AJSLP

Research Article

The Effect of Noise on Relationships Between Speech Intelligibility and Self-Reported Communication Measures in Tracheoesophageal Speakers

Tanya L. Eadie,[a] Devon Sawin Otero,[a] Susan Bolt,[a] Mara Kapsner-Smith,[a] and Jessica R. Sullivan[a]

Purpose: The purpose of this study was to examine how sentence intelligibility relates to self-reported communication in tracheoesophageal speakers when speech intelligibility is measured in quiet and noise.
Method: Twenty-four tracheoesophageal speakers who were at least 1 year postlaryngectomy provided audio recordings of 5 sentences from the Sentence Intelligibility Test. Speakers also completed self-reported measures of communication—the Voice Handicap Index-10 and the Communicative Participation Item Bank short form. Speech recordings were presented to 2 groups of inexperienced listeners who heard sentences in quiet or noise. Listeners transcribed the sentences to yield speech intelligibility scores.
Results: Very weak relationships were found between intelligibility in quiet and measures of voice handicap and communicative participation. Slightly stronger, but still weak and nonsignificant, relationships were observed between measures of intelligibility in noise and both self-reported measures. However, 12 speakers who were more than 65% intelligible in noise showed strong and statistically significant relationships with both self-reported measures (R^2 = .76–.79).
Conclusions: Speech intelligibility in quiet is a weak predictor of self-reported communication measures in tracheoesophageal speakers. Speech intelligibility in noise may be a better metric of self-reported communicative function for speakers who demonstrate higher speech intelligibility in noise.

Total laryngectomy continues to be a surgical treatment for individuals diagnosed with laryngeal cancer. The procedure results in an altered airway and a need for a new alaryngeal voice source, and leaves individuals to cope with physical, social, and psychological consequences (Doyle & Keith, 2005). After a total laryngectomy, voice and speech rehabilitation is achieved through a variety of methods. Although esophageal and electrolaryngeal speech are still common options, surgical-prosthetic (i.e., tracheoesophageal puncture; TEP) voice restoration is considered by some to be the gold-standard alaryngeal speech method (Kazi, Sayed, & Dwivedi, 2010).

TEP speech may be acquired after a puncture is surgically created between the posterior wall of the trachea and the anterior wall of the esophagus. A tracheoesophageal voice prosthesis is inserted within the puncture, allowing air to move from the trachea to the esophagus when the tracheostoma is occluded upon exhalation (Singer & Blom, 1980). The resulting increase in air pressure causes the pharyngoesophageal segment to vibrate and create a sound source; the sound is then modified by the vocal tract in a manner similar to typical speech production. TEP speech has been shown to be more similar to laryngeal speech than other alaryngeal modes on acoustic parameters such as fundamental frequency, speaking rate, sound level/intensity, and maximum phonation time (Doyle & Keith, 2005; Kazi et al., 2010). Unfamiliar listeners also judge TEP speakers as more intelligible and fluent than speakers who use other alaryngeal speech methods (D'Alatri, Bussu, Scarano, Paludetti, & Marchese, 2012). However, TEP speech is still identified as perceptually different and is less preferred by unfamiliar listeners than typical laryngeal speech (van As, Koopmans-van Beinum, Pols, & Hilgers, 2003).

Despite increased intelligibility among TEP speakers, there is some controversy about whether speakers who use

[a]University of Washington, Seattle
Correspondence to Tanya L. Eadie: teadie@u.washington.edu

Editor: Krista Wilkinson
Associate Editor: Preeti Sivasankar
Received June 17, 2015
Revision received October 29, 2015
Accepted December 22, 2015
DOI: 10.1044/2016_AJSLP-15-0081

