykstra, 2009). It is fortunate that speech amplification devices present a potential solution to the transfer-of-treatment issue.

Speech Amplification Devices

Speech amplification devices are a type of portable augmentative and alternative communication device that uses an individual's natural voice. According to the American Speech-Language-Hearing Association (1989), such devices "attempt to compensate (either temporarily or permanently) for the impairment and disability patterns of individuals with severe expressive communication disorders" (p. 107). In PD, speech amplification devices counteract the pattern of low speech intensity.

Early preliminary reports by Greene, Watson, Gay, and Townsend (1972) suggest that speech amplification devices may contribute to improvements in speech intelligibility by increasing the audibility of speech and by facilitating self-correction through self-monitoring. For example, it has been proposed that voice amplification may facilitate a reduction in the effort required to speech and enhance aural feedback, which may improve the accuracy of the speaker's perception of their speech and facilitate self-correction of imprecise articulation to improve intelligibility (Greene et al., 1972).

Speech amplification devices can vary in metrics such as signal-to-noise characteristics, amount of signal amplification, and audio frequency response range, and these differences in specifications may affect functional outcome measures such as speech intelligibility. In addition, specific disorder characteristics may be more or less sensitive to change on the basis of the use of a speech amplification device. For example, individuals with PD who speak in a whisper may have limited improvements in intelligibility from the use of speech amplification devices alone (Beukelman & Mirenda, 1998). Although amplification will enhance loudness and improve intelligibility of quiet speech, there are some features of speech sounds that are affected when speech is whispered (e.g., distinctions between voiced and voiceless stops) that would not be restored by simply increasing the loudness of the signal. It is unfortunate that the body of research on speech amplification devices is limited in terms of the disorder populations studied, conclusiveness of results, and number and types of devices compared.

In 2002, Roy et al. found that the ChatterVox voice amplifier, manufactured by Asyst Communications, was associated with better clarity, ease of voice production, and higher treatment compliance than a vocal hygiene program for teachers with voice disorders. In 2003, Roy et al. repeated this study with a larger sample size, but they included resonance therapy and respiratory muscle training as treatment alternatives. Again, the ChatterVox provided increased clarity and ease of voice production and overall voice improvement relative to the other treatments. Although the population included was limited to teachers, the results of these studies provide some support for the efficacy of speech amplification devices.

In 2002, Weiss compared the intelligibility of two severely dysarthric speakers using (a) the Speech Enhancer, (b) the ChatterVox, and (c) unamplified speech in various noise conditions. The Speech Enhancer, developed by Electronic Speech Enhancer, Inc., electronically filters and selectively amplifies a specific frequency range (800–4000 Hz) within a person's speech signal, whereas the ChatterVox and other traditional speech amplifiers do not filter out or amplify specific frequency ranges from the signal. The Speech Enhancer and ChatterVox were associated with higher intelligibility scores in loud background noise than those associated with unamplified speech.

In 2005, Bain, Ferguson, and Mathisen compared the effect of using the Speech Enhancer, the Voicette voice amplifier (Luminaud, Inc.), and unamplified speech on speech intelligibility in hyperkinetic dysarthria. For the less experienced and inexperienced judges, the unamplified condition produced significantly higher intelligibility than the Speech Enhancer did. However, use of the Speech Enhancer produced significantly more intelligible speech during the spontaneous speech task according to the experienced judges. Therefore, the results were inconclusive due to variations between the levels of experience of the judges and between speech tasks.

Although the findings in the previously discussed studies are limited, the general trend suggests that speech amplification devices may be an effective treatment option for individuals with PD and hypophonia. As efficacy research accumulates, the benefits of specific speech amplification devices may exceed the improvements reported from other treatments in some cases.

Speech Amplification Devices in Clinical Practice

In 2009, Bertrand conducted a study investigating which factors influence speech amplification device prescription in the United States and how prescription is distributed across clinical populations. The primary factor determining device prescription, as identified by 37 of the 62 participating speech-language pathologists (SLPs), was "patient preference and comfort" (Bertrand, 2009, p. 26). Although it may be assumed that patient preference correlates to successful device use, it appears that no published studies have investigated the existence or strength of this relationship. Another important finding was that the SLPs were not well informed about the variety of devices on the market. Almost 90% of SLPs surveyed had heard of the ChatterVox and about 73% had used it, making it the most commonly known and prescribed speech amplifier. As for other commonly prescribed speech amplification devices, such as the Speech Enhancer, EchoVoice EV4, and Spokeman, roughly a third of SLPs had never heard of these devices, and about half of SLPs had never used them.

In addition, when examining prescription distribution across different clinical populations, Bertrand (2009) found that speech amplification devices were more commonly prescribed for PD than all other communication disorders. Over 70% of the SLPs surveyed had prescribed speech amplification devices for PD and other motor speech disorders. Therefore, the lack of efficacy research and clinical knowledge regarding speech amplification devices is a significant concern for individuals with PD in particular.

The purpose of this study was to evaluate the efficacy of selected speech amplification devices in individuals with hypophonia and PD. Device effectiveness was examined during conversational speech in background noise using participant experience ratings and speech performance measures (i.e., SNR, speech intensity, and speech intelligibility scores). It was predicted that device-amplified speech would be more effective than unamplified speech and that there would be significant differences among the selected amplification devices in participant experience ratings and speech performance measures.

Method

Participants

This study included 11 individuals with hypophonia and mild to moderate idiopathic PD (aged 58–80 years; M = 70.9 years; 10 men, 1 woman). The average number of years since diagnosis of PD was 6.7 years (range = 1–16 years). Participants with PD were tested approximately 1 hr after their regularly scheduled anti-Parkinson medication. Two of the participants with PD were not on anti-Parkinson medications, whereas all other participants were on levodopa–carbidopa medication. None of the participants with PD had been previously prescribed a speech amplification device. The participants had no prior history of speech, language, or hearing problems. The Mini Mental State Examination (Folstein, Folstein, & McHugh, 1975) was used to exclude participants with dementia (cutoff score = 26/30). All participants with PD passed a bilateral 30 dB HL hearing screening at 500, 1000, and 2000 Hz.

This study also included 10 participants (9 women, 1 man) aged 21–25 years (M = 22.7) who served as listeners for an intelligibility test that used a visual analog scaling procedure. One additional female participant (23 years old) served as a listener for a transcription-based intelligibility test. None of the listener participants had previous experience with dysarthric speech. This study was approved by the Health Sciences Research Ethics Board at Western University, London, ON, Canada.

Speech Amplification Devices

This study evaluated seven speech amplification devices (see Figure 1): Addvox (Addvox, Waltham, MA), Boomvox (Griffin Laboratories, Temecula, CA), Chatterbox (Connections Unlimited, Nashville, TN), Oticon Amigo (Oticon, Smørum, Denmark), Sonivox (Griffin Laboratories, Temecula, CA), Spokeman (KEC Innovations, Singapore), and Voicette (Luminaud Inc., Mentor, OH). The ADDvox, ChatterVox, BoomVox, SoniVox, and Spokeman were selected for inclusion in this study because they have been previously recommended by their manufacturers for individuals with PD, and all five are approved by the Ontario Ministry of Health and Long-Term Care's Assistive Devices Program. The Voicette is an older device that was previously approved by the ministry's program. The Oticon Amigo is prescribed mainly to assist with hearing impairment; however, it was included in the present study because of its high-quality lightweight speaker accessory, which enabled it to function comparably to the other speech amplification devices. Thus, the Oticon Amigo represents a potentially effective type of amplification system that has rarely been considered for the treatment of PD.