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

this method actually report better outcomes in everyday contexts than other alaryngeal speakers (Eadie, Day, Sawin, Lamvik, & Doyle, 2013; Evans, Carding, & Drinnan, 2009; Law, Ma, & Yiu, 2009). Law et al. (2009) examined relationships among listener-rated speech intelligibility and acceptability measures with patient-reported communication outcomes for 49 alaryngeal speakers. Participants in the study used four different alaryngeal speech methods: esophageal speech, electrolaryngeal speech, pneumatic devices, and TEP speech. The researchers noted that there were no significant differences in speech intelligibility scores among esophageal, electrolaryngeal, and TEP speakers. To be specific, the averages across listeners and speaker groups were 58% (esophageal), 60% (TEP), and 70% intelligible (electrolaryngeal). It is notable that the subgroup of 15 speakers who used pneumatic devices showed significantly higher speech intelligibility scores on average (85%). However, whereas the 13 TEP speakers demonstrated significantly lower speech intelligibility and acceptability scores than those using pneumatic devices, these same TEP speakers had the best self-reported communication outcomes among the four groups. The authors concluded that high speech intelligibility does not always translate to high self-reported outcomes. That is, a communication partner's ability to understand a speaker in quiet may not always predict how well that person perceives his or her own communication abilities in everyday contexts.

Many factors may affect self-reported communication success in everyday contexts for people with communication disorders (Baylor, Burns, Eadie, Britton, & Yorkston, 2011; Eadie, 2007). The variables may range from physical symptoms and reduced capacity for performing tasks (e.g., reduced speech intelligibility, fatigue) to individual coping responses (Baylor et al., 2011; Eadie & Bowker, 2012) to changes in body image (Chen et al., 2015) to environmental facilitators and barriers. Factors in the environment may include reactions of communication partners as well as physical barriers such as background noise (Eadie, 2007). Thus, self-reported communication success is multidimensional, and even speech and voice impairments may not necessarily be the most important predictor.

Despite this multidimensionality, one reason past research may have shown uncertain relationships between intelligibility and self-reported communication outcomes may be due to the approaches used in measuring speech outcomes. In particular, listener-rated measures typically used in clinical and research environments (e.g., measures of intelligibility, acceptability) are most often assessed under quiet conditions. In contrast, self-reported outcomes such as voice handicap (e.g., Voice Handicap Index; Jacobson et al., 1997) or communicative participation (e.g., Communicative Participation Item Bank; Baylor et al., 2013) often consider background noise when assessing communication. These self-reported scales include background noise in communication situations because it is regularly encountered in everyday environments. However, it is unknown what role noise plays in affecting relationships between listener-rated and self-reported measures. Before these relationships may be better understood, we must briefly review TEP speaker outcomes.

TEP Speech and Communication Outcomes

Speech Intelligibility

One common measure of alaryngeal speech performance is ratings of intelligibility performed by clinicians and unfamiliar listeners (Bennett & Weinberg, 1973; Law et al., 2009; Miralles & Cervera, 1995; Tardy-Mitzell, Andrews, & Bowman, 1985). Speech intelligibility has been defined as "the degree to which the speaker's intended message is recovered by the listener" (Kent, Weismer, Kent, & Rosenbek, 1989, p. 484). Intelligibility in TEP speakers is primarily affected by poor voice quality, altered speech rate, and the noise associated with improper timing and occlusion of the tracheostoma (van As et al., 2003).

The intelligibility of TEP speech has been shown to be at least equal to or better than that of esophageal speech and better than that of electrolaryngeal speech, usually resulting in high speech performance (Hillman, Walsh, & Heaton, 2005; Tardy-Mitzell et al., 1985). In one study, 46 unfamiliar listeners understood 15 TEP speakers on average 93% of the time (Tardy-Mitzell et al., 1985). Similar results were shown in a recent study that included 16 TEP speakers (mean intelligibility = 92%; range = 73%–100%; Eadie et al., 2013). These results indicate that many TEP speakers are highly intelligible to unfamiliar listeners, one type of listener they would likely encounter on a day-to-day basis.

In summary, results show that TEP speakers usually demonstrate high intelligibility scores as a group, but there is wide variability within this group (Eadie et al., 2013; Hillman et al., 2005; Tardy-Mitzell et al., 1985). Listener-rated intelligibility is important to consider when measuring postlaryngectomy outcomes because it may influence the way a communication partner interacts with a TEP speaker. However, to understand its impact, we need to examine how TEP speakers gauge their own communication in everyday settings.