The ADDvox (Stanton Magnetics, Inc.), SoniVox (Griffin Laboratories, Inc.), and ChatterVox are similar in dimensions, weight, and placement style. The dimensions of the three amplifiers are approximately 8 × 3 × 3 in. (20.32 × 7.62 × 7.62 cm) with an approximate weight of 1–2 lbs. (0.45–0.91 kg), and all three devices come with an adjustable belt strap to be worn around the waist. The Spokeman (KEC Innovations) can be worn around the waist or the arm, and it is the smallest and lightest speech amplifier included in this study with dimensions of 2.76 × 2.76 × 1.34 in. (7.01 × 7.01 × 3.4 cm) and a weight of 3.7 oz (0.10 kg). The Voicette is a larger speech amplifier with dimensions of 6.5 × 6.5 × 3 in. (16.51 × 16.51 × 7.62 cm) and a weight of 2 lbs. (0.91 kg); it is designed to be worn over the shoulder. The BoomVox (Griffin Laboratories, Inc.) is the largest of all the speech amplification devices, with dimensions of 7 × 4 × 11 in. (17.78 × 10.16 × 27.94 cm) and a weight of 5 lbs. (2.27 kg). The lightweight, wireless transmitter has a belt clip, and the speaker accessory can be carried by a handle. The Oticon Amigo (Oticon A/S) has a lightweight, wireless transmitter with a belt clip and a receiver that also has a belt clip, and it can be connected to headphones or a lightweight speaker accessory; the latter was used in this study.

Microphones were paired with the speech amplification devices according to the microphone options included by default with the devices. The Voicette was the only device that utilized a handheld microphone because it was considered to be of superior quality to the headset microphone option.

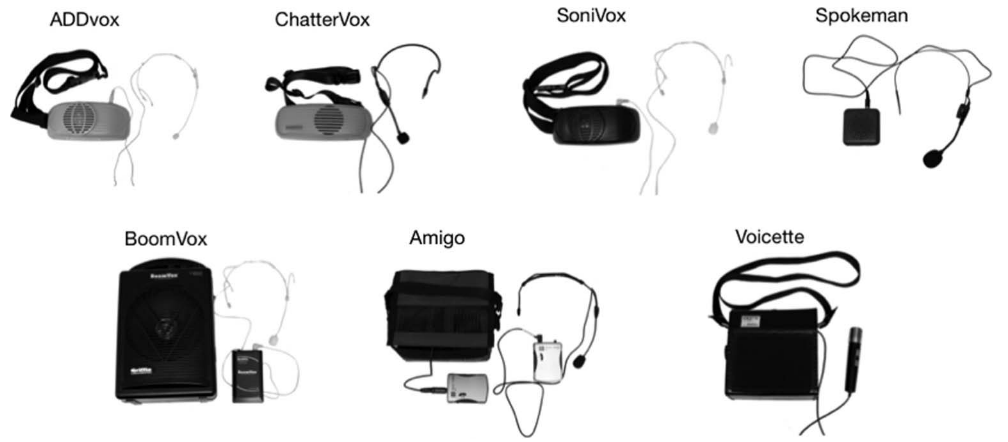
Two important variables in device specification captured by the devices selected for this study were speaker size and connection type. In general, the output from smaller speakers have lower intensities than larger speakers of the same amplification power. In the present study, the speaker volume was set at a level just below that which triggered continuous electronic or auditory feedback distortions for each participant. The FM technology used with the BoomVox and Oticon Amigo is an important variable due to the compression of the speech signal necessary for wireless transmission. This compression could affect the fidelity of the signal by reducing the frequency range of the audio (Bell, 2012). In addition, FM technology is subject to interference from any other signals being broadcast on the FM wireless spectrum (Bell, 2012).

Procedure

Speech Tasks

Speech samples were collected in the Speech Movement Disorders Lab in Elborn College at Western University. Conversational speech samples were gathered in 65-dB SPL multitalker background noise during eight device conditions. The experiment consisted of seven conditions in which each of the devices (ADDvox, BoomVox, ChatterVox, Oticon Amigo, SoniVox, Spokeman, and Voicette) was used and one condition of unamplified speech where no device was used. The last device condition (no device) served as a control for speech performance measures (i.e., SNR, speech intensity, and conversational speech intelligibility) for each speaker.

Figure 1. Pictures of the speech amplification devices. Top (left to right): ADDvox, ChatterVox, SoniVox, Spokeman. Bottom (left to right): BoomVox, Oticon Amigo, Voicette.



The eight device conditions were randomized for each participant with PD. It should be noted that the participants performed in three additional conditions that were part of a larger unpublished study and that will not be reported in the present study (Andreetta, 2013). These unreported conditions included conversational speech in no background noise and the Sentence Intelligibility Test (Yorkston, Beukelman, & Tice, 1996) in no-noise and 65-dB noise conditions.

Before each new device condition, speakers were reminded to use their typical speech intensity and rate. Speakers engaged in an unscripted conversation with one of the examiners for approximately 2 min per device condition. The purpose of this task was to simulate an ecologically valid speech context by mimicking the demands of real-life situations in which the performance of the device is most pertinent. The examiner elicited extemporaneous speech from the participant by posing selected open-ended questions and logical follow-up questions to prompt additional utterances as needed. The topics presented in the examiner's questions were those with which the speaker was likely to be comfortable and familiar (e.g., a memorable or recent trip, hobbies, current or former occupation).

Experimental Setup

Speech was recorded with a headset microphone (AKG c520) placed 6 cm from the participant's mouth and a stand-mounted omnidirectional dynamic microphone placed 4 m from the participant. Two loudspeakers remained at a constant distance from the speaker to distribute free-field background noise. Prerecorded, four-talker background noise (Audiotech Corp.) was played through the two loudspeakers to produce 65-dB SPL multitalker noise. The multitalker noise was calibrated to 65 dB SPL using a sound level meter placed beside the stand-mounted microphone (4 m), beside the participant's head, and beside the examiner's head. This noise intensity level was chosen because it exceeds the threshold to trigger the Lombard effect and it represents a noise level that is equivalent to what speakers may experience in daily communication situations (e.g., moderate cafeteria noise; Adams et al., 2005; Adams et al., 2006; Dykstra et al., 2012). A second floor microphone (2 m) was also calibrated to 65 dB SPL. A preliminary evaluation of the recordings from this floor microphone indicated that many of the devices were associated with 100% intelligibility. To avoid a ceiling effect and to increase the sensitivity of the outcome measures, the more distant 4-m floor microphone recordings were used in the present study.

Participant Experience Questionnaire

Immediately after using each device, speakers completed a device-based experience questionnaire to rate their experience on a visual analog scale (VAS) for each of the following dimensions: (a) physical comfort, (b) visual presentation, (c) sound quality, (d) perceived amplification power, and (e) overall preference (see Appendix A). This sequence allowed participants to gather a more cogent analysis of their experience with each device when filling out the questionnaire, despite the large number of devices compared in a limited time frame.

Acoustic Analyses

Two objective acoustic measures (average speech intensity and average SNR) were of primary interest in this study. Speech intensity and noise intensity values were obtained from the 4-m recordings using the intensity analysis functions in PRAAT software (Boersma & Weenink, 2008). Speech intensity values were obtained for 11 utterances (minimum five-word utterance lengths) in each of the eight device conditions. The utterances were continuous (fluent), and pauses greater than 250 ms were excluded from the average intensity measurements.

Average noise values were obtained by measuring intensity for three nonspeech selections (of the 65-dB SPL multitalker background noise) that occurred between spoken utterances. SNR was calculated by subtracting the average intensity of the three nonspeech noise selections from the average intensity of the 11 utterances to obtain a single ratio for each participant in each of the eight device conditions.