Self-Reported Measures

To understand the functional consequences of a health condition or disorder, it is essential to include the individual's perspective as part of a comprehensive assessment approach (Eadie, 2007). Self-reported measures have become increasingly valued and common in health care and research, and may include a composite measure of a person's well-being, such as quality of life (QOL). QOL measures may be disease specific (e.g., relating to all symptoms of head and neck cancer) or discipline specific (e.g., voice-specific QOL or voice handicap, measuring the psychosocial impact of voice-specific symptoms). Because disease-specific QOL measures may ask only one or two particular questions regarding certain functions, a voice- or speech-specific QOL measure is better suited to capturing QOL as it relates to the voice or speech symptoms.

One commonly used voice-specific QOL measure is the Voice Handicap Index (VHI; Jacobson et al., 1997) and its validated short form, the Voice Handicap Index–10 (VHI-10; Rosen, Lee, Osborne, Zullo, & Murry, 2004). This self-reported tool has been used to measure outcomes in individuals who have undergone total laryngectomy, including TEP speakers (Azevedo, Montoni, Filho, Kowalski, & Carrara-de Angelis, 2012; Evans et al., 2009; Kazi et al., 2007; Oridate et al., 2009). These studies have shown that many TEP speakers demonstrate a chronic mild–moderate voice handicap postlaryngectomy (Azevedo et al., 2012; Evans et al., 2009; Kazi et al., 2007; Oridate et al., 2009).

Increased attention has recently been paid to developing a self-reported measure that goes beyond voice-related symptoms and also captures speech-related difficulties in everyday settings. The Communicative Participation Item Bank (CPIB; Baylor et al., 2013; Baylor, Yorkston, Eadie, Miller, & Amtmann, 2009; Eadie et al., 2014) is a validated self-report instrument that was designed to measure communicative participation, which is defined as "taking part in life situations in which knowledge, information, ideas or feelings are exchanged" (Eadie et al., 2006, p. 309). For example, items include talking to a clerk in a store, negotiating a raise at work, and ordering food in a restaurant.

The construct validity of the CPIB and its 10-item short form was demonstrated recently in a large sample of individuals with communication disorders, including 66 individuals who used TEP as their primary method of communication (Baylor et al., 2013; Eadie et al., 2014). TEP speakers reported worse voice handicap (measured by the VHI-10) and worse communicative participation than those who had also undergone treatment for head and neck cancer but used natural speech for communication (Eadie et al., 2014). As hypothesized, a strong relationship was found between the self-reported measures (VHI-10 vs. CPIB; $r = -.79$). However, to better understand factors that predict self-rated communication success, it is important to determine the strength of the association between traditional voice and speech measures and self-reported outcomes. In particular, we need to consider the role that noise might play in affecting these measures, as well as the relationships among them.

Relationships Between Speech Intelligibility and Self-Reported Measures: Effect of Noise?

Some studies in the literature on head and neck cancer have examined the relationship between speech intelligibility or voice quality and self-reported measures, such as QOL specific to head and neck cancer or to the voice (Eadie & Doyle, 2004; Meyer et al., 2004). In general, these studies have found weak to moderate relationships between these measures (e.g., Law et al., 2009; Meyer et al., 2004). These findings are consistent with those reported in the motor speech literature for dysarthria secondary to Parkinson's disease (Donovan, Kendall, Young, & Rosenbek, 2008) and traumatic brain injury (McCauliffe, Carpenter, & Moran, 2010).

We previously investigated associations among alaryngeal speech intelligibility with disease- and voice-specific QOL measures in 25 individuals who had undergone total laryngectomy and used different alaryngeal speech modes (Eadie et al., 2013). Results of this study revealed weak correlations among the measures: The relationship between speech intelligibility (percentage of words understood from the Sentence Intelligibility Test; Yorkston, Beukelman, & Tice, 1996) and the VHI-10 was extremely weak to nonexistent ($r = .04$). The correlation between speech intelligibility and one item measuring self-rated speech understandability on a disease-specific scale was also weak ($r = .22$). Together, these results suggest that listener-rated intelligibility may not strongly predict self-rated communication success within an alaryngeal population.

Relatively weak relationships have been found between speech intelligibility and self-reported measures, but we need to consider how they are both obtained. Clinical assessment of speech intelligibility, for people with speech disorders (vs. hearing disorders), is typically performed in a quiet environment. Yet in daily life, events often occur in suboptimal listening conditions, which may negatively affect a communication partner's ability to process the speech signal during a communication exchange. For example, it is well known that presence of noise adversely affects typical speech intelligibility (e.g., Sperry, Wiley, & Chial, 1997; Van Engen & Bradlow, 2007).

How background noise affects speakers with speech and voice disorders has had limited study (e.g., Bunton, 2006; McColl, Fucci, Petrosino, Marin, & McCaffrey, 1998; Tjaden, Sussman, & Wilding, 2014). Even when noise is added to a speech intelligibility assessment, it is often used as an approach to reduce ceiling effects of speech performance, which assumes equal reduction across speakers (Bunton, 2006). However, results from limited studies appear to show that background noise may differentially penalize those with already compromised speech intelligibility (McColl et al., 1998; McAuliffe, Schaefer, O'Beirne, & LaPointe, 2009). For example, in the dysarthria literature, one study has found that for three speakers with a variety of dysarthria types, background noise affected their intelligibility more than speakers without dysarthria (McAuliffe et al., 2009).

McColl et al. (1998) performed one of the few studies investigating the effect of noise on TEP speech intelligibility. They recorded speech samples from one typical laryngeal speaker and one superior TEP speaker, who read two sentences from a reading passage. Fifty listeners were asked to subjectively judge how well they understood the speech sample using a rating scale. The speech samples were presented at nine signal-to-noise ratio levels; the noise masker included multitalker babble. Results of the study revealed that the TEP speaker was significantly less intelligible than the laryngeal speaker across all conditions, and that both speakers had significantly lower intelligibility in noise. There was a significant interaction, however, between the variables. The TEP speaker's intelligibility was more affected by noise than was that of the typical laryngeal speaker.

The results of these studies appear to suggest that individuals with voice or speech disorders, such as TEP speakers, may be more susceptible to speech degradation in noise than speakers without communication disorders. This is an important factor to consider because most TEP speakers report particular difficulty communicating in noise in everyday environments (Baylor et al., 2011; Op de Coul et al., 2005). The purpose of this study, therefore, was to investigate how TEP speech intelligibility relates to self-reported communication measures when speech intelligibility is measured in quiet and in noise. It was hypothesized that the addition of noise to speech intelligibility measures would strengthen relationships among measures. However, because self-reported communication success is affected by multiple factors (e.g., coping strategies, topic of conversation, familiarity of communication partner), it also was expected that these relationships would remain moderate even after controlling for noise.

Method

This study included two groups of subjects: TEP speakers and inexperienced listeners. All participants were native English speakers, and none reported any other speech, language, or voice symptoms beyond those associated with laryngectomy in the TEP speakers. The listeners all passed hearing screening tests at 20 dB for the octave frequencies of 250 to 8000 Hz. All participants were paid for their participation. The University of Washington Institutional Review Board approved the procedure in this study.

Subjects

TEP Speakers

Twenty-four individuals (21 men, three women) who had undergone total laryngectomy secondary to cancer were recruited through support groups, professional e-mailing lists, and professional contacts. The mean age of the TEP speaker group was 64 years (range: 39–86). Individuals had undergone total laryngectomy at least 1 year prior to participation in this study, to allow time for adjustment and adaption of their new speech method (Campbell, Marbella, & Layde, 2000). Only individuals who reported using TEP speech as their primary method of communication were included in this study.

Listeners

Sixty-six inexperienced listeners (50 women, 16 men) were recruited from among the student population at the University of Washington and the broader Seattle community. The average age of the listeners was 24 years (SD = 5.4), with a range of 19–45. They were all inexperienced in that they were individuals with no prior experience with or course work related to alaryngeal speech.

Speaker Data Collection

Demographic Measures

Speakers completed a set of questionnaires that included demographic information related to age, sex, ethnicity, education level, living situation, date of total laryngectomy, type of cancer treatment, and primary speech method. They also completed two self-reported measures: the VHI-10 (Rosen et al., 2004) and the CPIB short form (Baylor et al., 2013). Questionnaires were completed at the time of the speech recordings or were returned within a 2-month window of recording.