Listening Tasks

The assessment of the intelligibility of conversational speech is associated with an important methodological challenge related to uncertainty of the intended utterance. Without knowing the intended utterance, it is very difficult to score the accuracy of listener transcriptions. Because of this challenge, most attempts to measure conversational speech intelligibility in dysarthric participants have used listener rating scales (i.e., equal-appearing interval scales or VASs) instead of transcription procedures. One exception is a novel transcription procedure that was used with hypophonic PD participants (Adams et al., 2008). This transcription procedure used a headset microphone to estimate the intended utterances of the speakers with hypophonia as they spoke in different levels of background noise (Adams et al., 2008). The present study used this new transcription procedure and the traditional VAS procedure to measure conversational speech intelligibility. Thus, two different listening tasks were used to obtain two separate measures of intelligibility: (a) a transcription-based intelligibility score and (b) an intelligibility score that was based on a VAS listening procedure. For both listening tasks, the naïve listeners were not present at the time of the recordings and were blinded to all information about the speakers and to the content of the sentences. In the transcription task, the listener orthographically transcribed randomized speech samples for all 88 utterances (11 utterances × 8 devices) obtained from each PD participant for a total of 968 conversational utterances (88 utterances × 11 participants). The listener transcribed the speech recordings, which were obtained from the stand-mounted microphone placed 4 m from the participants, over the course of approximately four 2-hr listening sessions. One of the investigators independently transcribed all of the same utterances using the simultaneous recordings obtained from the participants' headset microphone. These headset recordings contained minimal noise and therefore were used as the "correct" reference for scoring the listener's speech-in-noise transcriptions of the PD participants' utterances. Thus, a conversational speech intelligibility percentage score was obtained by dividing the number of words that the listener correctly transcribed by the total number of words in the more accurate investigator transcription.

For the VAS-based intelligibility listening task, the 10 listeners were presented with one randomly selected utterance from each of the 11 PD participants' eight device conditions. These 88 utterances were randomly presented to each of the 10 listeners (different random order for each listener), who then provided a VAS rating of intelligibility for each utterance. The VAS involved a 10-cm horizontal line marked with end points of 0% and 100% intelligible. To examine the listeners' intrajudge reliability, each of the 88 utterances was randomly repeated. Thus, each of the 10 listeners was required to make 176 separate VAS intelligibility judgments during a 60-min listening session.

Results

Reliability

To examine reliability, approximately 10% of the speech intensity and SNR data and 5% of the transcribed intelligibility measures were reanalyzed by the same individual and analyzed independently by a different individual. Intraclass correlational analyses, using a two-way mixed-effects model and an absolute agreement definition for the intraclass correlation coefficient (ICC), were conducted to examine intrarater and interrater reliability for all dependent measures. Intrarater reliability for the measurement of speech intensity was high, ICC = .999, $p < .001$; and interrater reliability was high, ICC = .990, $p < .001$. Intrarater reliability for the measurement of SNR was high, ICC = .991, $p < .001$; and interrater reliability was high, ICC = .982, $p < .001$. Intrarater reliability for the measurement of intelligibility by transcription was high, ICC = .974, $p < .001$; and interrater reliability was high, ICC = .970, $p < .001$. Intrarater reliability for the measurement of intelligibility by VAS was good, with an overall average correlation coefficient of .901 and a range of .628–.971; and interrater reliability was high, ICC = .971, $p < .001$. Overall, these correlation coefficients demonstrated high intrarater and interrater reliability for all dependent measures.

Analyses of Outcome Measures

A series of repeated measures one-way analyses of variance (ANOVAs), combined with Bonferroni-corrected post hoc comparisons, were used to compare device conditions across each of the outcome measures. The outcome measures related to the experience ratings involved seven device conditions, and this was associated with 21 post hoc comparisons and a Bonferroni-corrected p value of .0023 (.05/21). The outcome measures related to SNR, intensity, and intelligibility involved eight device conditions, and this was associated with 28 post hoc comparisons and a Bonferroni-corrected p value of .0017 (.05/28). It should be noted that these Bonferroni corrections produce fairly low critical p values, and these may raise concerns about the risk of producing a Type 2 error (i.e., failing to find a significant difference in devices when a difference actually exists; false negative). This potential concern about the risk of Type 2 errors with the use of Bonferroni corrections in small-sample studies has been previously discussed (Nakagawa, 2004). In an attempt to address this potential concern, the uncorrected p values for each post hoc comparison related to each of the outcome measures have been provided in Appendix B.

Experience Ratings

Table 1 depicts the means and standard deviations for each device across each experiential domain. For the dimension of physical comfort, the Spokeman (M = 75.5, SD = 5.35) received the highest average rating, followed by the Oticon Amigo (M = 74.8, SD = 5.33), BoomVox

Table 1. Means and standard deviations of experience ratings by device type.

	Device						
Characteristic	ADDvox	BoomVox	ChatterVox	Amigo	SoniVox	Spokeman	Voicette
Physical comfort							
M	69.6	74.6	67.2	74.8	69.4	75.5	49.1
SD	6.74	5.83	7.11	5.33	5.91	5.35	9.65
Visual presentation							
M	60.4	65.5	62.8	63.8	71.1	72.8	45.9
SD	6.38	7.77	7.07	6.83	5.24	5.03	8.94
Sound quality							
M	73.0	77.2	78.4	75.8	73.3	65.2	65.6
SD	4.86	4.12	5.02	4.06	5.61	6.31	6.37
Amplification power							
M	68.2	75.9	78.2	74.2	72.2	68.2	65.6
SD	5.44	3.07	4.40	3.98	4.88	6.07	5.73
Overall preference							
M	57.3	59.8	67.3	62.4	58.5	68.3	40.0
SD	6.34	7.39	5.21	7.16	8.03	6.60	8.01

Note. Experience ratings represent the percentage on a visual analog scale ranging from 0% to 100%.

($M = 74.6$, $SD = 5.83$), ADDvox ($M = 69.6$, $SD = 6.74$), SoniVox ($M = 69.4$, $SD = 5.91$), ChatterVox ($M = 67.2$, $SD = 7.11$), and Voicette ($M = 49.1$, $SD = 9.65$). The results of a one-way repeated measures ANOVA found a significant difference in the participants' ratings of physical comfort across the seven devices, $F(1, 10) = 2.94$, $p = .013$. None of the 21 Bonferroni-corrected post hoc comparisons were significant ($p < .0023$). If a less conservative, uncorrected p value of .05 were to be applied to these comfort ratings, there would be five of 21 significant comparisons. All of these relate to low comfort ratings for the Voicette relative to the other devices. These uncorrected p values and the related device comparisons are provided in Appendix B.

For the dimension of visual presentation, the Spokeman ($M = 72.8$, $SD = 5.03$) received the highest average rating, followed by the SoniVox ($M = 71.1$, $SD = 5.24$), BoomVox ($M = 65.5$, $SD = 7.77$), Oticon Amigo ($M = 63.8$, $SD = 6.83$), ChatterVox ($M = 62.8$, $SD = 7.07$), ADDvox ($M = 60.4$, $SD = 6.38$), and Voicette ($M = 45.9$, $SD = 8.94$). The results of a one-way repeated measures ANOVA found a significant difference in the participants' ratings of visual presentation across the seven devices, $F(1, 10) = 2.70$, $p = .022$. None of the 21 Bonferroni-corrected post hoc comparisons were significant ($p < .0023$). If a less conservative, uncorrected p value of .05 were to be applied to these visual presentation ratings, there would be four of 21 significant comparisons. All of these relate to low visual presentation ratings for the Voicette relative to the other devices. These uncorrected p values and the related device comparisons are provided in Appendix B.

For the dimension of sound quality, the ChatterVox ($M = 78.4$, $SD = 5.02$) received the highest average rating, followed by the BoomVox ($M = 77.2$, $SD = 4.12$), Oticon Amigo ($M = 75.8$, $SD = 4.06$), SoniVox ($M = 73.3$, $SD = 5.61$), ADDvox ($M = 73.0$, $SD = 4.86$), Voicette ($M = 65.6$, $SD = 6.37$), and Spokeman ($M = 65.2$, $SD = 6.31$). The results of a one-way repeated measures ANOVA failed to find a significant difference in the participants' ratings of sound quality across the seven devices, $F(1, 10) = 1.66$, $p = .144$.

For the dimension of perceived amplification power, the ChatterVox ($M = 78.2$, $SD = 4.40$) received the highest average rating, followed by the BoomVox ($M = 75.9$, $SD = 3.07$), Oticon Amigo ($M = 74.2$, $SD = 3.98$), SoniVox ($M = 72.2$, $SD = 4.88$), ADDvox ($M = 68.2$, $SD = 5.44$), Spokeman ($M = 68.2$, $SD = 6.07$), and Voicette ($M = 65.6$, $SD = 5.73$). The results of a one-way repeated measures ANOVA failed to find a significant difference in the participants' ratings of perceived amplification power across the seven devices, $F(1, 10) = 1.24$, $p = .295$.