VHI-10

The VHI-10 is a validated 10-item questionnaire that measures the impact of voice disorders, such as those experienced secondary to total laryngectomy, on voice-related QOL (Rosen et al., 2004). This self-reported tool has been used to measure outcomes in TEP speakers (Kazi et al., 2007; Oridate et al., 2009). Respondents indicate how frequently they have had an experience using a 5-point Likert scale. Items are summed to derive a composite score that ranges from 0 (*minimal voice handicap*) to 40 (*significant/severe voice handicap*).

CPIB Short Form

The CPIB short form is a validated 10-item instrument derived from the CPIB, an item bank validated using item response theory (Baylor et al., 2009, 2013). The CPIB short form measures communicative participation in community-dwelling adults with a range of speech-related communication disorders. Validation studies have included individuals who have undergone total laryngectomy (Baylor et al., 2013; Eadie et al., 2014). The CPIB short form items ask individuals to rate how much their condition (e.g., laryngectomy) interferes with participation in a wide range of daily speech communication activities using a 4-point Likert scale (3 = *not at all*, 2 = *a little*, 1 = *quite a bit*, 0 = *very much*). For example, items include talking to people you do not know, ordering a meal in a restaurant, talking in groups of people, and having a conversation in a noisy place.

Baylor et al. (2013) report a strong and significant correlation between scores on the CPIB full item bank and the CPIB short form ($r = .971$, $p < .001$), demonstrating strong concurrent validity for the short form. CPIB short form summary scores are derived by summing scores from across the 10 items; total scores range from 0 (*severely restricted communicative participation*) to 30 (*high levels of communicative participation*). A person (theta) score is then derived using a translation table, on the basis of a normed sample (Baylor et al., 2013). Scores typically range from −3.0 to +3.0 logits, with 0 representing the mean of the sample used for item-bank calibration; higher scores are better.

Speech Recordings and Preparation

TEP speakers provided speech recordings that included sentences from the Sentence Intelligibility Test (SIT; Yorkston et al., 1996). Sentences for each speaker were randomly chosen and comprised sentences of seven, nine, 11, 13, and 15 words in length, resulting in 55 words per speaker. Similar sample lengths (e.g., 50–65 words) have been used to measure intelligibility in previous research (Hustad, 2008; Nagle, Eadie, Wright, & Sumida, 2012).

Speech samples were recorded in a sound-treated room using a headset microphone (Shure PG-81, Shure, Niles, IL; or AKG-C20, AKG Acoustics, Vienna, Austria) with a 3-cm (offset) mouth-to-microphone distance. The microphone was connected to a preamplifier (M-Audio Fast Track Pro, Avid Technology, Burlington, MA), and the sample was acquired on a laptop computer using a specialized sound card and acoustic software (Sona-Speech II, Model 3650, KayPENTAX, Montvale, NJ). Additional samples were obtained in the same type of room, using the same microphone and mouth-to-microphone distance, but on a portable digital audiotape recorder (TASCAM DAP1, TASCAM, Montebello, CA). All speech samples were recorded at a sampling rate of 44.1 kHz with 16-bit quantization. Speech samples were transferred to a computer and converted into WAV files using acoustic software (Sony Soundforge, Sony Creative Software Inc., Middleton, WI).

To control the signal-to-noise ratio (SNR), each sentence from the SIT was equated for peak amplitude (root-mean-square normalized) using sound-editing software (Sony Soundforge). One set of sentences was saved as stimuli to be presented in quiet, with the speech signal held constant at the same root-mean-square amplitude (65 dBA SPL; McColl et al., 1998). The same sentences were then mixed with four-talker babble (one male speaker and three female; Audiotec of St. Louis) from the QuickSIN Speech-in-Noise test (Version 1.3, Etymotic Research, Elk Grove Village, IL) to create a noise set. Multitalker babble was selected as the noise because previous research has found that meaningful speech competitors had a significantly more adverse effect on word recognition performance compared with nonmeaningful competitors (e.g., white noise; Sperry et al., 1997). Multitalker babble also is representative of the most challenging adverse listening environment encountered in everyday speech communication situations (Gilbert, Tamati, & Pisoni, 2013).