For the dimension of overall preference, the Spokeman ($M = 68.3$, $SD = 6.60$) received the highest average rating, followed by the ChatterVox ($M = 67.3$, $SD = 5.21$), Oticon Amigo ($M = 62.4$, $SD = 7.16$), BoomVox ($M = 59.8$, $SD = 7.39$), SoniVox ($M = 58.5$, $SD = 8.03$), ADDvox ($M = 57.3$, $SD = 6.34$), and Voicette ($M = 40.0$, $SD = 8.01$). The results of a one-way repeated measures ANOVA found a significant difference in the participants' ratings of overall preference across the seven devices, $F(1, 10) = 2.71$, $p = .021$. Post hoc comparisons using the Bonferroni correction for multiple comparisons indicated that the average overall preference rating for the ChatterVox ($M = 67.3$, $SD = 5.21$) was significantly higher than that for the Voicette ($M = 40.0$, $SD = 8.01$; $p = .018$). None of the other 20 Bonferroni-corrected post hoc comparisons were significant. If a less conservative, uncorrected p value of .05 were to be applied to these overall preference ratings, five of 21 comparisons would be significant. All of these relate to low overall preference ratings for the Voicette relative to the other devices. These uncorrected p values and the related device comparisons are provided in Appendix B.

In summary, although only one of the experience-based post hoc comparisons reached significance, the following hierarchical observations are noted: (a) The ChatterVox had the highest sound quality and perceived amplification power ratings, (b) the Spokeman had the highest comfort, appearance, and overall preference ratings but received low sound quality and perceived amplification power ratings, and (c) the Voicette received the lowest comfort, appearance, amplification power, and overall preference ratings.

SNR

The results of a one-way repeated measures ANOVA found a significant difference in the SNR across the eight device conditions, $F(1, 10) = 13.91$, $p < .001$. Table 2 depicts the means and standard deviations for the SNR for each device condition.

Post hoc comparisons using the Bonferroni correction for multiple comparisons indicated that the average SNR for the BoomVox ($M = 6.04$, $SD = 2.66$) was significantly higher than that for the ChatterVox ($M = 3.62$, $SD = 2.26$; $p = .033$), the Spokeman ($M = 3.11$, $SD = 2.72$; $p = .006$), the Oticon Amigo ($M = 3.05$, $SD = 2.27$, $p < .001$), the SoniVox ($M = 2.80$, $SD = 2.37$; $p = .006$), the ADDvox ($M = 2.38$, $SD = 1.18$; $p = .002$), and the unamplified, no-device condition ($M = 1.31$, $SD = 0.80$; $p = .001$). The SNR for the BoomVox exceeded that of the unamplified speech condition by approximately 5 dB SPL, whereas the SNR of the ADDvox was only approximately 1 dB SPL higher than the unamplified speech condition. Figure 2 shows the SNR values for each device condition arranged hierarchically from lowest to highest.

These results relate to the Bonferroni-corrected post hoc comparisons ($p = .0017$). If a less conservative, uncorrected p value of .05 were to be applied to these SNR data, an additional 11 device comparisons would reach significance. These uncorrected p values and the related device comparisons are provided in Appendix B.

Speech Intensity

The results of the one-way repeated measures ANOVA found a significant difference in the speech intensity across the eight device conditions, $F(1, 10) = 14.04$, $p < .001$. Table 3 depicts the means and standard deviations for the speech intensity for each device condition.

Post hoc comparisons using the Bonferroni correction for multiple comparisons indicated that the average speech intensity for the BoomVox ($M = 71.3$, $SD = 4.57$) was significantly higher than that for the ChatterVox ($M = 68.2$, $SD = 3.75$; $p = .004$), Voicette ($M = 68.1$, $SD = 4.04$; $p = .011$), the SoniVox ($M = 68.1$, $SD = 4.68$; $p = .004$), the Oticon Amigo ($M = 68.0$, $SD = 4.10$; $p < .001$), the Spokeman ($M = 67.7$, $SD = 4.17$; $p = .001$), the ADDvox ($M = 67.3$, $SD = 3.52$; $p = .001$), and unamplified speech ($M = 66.2$, $SD = 2.79$; $p = .002$). The speech intensity from the BoomVox exceeded that of the unamplified speech condition by approximately 5 dB SPL, whereas the speech intensity from the ADDvox was only approximately 1 dB SPL higher than unamplified speech, which was not found to be statistically significant.

These results relate to the Bonferroni-corrected post hoc comparisons ($p = .0017$). If a less conservative, uncorrected p value of .05 were to be applied to these speech intensity data, an additional nine device comparisons would reach significance. These uncorrected p values and the related device comparisons are provided in Appendix B.

Intelligibility: Transcribed Conversational Intelligibility

The results of the one-way repeated measures ANOVA found a significant difference in the average transcribed intelligibility score across the eight device conditions, $F(1, 10) = 14.30$, $p < .001$. Table 4 depicts the means and standard deviations for the intelligibility from each device condition.

Post hoc comparisons using the Bonferroni correction for multiple comparisons indicated that the average intelligibility score for the BoomVox ($M = 73.6$, $SD = 29.4$) was significantly higher than that for the SoniVox ($M = 32.4$, $SD = 29.5$; $p = .003$), the ADDvox ($M = 28.5$, $SD = 22.5$; $p = .007$), and the no-device condition ($M = 14.7$, $SD = 16.5$; $p = .001$). In addition, the average intelligibility score for the ChatterVox ($M = 47.8$, $SD = 30.9$) was significantly higher than for the SoniVox ($M = 32.4$, $SD = 29.5$; $p = .034$) and the unamplified, no-device condition ($M = 14.7$, $SD = 16.5$; $p = .018$). Figure 3 shows the transcribed speech intelligibility values for each device condition arranged hierarchically from lowest to highest.

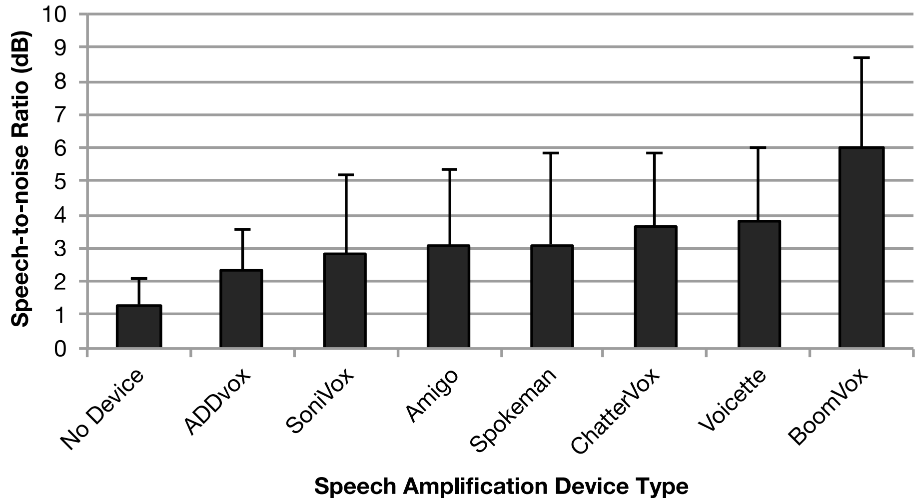
These results relate to the Bonferroni-corrected post hoc comparisons ($p = .0017$). If a less conservative, uncorrected p value of .05 were to be applied to these intelligibility data, an additional 16 device comparisons would reach significance. These uncorrected p values and the related device comparisons related to the average transcribed intelligibility scores are provided in Appendix B.

Table 2. Speech-to-noise ratio means and standard deviations by device condition.

	Device							
Statistic	ADDvox	BoomVox	ChatterVox	Amigo	SoniVox	Spokeman	Voicette	Unamplified
M	2.38	6.04*	3.62	3.05	2.80	3.11	3.84	1.31
SD	1.18	2.66	2.26	2.27	2.37	2.72	2.13	0.80

Note. Speech-to-noise ratio values are presented in dB SPL. An asterisk indicates that the device performed significantly better than at least one other device condition.