Speech samples were mixed with multitalker babble at an SNR of +6 dB for each sentence. The speech signal was held constant at a level of 65 dB SPL, and the SNR was adjusted by reducing the noise level. Each sentence started with 500 ms of noise alone, followed by the speech signal mixed with the noise, and finally 500 ms of noise at the end of each trial (Van Engen & Bradlow, 2007). The +6 dB SNR level was identified from the McColl et al. (1998) study and was pilot tested to ensure no ceiling or floor effects. The WAV files were entered into a software program (EcosWin, Avaaz Innovations, London, ON, Canada) that randomizes the presentation order of the speech samples and allows listeners to transcribe the speech sample (intelligibility).

Listener Procedure

Listeners were randomized into two groups: One group transcribed SIT sentences in quiet, and the other transcribed the same SIT sentences presented at +6 dB SNR (i.e., the noise condition). Whereas group assignment was random, the experimenters also matched the proportion of female to male listeners between the groups to control for any possible gender effects. Because the groups transcribed the same sentences (i.e., one group transcribing the sentences in quiet, one group transcribing the same sentences in noise), a between-subjects design was used. This design ensured that listeners were not exposed to the same sentence more than once.

Before the transcription task, both groups of listeners were provided instructions about the task and the speech samples as follows:

> You will be listening to adult speakers who have had total removal of their voice box due to cancer. These speakers are using a new method of speech called "tracheoesophageal speech." We are interested in how well listeners can understand these speakers in both quiet and background noise. You will only hear samples presented in [quiet or noise]. We will play some sentences, and we would like you to type out the words that you hear. You may listen to the sentences up to two times. Some of these sentences will be difficult to understand. Do your best, and guess when you need to. You may listen to each sentence two times.

Listeners in both groups were presented several samples produced by TEP speakers who were not otherwise included in the study. Then they were exposed once to the first speaker they would transcribe. This familiarization sample included the speaker reading a novel five-word sentence in quiet that immediately preceded the transcription task. Listeners in the noise group were informed that this same speaker would be presented along with several other talkers in the background. This familiarization protocol was used to prepare listeners in both groups for the general quality of TEP speech, the speaker they were being asked to transcribe, and the presence of multitalker babble (for the noise group only). Although this type of brief familiarization (i.e., fives words of audio only from a specific speaker) has not shown consistent effects on intelligibility (Hustad & Cahill, 2003; Yorkston & Beukelman, 1983), equal exposure between the two listener groups ensured that any potential differences between the groups would be systematic. This procedure (presentation of the familiarization sample followed by the transcription of five SIT sentences per speaker) was repeated two more times for each listener, for a total of three speakers per listener.

Listeners in each group (quiet or noise) provided a score of speech intelligibility (percentage of words understood) for 15 sentences (165 words total). Different speakers were randomly assigned to each listener. No sentences were repeated across speakers (within each listener set), to control for learning effects. In a few cases, listeners transcribed sentences for only two speakers (110 words total), to avoid any sentence repetition. The recordings were presented over headphones (Samson RH600 Samson Technologies, Hauppauge, NY), with the speech signal held at 65 dBA SPL, which was verified by measuring the output of the speech signals in quiet from the headphones using an AEC100 Acoustic Ear (Larson Davis, PCB Piezotronics, Inc., Provo, UT) coupled to a sound-level meter (Larson Davis Model 824). Listeners were asked to orthographically transcribe the sentences they heard into a software program on a desktop computer (EcosWin, Avaaz Innovations, London, ON, Canada). Each speaker's intelligibility judgments were based on an average score derived from three listeners using a total-word phonemic-match model in scoring (Hustad & Cahill, 2003).

Reliability of Listeners' Transcriptions

Measures of intrarater reliability were not included in this study, due to learning effects with presentation of a repeated sentence. To assess interrater reliability of transcriptions for each set of three speakers evaluated by the listeners, intraclass correlation coefficients were calculated. For listeners in quiet, the mean intraclass correlation coefficient across listeners was .72 (95% CI [.61, .79]). For listeners in noise, the mean intraclass correlation coefficient was .83 (95% CI [.77, .88]). These levels are consistent with prior research and are acceptable levels for further data analysis (Sussman & Tjaden, 2012).

Data Analysis

The predictor variables in this study were speech intelligibility in quiet and in noise. The predicted variables were self-reported measures of the VHI-10 (total score) and CPIB short form (person/theta scores). Each speaker's intelligibility for each stimulus set (quiet or noise) was based on the percentage correctly understood out of five sentences, averaged across three listeners (55 words × 3 listeners per speaker = 165 words per speaker). Using the average of three listeners in this analysis helps reduce the influence of potential outlying listener ratings on the predictor variable of intelligibility (Shrivastav, Sapienza, & Nandur, 2005). Total scores for the VHI-10 and theta scores for the CPIB short form were then obtained by adding up all responses to items in each self-reported measure for each speaker, and then transforming the summary scores to theta scores for the CPIB short form (Baylor et al., 2013). To determine relationships between the variables, four Pearson correlations were calculated: between intelligibility in quiet or in noise and the VHI-10 total scores or CPIB short form theta scores. The strength of relationships also was assessed using variance scores (R^2).

Results

Demographics of the TEP Speakers

The 21 male and three female TEP speakers were on average 6.91 years (range: 1–18) postlaryngectomy. Sixty-seven percent (n = 16) received radiation treatment, 75% had completed at least some college education, 79% (n = 19) lived with family, and the majority (92%) were White (n = 22). Additional demographic information about the TEP speakers is presented in Table 1.

Summary Scores

VHI-10 and CPIB Short Form

Overall, the TEP speakers reported a mean VHI-10 total score of 16.52 (SD = 7.02), consistent with a moderate voice handicap. One speaker did not complete the VHI-10; as a result, the mean VHI-10 total score is based on data from 23 speakers. Data from all 24 TEP speakers were obtained for the CPIB short form. On average, they reported summary scores of 20.67 (SD = 6.17), which corresponds to a mean θ score of 0.38 (SD = 0.83). Mean scores and data for individual speakers are presented in Table 2.

Speech Intelligibility in Quiet and in Noise

The mean speech intelligibility score for TEP speakers in quiet was 93.27% (SD = 5.70%); for speech intelligibility in noise, the mean was 68.64% (SD = 17.56%). Scores revealed an average decrease of 24.63% in intelligibility with the introduction of background noise. For four out of 24 speakers (M, N, P, SC), intelligibility did not appear to change (less than 5% decrease) in noise. Data from all 24 speakers are presented in Table 2.

Relationships Between Intelligibility and Self-Reported Measures

Before performing statistical analyses, scatter plots demonstrating the relationships between intelligibility and the self-reported scores were visually inspected. They revealed a wide distribution of scores but included one speaker (Q) who was a significant outlier from the group (e.g., z = 3 or more for predictor and predicted measures; Osborne & Overbay, 2004). In accordance with recommendations for outliers (Osborne & Overbay, 2004), analyses were performed with this speaker's data included as well as removed.

To ensure that relationships among variables were not affected by the distribution of scores, data also were log transformed and the strength of relationships among variables examined (Keene, 1995). No changes in relationships were observed with use of the transformed scores; changes in relationships were $r < .03$, which is equivalent to less than 0.10% of the variance that changed as a result of using log transformation. As a consequence, all analyses reported in this study include mean scores for intelligibility (%), VHI-10 total scores, and CPIB short form theta scores.

Relationships Between Intelligibility and the VHI-10

Relationships between intelligibility in quiet and in noise and VHI-10 scores were both weak and nonsignificant ($r = .18$, $p > .05$ for intelligibility in quiet; $r = -.25$, $p > .05$ for intelligibility in noise). Using a cutoff score of z = 3 (Osborne & Overbay, 2004), one outlier speaker (Q) was removed. The relationships between intelligibility and the VHI-10 total scores are shown in Figure 1 for the remaining 23 TEP speakers. Overall, the relationship between intelligibility in quiet and VHI-10 scores remained nonsignificant

Table 1. Demographics of tracheoesophageal speakers.