Figure 2. Average speech-to-noise ratio by device condition.



Intelligibility: VAS of Intelligibility

The results of the one-way repeated measures ANOVA found a significant difference in the average intelligibility score obtained with the VAS across the eight device conditions, $F(1, 10) = 8.946$, $p < .001$. Table 5 depicts the means and standard deviations for the intelligibility from each device condition.

Post hoc comparisons using the Bonferroni correction for multiple comparisons indicated that the average intelligibility score for the BoomVox ($M = 46.8$, $SD = 7.8$) was significantly higher than the Spokeman ($M = 25.3$, $SD = 8.1$; $p = .025$), the Oticon Amigo ($M = 22.9$, $SD = 6.5$; $p = .015$), the ADDvox ($M = 15.6$, $SD = 5.1$; $p = .008$), the SoniVox ($M = 13.2$, $SD = 3.5$; $p = .020$), and the unamplified, no-device condition ($M = 7.2$, $SD = 2.2$; $p = .018$). Figure 4 shows the average VAS speech intelligibility values for each device condition arranged hierarchically from lowest to highest.

These results relate to the Bonferroni-corrected post hoc comparisons ($p = .0017$). If a less conservative, uncorrected p value of .05 were to be applied to these intelligibility data, an additional 11 device comparisons would reach significance. These uncorrected p values and the related device comparisons are provided in Appendix B.

Discussion

The purpose of this study was to provide a preliminary investigation into speech amplification device efficacy for individuals with PD in an ecologically valid context (i.e., during conversation in moderate background noise). Speech outcome measures included experience ratings and speech performance measures (i.e., SNR, speech intensity, and conversational intelligibility scores). The present study found that amplification devices effectively increased SNR, speech intensity, and speech intelligibility compared with the unamplified speech of individuals with PD in moderate background noise. In addition, the type of device used significantly affected all outcome measures. However, device effectiveness varied across outcome measures.

An SNR of 1.8 dB SPL has been reported to be associated with conversational speech intelligibility of less than 50%, and an SNR of 5–7 dB SPL was found to be associated with 80% intelligibility (Adams et al., 2008). Overall, the present study found that a higher average SNR predicted higher average conversational speech intelligibility. The BoomVox had an average SNR of 6.04 dB SPL and an average transcribed intelligibility score of 73.6%. All other devices had SNRs in the range of 2.38–3.84 dB SPL and corresponding transcribed intelligibility scores in the range

Table 3. Speech intensity means and standard deviations by device condition.

	Device							
Statistic	ADDvox	BoomVox	ChatterVox	Amigo	SoniVox	Spokeman	Voicette	Unamplified
M	67.3	71.3*	68.2	68.0	68.1	67.7	68.1	66.2
SD	3.52	4.57	3.75	4.10	4.68	4.17	4.04	2.79

Note. Speech intensity values are presented in dB SPL. An asterisk indicates that the device performed significantly better than at least one other device condition.

Table 4. Transcribed intelligibility means and standard deviations by device condition.

	Device							
Statistic	ADDvox	BoomVox	ChatterVox	Amigo	SoniVox	Spokeman	Voicette	Unamplified
M	28.5	73.6*	47.8*	44.9	32.4	38.9	49.2	14.7
SD	22.5	29.4	30.9	34.7	29.5	37.9	35.5	16.5

Note. Scores represent the percentage of words correctly transcribed ranging from 0% to 100%. An asterisk indicates that the device performed significantly better than at least one other device condition.

of 28.5%–49.2%. Unamplified speech had an average SNR of 1.31 dB SPL and an average transcribed intelligibility score of 14.7%. Therefore, the results of this study suggest that speech amplification devices measurably increase the speech intensity (and corresponding SNR) sufficiently to increase both transcribed and VAS conversational speech intelligibility scores for individuals with PD in moderate background noise. These results are consistent with previous research findings that the use of speech amplification devices in background noise was associated with higher intelligibility scores than unamplified speech for individuals with dysarthria (Weiss, 2002).

With regard to the intelligibility testing procedures, it should be noted that the results for the two intelligibility tests found a fairly similar device hierarchy, but the VAS scores were generally lower than the transcribed scores for each device. This result is not consistent with a previous study of hypophonia that found that VAS scores were higher than transcribed intelligibility scores (Adams et al., 2008). This inconsistency may be related to the fact that device-amplified speech was used in the present study but not in the previous study. Perhaps, when speech is transmitted through a portable amplification system, there are signal distortions that cause the listener to judge speech to be less intelligible than it actually is. Another possible explanation is that the use of a different number of listeners in the two intelligibility tasks could have influenced the results. We used one listener in the transcription task and 10 listeners in the VAS task. Perhaps our one listener tended to give higher intelligibility values than the average of the 10 listeners.

Patient preference has been reported as the primary basis for device prescription (Bertrand, 2009). Thus, because the correspondence between patient preference and device effectiveness was unknown, the present study compared the results from experience ratings to those from speech performance measures (i.e., speech intensity, SNR, and speech intelligibility). Table 6 displays a summary of the device hierarchies across all experience ratings, and Table 7 displays a summary of the device hierarchies across all speech performance measures, where 1 is the highest rank and 7 is the lowest in both tables.

When considering the results of all outcome measures addressed in the present study together, the BoomVox received the highest scores overall. In particular, the BoomVox had the highest scores for speech intensity, SNR, and conversational speech intelligibility. We find it interesting that the device preference ratings did not produce the highest scores for the BoomVox. The participants with PD rated the BoomVox second highest for power and sound quality, third for comfort and visual appearance, and fourth for

Figure 3. Average transcribed intelligibility scores by device condition.

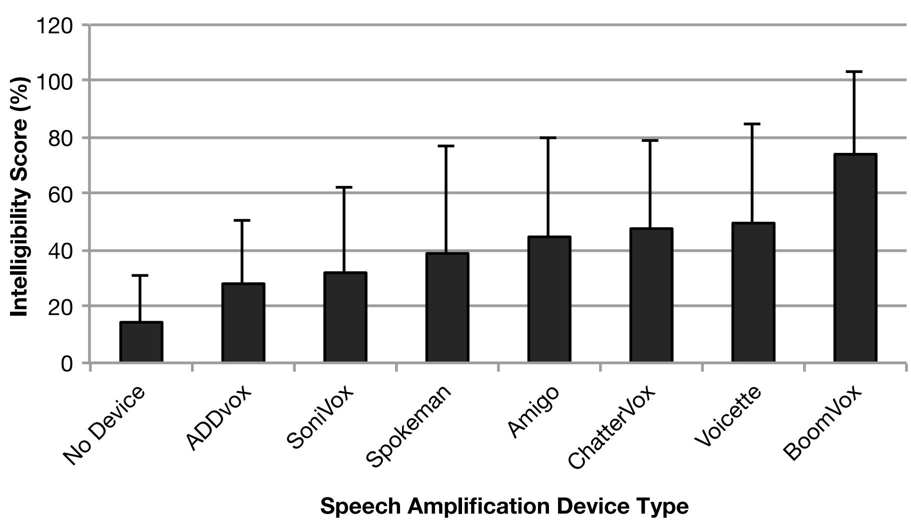


Table 5. Visual analog scale intelligibility means and standard deviations by device condition.

	Device							
Statistic	ADDvox	BoomVox	ChatterVox	Amigo	SoniVox	Spokeman	Voicette	Unamplified
M	15.6	46.8*	23.7	22.9	13.2	25.3	28.2	7.2
SD	5.1	7.8	6.3	6.5	3.5	8.1	8.0	2.2

Note. Scores represent the estimated percentage on a visual analog scale ranging from 0% to 100%. An asterisk indicates that the device performed significantly better than at least one other device condition.

overall preference. These results for the BoomVox provide an interesting example of the potential discrepancy between speech performance-based measures and experience-based preference ratings in the evaluation of amplification devices.

The ChatterVox was consistently in the top four highest positions in the device hierarchy for all outcome measures. The ChatterVox scores were second highest for speech intensity, third for SNR, and fourth for speech intelligibility. Although the ChatterVox was rated as having the highest perceived amplification power and sound quality, it was rated the sixth highest for comfort, fifth for appearance, and second for overall preference. Thus, the ChatterVox provides another interesting example of the potential discrepancy between speech performance measures and user experience ratings.

At the other end of the device hierarchy, the Voicette received the lowest physical comfort, visual presentation, and overall preference ratings but achieved second or third highest place in the device hierarchy for all speech performance measures. Conversely, the Spokeman received higher ratings on overall preference and aesthetic dimensions (comfort and visual presentation), which may be attributed to its smaller size, but low ratings on functional dimensions (amplification power and sound quality), and it fell into the lowest four positions on the device hierarchy for most speech performance measures.

Overall, it appears that speaker preference (i.e., on the basis of experience ratings) does not appear to be associated with higher speech performance measures for most devices. One possible explanation for this discrepancy may be that preference is highly influenced by the appearance and size of the device. However, additional research is needed to confirm whether aesthetics outweigh performance in determining device preference. A second possible explanation is that individuals with PD may have difficulty with the accurate perception of their own speech production (i.e., accurately appraising their own loudness, clarity, sound quality; Clark et al., 2014) and that these perceptual difficulties may be further exacerbated when attempting to evaluate speech produced in background noise or projected through amplification devices.

Strengths and Limitations

Previous studies investigating the effectiveness of speech amplification devices have included a limited number of devices and outcome measures. The present study is the first to systematically examine the effectiveness of a

Figure 4. Average visual analog scale intelligibility scores by device condition.

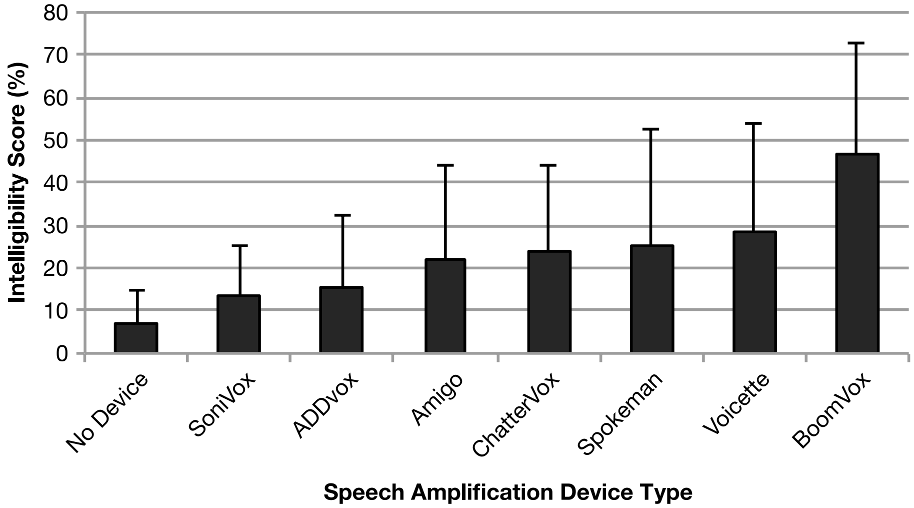


Table 6. Device hierarchies across experience ratings.

Rating	Physical Comfort	Visual	Sound Quality	Power	Overall
1	Spokeman	Spokeman	ChatterVox	ChatterVox	Spokeman
2	Amigo	SoniVox	BoomVox	BoomVox	ChatterVox*
3	BoomVox	BoomVox	Amigo	Amigo	Amigo
4	ADDvox	Amigo	SoniVox	SoniVox	BoomVox
5	SoniVox	ChatterVox	ADDvox	ADDvox	SoniVox
6	ChatterVox	ADDvox	Voicette	Spokeman	ADDvox
7	Voicette	Voicette	Spokeman	Voicette	Voicette

Note. An asterisk indicates that the device performed significantly better than at least one other device.

broad range of different device types for individuals with PD. In addition, this study incorporated a variety of outcome measures, including speaker preference, speech intensity, SNR, and conversational speech intelligibility. Thus, this study is the first to simultaneously collect data on speakers' preference and device performance. The present study also used an ecologically valid speech context (conversation in multitalker noise) to evaluate outcome measures related to speech performance.

There were two potential limitations of this study related to participant inclusion. The first potential limitation is the unequal representation of men and women in the PD group, which comprised 10 men and only one woman. This ratio of men to women is not consistent with that of the overall population of individuals with PD, and sex differences may have influenced speakers' outcome measures. Second, there were a relatively small number of speaker participants ($N = 11$). Therefore, individual differences may have caused increased variability in the results. Recruitment and inclusion of a greater number of speakers may have increased the power to detect significant effects and may have made significant findings more consistent and definitive. In consideration of the small number of participants included in this study, the power to detect significant differences may have been reduced further, given the large number of outcome measures.

Another limitation of this study is the large number of post hoc comparisons that were required to examine all of the pairwise differences involving the seven devices or eight device conditions (a total of 21 or 28 paired comparisons). The generally accepted statistical correction procedure (i.e., Bonferroni correction for multiple comparisons) that is required to avoid obtaining a false-positive (i.e., Type 1 error or a significant comparison by random chance alone) can result in a greater risk of causing a false-negative (i.e., Type 2 error or failing to detect a significant comparison when one actually exists). This risk of false-negatives (Type 2 error) increases as the number of multiple comparisons increases. Thus, when a large number of comparisons are examined, such as the 21 or 28 comparisons in the present study, there can be a shift in the balance between Type 1 and Type 2 errors toward a greater risk of Type 2 errors. This risk can be further compounded by the use of a fairly small sample size. In an attempt to address this potential imbalance, we elected to report the results of the traditional Bonferroni correction procedure for our multiple comparisons but also to include the uncorrected p values from the pairwise comparisons as supplementary information in Appendix B. This supplementary information is provided for those readers who may have concerns about the effects of the statistical correction procedure on the Type 2 statistical error and would like to examine the uncorrected p values for each of the specific device paired comparisons. The results of this preliminary, exploratory study indicate that future amplification device comparison studies should consider using a larger number of participants and examining a smaller number of devices.

A possible methodological limitation of this study is that speaker participants remained seated while using the devices. In typical use, devices are worn while standing and walking as well as sitting. These constraints on movement may have influenced outcome measures and limited the utility of the experience-based ratings, especially in terms of physical comfort. These additional device-use conditions

Table 7. Device hierarchies across speech performance measures.

Rating	Speech-to-Noise Ratio	Speech Intensity	Speech Intelligibility (Transcribed)	Speech Intelligibility (VAS)
1	BoomVox*	BoomVox*	BoomVox*	BoomVox*
2	Voicette	ChatterVox	Voicette	Voicette
3	ChatterVox	Voicette	ChatterVox	Spokeman
4	Spokeman	SoniVox	Amigo	ChatterVox
5	Amigo	Amigo	Spokeman	Amigo
6	SoniVox	Spokeman	SoniVox	ADDvox
7	ADDvox	ADDvox	ADDvox	SoniVox

Note. An asterisk indicates that the device performed significantly better than at least one other device. VAS = visual analog scale.

were not incorporated into the protocol in order to maintain consistency in the audio recording and background noise setup (e.g., stationary loudspeakers were set at a constant distance).

Another limitation related to the experimental protocol was that speakers rated their experience with each device after using it for only a short period of time (approximately 2–3 min total). The apparent tendency for device aesthetics to influence preference may be attributed to this limited time frame, whereas functional aspects may be more salient with increased use.

In addition, there were limitations associated with the listeners' perceptual judgments of sound quality. If the speech signal could not be identified above the background noise, the sound quality of the signal could not be accurately judged. This may have had the effect of decreasing the sound-quality scores for the devices that had lower amplification power. Related to this concern is the potential effect of a mild hearing loss on the perceptual judgments of some of the participants. Although all participants passed a 30-dB pure-tone screening at 500, 1000, and 2000 Hz, some of the participants may have had a mild hearing loss related to pure-tone frequencies above 2000 Hz. In the present study, a fairly conservative hearing status exclusion criteria was used to include as many participants with hypophonia and idiopathic PD as possible. The potential effect of mild hearing loss on judgments of device sound quality and performance should be examined in future studies.

A limitation of acoustic data collection was that many speech samples may have dropped below the 65 dB SPL multitalker noise, but they would have been measured as having a speech intensity equivalent to that of the noise. This would have the effect of inflating the speech intensity values for the unamplified speech condition. Related to this limitation is that the lowest possible SNR that could be obtained in this study is zero. This would be the case even if the speech intensity fell to values that were lower than the background noise. For example, if the speech signal was −5 dB below the noise, it would be measured as 0 dB SNR. The methodology used in the present study did not allow for the measurement of negative SNR values. Future studies are required to examine the potential effect of the negative SNR values on the device outcome measures.

Future Directions

This preliminary study provides a novel framework for future device efficacy research in PD. Future studies should focus on either increasing the number of participants or reducing the number of devices to increase the power to detect significant effects. Future research may also include different permutations of microphones with a select number of devices. In addition to ecologically valid speech tasks and noise conditions, researchers may want to consider incorporating a variety of speaking postures (e.g., standing and walking) to increase ecological validity. Another consideration is the potential role of a loudness perception deficit in the PD participants' evaluation of amplification power and judgments of device effectiveness (Clark et al., 2014). Participants with significant deficits in the perception of their own speech loudness may have difficulty making accurate judgments about the loudness of speech amplification devices. Therefore, it may be appropriate to assess device preference ratings from the perspective of the primary communication partners of individuals with PD.

Most important, future studies should investigate factors affecting long-term compliance with various devices to assess the influence of an individual's initial preference compared with the effectiveness of the device as demonstrated in speech performance measures.

Clinical Implications and Tentative Recommendations

As previously discussed, it has been reported that the ChatterVox is the most commonly prescribed amplification device and that most SLPs determine prescription primarily on the basis of "patient preference and comfort" (Bertrand, 2009). The results of the present study found that the ChatterVox was one of the most preferred devices and was associated with fairly consistent positive speech performance measures, surpassing most of the other devices. Therefore, device prescription on the basis of preference may not result in unsuccessful device use. Moreover, the results of this study found that the Spokeman surpassed the ChatterVox in both comfort and overall preference. Therefore, the reason the Spokeman was not reported to be the most commonly prescribed device may be related to the finding that fewer SLPs had heard of or used the Spokeman than the ChatterVox (Bertrand, 2009). Thus, clinician knowledge and experience may be a greater factor in device prescription than patient preference in current practice. However, clinicians may not be prescribing the Spokeman as often as other devices because of the poorer performance of the Spokeman, which may have been evident to clinicians even with limited experience with the device.

Although the present study found some variations in the device hierarchy across the outcome measures assessed, a tentative hierarchy of recommendation for device prescription is presented. This recommendation captures an overview of the individually identified hierarchies, but it should be viewed as a hypothesis that requires verification with additional research. The BoomVox is the highest recommended device on the basis of its speech performance measures surpassing those of all other devices and its good user experience scores. The loudspeaker component of the device makes it the largest and heaviest device, but the FM technology obviates the need for portability in most circumstances. The ChatterVox is recommended over the Oticon Amigo due to the generally higher scores associated with the ChatterVox across most outcome measures. The BoomVox, ChatterVox, and Oticon Amigo received high ratings fairly consistently across all experiential dimensions, but the ChatterVox was rated higher than all other devices on functional dimensions (sound quality and amplification power).

Although the Spokeman received some of the highest experience ratings, it is the fourth most recommended device because of inconsistencies in performance across outcome measures. The Voicette received the lowest ratings according to speakers' experience with the device, which offset its consistently high performance scores. The users' confidence in and comfort with the device may have been diminished by use of the handheld microphone. In addition, because the handheld microphone raises concerns about user fatigue, a headset microphone may be recommended, but this requires additional testing with the headset microphone to confirm that outcome measures are consistent with those obtained in this study. Neither the SoniVox nor the ADDvox are highly recommended on the basis of the results of this study. However, the experience ratings and speech performance measures associated with SoniVox were consistently higher than those of the ADDvox.

Overall, the results from this study may provide some insights for clinical practice. Most notably, the results suggest that device prescription on the basis of user preference and user comfort may not accurately predict device performance or effectiveness. Therefore, it is suggested that SLPs consider exploring device options that optimize speech performance while also considering aesthetic qualities of comfort and size. In addition, SLPs should consider providing patients ample time to develop an informed preference by testing the device in ecologically valid contexts. These recommendations are especially important to consider before using patient preference as the basis for prescription. Future research is needed to determine which performance and preference measures will best predict long-term speech amplification device acceptance in PD.

Acknowledgments

This research was supported by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada, awarded to Monika D. Andreetta.

References

Adams, S. G., & Dykstra, A. (2009). Hypokinetic dysarthria. In M. R. McNeil (Ed.). *Clinical management of sensorimotor speech disorders* (pp. 166–186). New York, NY: Thieme Medical.

Adams, S., Dykstra, A., Abrams, K., Winnell, J., Jenkins, M., & Jog, M. (2006). Conversational speech intensity under different noise conditions in hypophonia and Parkinson's disease. *Canadian Acoustics, 34,* 96–97.

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,* 165–172.

Adams, S., Haralabous, O., Dykstra, A., Abrams, K., & Jog, M. (2005). Effects of multi-talker background noise on the intensity of spoken sentences in Parkinson's disease. *Canadian Acoustics, 33,* 94–97.

Adams, S. G., Winnell, J., & Jog, M. (2010). Effects of interlocutor distance, multi-talker background noise, and a concurrent manual task on speech intensity in Parkinson's disease. *Journal of Medical Speech-Language Pathology, 18,* 1–8.

American Speech-Language-Hearing Association. (1989). Competencies for speech-language pathologists providing services in augmentative communication. *ASHA, 31,* 107–110.

Andreetta, M. (2013). *A comparison of speech amplification devices for individuals with Parkinson's disease and hypophonia* (Unpublished master's thesis). University of Western Ontario, London, ON, Canada.

Bain, C., Ferguson, A., & Mathisen, B. (2005). Effectiveness of the Speech Enhancer on intelligibility: A case study. *Journal of Medical Speech-Language Pathology, 13*(2), 85–95.

Bell, D. A. (2012). Wireless microphone systems. *Broadcast Engineering, 54*(11), 24–26.

Bertrand, A. N. (2009). *Speech-language pathologists decision making pertaining to speech amplification devices: A survey*. (Master's thesis). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 1466986)

Beukelman, D. R., & Mirenda, P. (1998). *Augmentative and alternative communication* (2nd ed.). Baltimore, MD: Brookes.

Boersma, P., & Weenink, D. (2008). PRAAT (version 5.0.20) [Software]. Available from http://www.fon.hum.uva.nl/praat/

Clark, J., Adams, S. G., Dykstra, A., Moodie, S., & Jog, M. (2014). Loudness perception and speech intensity control in Parkinson's disease. *Journal of Communication Disorders, 51,* 1–12.

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

DeBonis, D. A., & Donohue, C. L. (2008). *Survey of audiology: Fundamental for audiologists and health professionals* (2nd ed.). Boston, MA: Pearson Education.

Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management* (3rd ed.). St. Louis, MO: Elsevier Mosby.

Dykstra, A. (2007). *The effects of hypophonia on speech intelligibility, communication effectiveness and communication-related quality of life in Parkinson's disease* (Unpublished doctoral dissertation). University of Western Ontario, London, Ontario, Canada.

Dykstra, A., Adams, S. G., & Jog, M. (2012). Examining the conversational speech intelligibility of individuals with hypophonia associated with Parkinson's disease. *Journal of Medical Speech-Language Pathology, 20,* 53–57.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198.

Fox, C., & Ramig, L. (1997). Vocal sound pressure and self-perception of speech and voice in men and women with idiopathic Parkinson disease. *American Journal of Speech-Language Pathology, 6*(2), 85–93.

Frearson, B. (1985). A comparison of the A.I.D.S. sentence list and spontaneous speech intelligibility scores for dysarthric speech. *Australian Journal of Human Communication Disorders, 13,* 5–21.

Greene, M. C. L., Watson, B. W., Gay, P., & Townsend, D. B. K. (1972). A therapeutic speech amplifier and its use in speech therapy. *The Journal of Laryngology and Otology, 86,* 595–605.

Ho, A. K., Bradshaw, J. L., & Iansek, R. (2000). Volume perception in parkinsonian speech. *Movement Disorders, 15,* 1125–1131.

Ho, A. K., Bradshaw, J. L., Iansek, R., & Alfredson, R. (1999). Speech volume regulation in Parkinson's disease: Effects of implicit cues and explicit instructions. *Neuropsychologia, 37,* 1453–1460.

Kent, R., Weismer, G., Kent, J., & Rosenbek, J. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54,* 482–499.

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,* 449–464. doi:10.1006/brln.2001.2602

Kryter, K. D. (1994). *The handbook of hearing and the effects of noise*. San Diego, CA: Academic Press.

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research, 14,* 677–709.

Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, B. (1978). Frequency and co-occurrence of vocal tract dysfunctions in the speech sample of a large sample of Parkinson patients. *Journal of Speech and Hearing Disorders, 46,* 348–352.

Miller, N., Noble, E., & Jones, D. (2006). Life with communication changes in Parkinson's disease. *Age and Ageing, 35,* 235–239.

Nakagawa, S. (2004). A farewell to Bonferroni: The problems of a low statistical power and publication bias. *Behavioral Ecology, 15,* 1044–1045.

National Institute for Clinical Excellence. (2005). *Parkinson's disease: Diagnosis and management in primary and secondary care: National Collaborating Centre for Chronic Conditions: Draft for First Consultations*. London, United Kingdom: Royal College of Physicians.

Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R., & Kearney, J. K. (1989). Inhibiting the Lombard effect. *The Journal of the Acoustical Society of America, 85,* 848–900.

Ramig, L. O., Fox, C., & Sapir, S. (2004). Parkinson's disease: Speech and voice disorders and their treatment with the Lee Silverman voice treatment. *Seminars in Speech and Language, 25,* 169–180.

Roy, N., Weinrich, B., Gray, S. D., Tanner, K., Stemple, J. C., & Sapienza, C. M. (2003). Three treatments for teachers with voice disorders: A randomized clinical trial. *Journal of Speech, Language, and Hearing Research, 46,* 670–688.

Roy, N., Weinrich, B., Gray, S. D., Tanner, K., Toledo, S. W., Dove, H., . . . Stemple, J. C. (2002). Voice amplification versus vocal hygiene instruction for teachers with voice disorders: A treatment outcomes study. *Journal of Speech, Language, and Hearing Research, 45,* 625–638.

Rubow, R., & Swift, E. (1985). A microcomputer-based wearable biofeedback device to improve transfer of treatment in parkinsonian dysarthria. *Journal of Speech & Hearing Disorders, 50,* 178–185.

Scott, S., & Caird, F. I. (1983). Speech therapy for Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry, 46,* 140–144.

Spencer, K., Morgan, K., & Blond, E. (2009). Dopaminergic medication effects on the speech of individuals with Parkinson's disease. *Journal of Medical Speech-Language Pathology, 17,* 125–144.

Tjaden, K. (2006, March). *Intelligibility in Parkinson's disease: Effects of speech task*. Presentation at the Conference on Motor Speech, Austin, TX.

Weiss, L. (2002). Speech intelligibility in quiet and noise environments with the Speech Enhancer™ amplification and natural speech. *Journal of Medical Speech-Language Pathology, 10,* 327–331.

Wirdefeldt, K., Adami, H. O., Cole, P., Trichopoulos, D., & Mandel, J. (2011). Epidemiology and etiology of Parkinson's disease: A review of the evidence. *European Journal of Epidemiology, 26*(Suppl. 1), S1–S8.

Yorkston, K., Beukelman, D., & Tice, R. (1996). Sentence Intelligibility Test (Version 1.0) [Computer software]. Lincoln, NE: Communication Disorders Software.

Appendix A

Experience Questionnaire

1. Participants' comfort ratings

Device	Spokeman	Amigo	BoomVox	ADDvox	SoniVox	ChatterVox
Voicette	.011	.049	.008	.042	NS	.048
ChatterVox	NS	NS	NS	NS	NS	
SoniVox	NS	NS	NS	NS		
ADDvox	NS	NS	NS			
BoomVox	NS	NS				
Amigo	NS					

2. Participants' visual presentation ratings

Device	Spokeman	SoniVox	BoomVox	Amigo	ChatterVox	ADDvox
Voicette	.027	.020	.044	NS	.040	NS
ADDvox	NS	NS	NS	NS	NS	
ChatterVox	NS	NS	NS	NS		
Amigo	NS	NS	NS			
BoomVox	NS	NS				
SoniVox	NS					

3. Participants' sound quality ratings

Device	ChatterVox	BoomVox	Amigo	SoniVox	ADDvox	Voicette
Spokeman	.042	NS	NS	NS	NS	NS
Voicette	.020	NS	NS	NS	NS	
ADDvox	NS	NS	NS	NS		
SoniVox	NS	NS	NS			
Amigo	NS	NS				
BoomVox	NS					

4. Participants' amplification ratings

Device	ChatterVox	BoomVox	Amigo	SoniVox	ADDvox	Spokeman
Voicette	.030	NS	NS	NS	NS	NS
Spokeman	.027	NS	NS	NS	NS	
ADDvox	NS	NS	NS	NS		
SoniVox	NS	NS	NS			
Amigo	NS	NS				
BoomVox	NS					

5. Participants' overall preference ratings

Device	Spokeman	ChatterVox	Amigo	BoomVox	SoniVox	ADDvox
Voicette	.006	.001	.016	.006	NS	.013
ADDvox	NS	NS	NS	NS	NS	
SoniVox	NS	NS	NS	NS		
BoomVox	NS	NS	NS			
Amigo	NS	NS				
ChatterVox	NS					

Note. NS = not significant.

Appendix B

Uncorrected Post Hoc *p* Values

1. Speech-to-noise ratios

Condition	BoomVox	Voicette	ChatterVox	Spokeman	Amigo	SoniVox	ADDvox
No device	< .001	.003	.002	.040	.023	NS	.005
ADDvox	< .001	.010	.034	NS	NS	NS	
SoniVox	< .001	.006	NS	NS	NS		
Amigo	< .001	.015	NS	NS			
Spokeman	< .001	.023	NS				
ChatterVox	.001	NS					
Voicette	.004						

2. Speech intensity

Condition	BoomVox	ChatterVox	Voicette	SoniVox	Amigo	Spokeman	ADDvox
No device	< .001	.005	.033	NS	.027	NS	NS
ADDvox	< .001	.003	.020	NS	NS	NS	
Spokeman	< .001	.012	.045	NS	NS		
Amigo	< .001	.017	.040	NS			
SoniVox	< .001	NS	NS				
Voicette	< .001	NS					
ChatterVox	< .001						

3. Intelligibility scores obtained by transcription

Condition	BoomVox	Voicette	ChatterVox	Amigo	Spokeman	SoniVox	ADDvox
No device	< .001	.003	.001	.002	.017	.015	.024
ADDvox	< .001	.017	.011	.014	NS	NS	
SoniVox	< .001	.005	.001	.014	NS		
Spokeman	.003	.033	.020	NS			
Amigo	.002	NS	NS				
ChatterVox	.005	NS					
Voicette	.026						

4. Intelligibility scores obtained by visual analog scale ratings

Condition	BoomVox	Voicette	Spokeman	ChatterVox	Amigo	ADDvox	SoniVox
No device	< .001	.012	.034	.007	.024	NS	NS
SoniVox	< .001	.014	NS	.026	NS	NS	
ADDvox	< .001	.017	.030	.049	NS		
Amigo	.001	NS	NS	NS			
ChatterVox	.016	NS	NS				
Spokeman	< .001	NS					
Voicette	.040						

Note. NS = not significant.