Characteristic	No. (%)	*M* (*SD*)	Range
Sex			
Male	21 (87.50)		
Female	3 (12.50)		
Ethnicity			
White	22 (91.66)		
American Indian/Alaskan Native	1 (4.17)		
Other	1 (4.17)		
Age (years)		63.75 (10.06)	39–86
Living situation			
Alone	5 (21)		
With family	19 (79)		
Education			
High school graduate	5 (21)		
Some college	11 (46)		
College graduate	6 (25)		
Postgraduate	1 (4)		
Not reported	1 (4)		
Time since total laryngectomy (years)		6.91 (5.56)	1–18
Cancer treatment			
Surgery alone	2 (8)		
Surgery + radiation	16 (67)		
Surgery + radio(chemo)therapy	5 (21)		
Not reported	1 (4)		

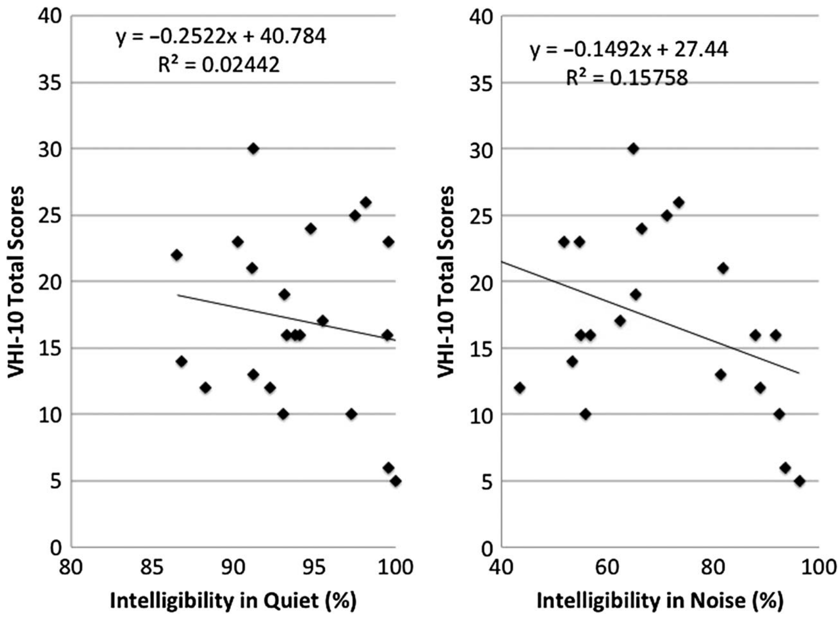
Table 2. Summary results across TEP speakers: mean (and standard deviation) intelligibility in quiet and in noise, and self-reported scores (VHI-10 total and CPIB short form person/theta score).

Speaker	Intelligibility in quiet	Intelligibility in noise	VHI-10 total	CPIB short form person/theta
D	99.56 (1.72)	54.77 (40.55)	23	–0.33
G	90.25 (15.01)	51.87 (41.25)	23	–0.89
GG	91.21 (12.80)	81.41 (27.64)	13	0.92
I	94.77 (7.29)	66.58 (28.54)	24	–0.89
II	97.48 (5.24)	71.22 (24.19)	25	–0.56
J	97.27 (5.06)	55.81 (35.07)	10	1.22
JJ	99.49 (1.99)	54.98 (35.62)	16	1.42
L	94.12 (6.51)	56.72 (28.53)	16	0.27
M[a]	92.24 (6.43)	88.83 (10.04)	12	1.67
N[a]	93.78 (10.79)	91.77 (11.67)	16	0.78
O	91.14 (11.68)	81.94 (17.20)	21	0.27
P[a]	93.08 (12.58)	92.55 (12.59)	10	1.06
Q[b]	74.13 (19.17)	45.97 (27.80)	4	1.42
S	86.53 (15.99)	37.23 (28.13)	22	0.15
SC[a]	100.00 (0.00)	96.26 (7.12)	5	2.10
SE	86.82 (15.33)	53.29 (34.02)	14	–0.10
SG	93.14 (14.22)	65.39 (24.79)	19	0.27
SH	95.46 (5.24)	62.35 (37.66)	17	0.03
SM	88.29 (12.22)	43.32 (27.21)	12	1.22
SN	98.18 (3.76)	73.40 (35.30)	26	–0.22
SO	93.28 (13.22)	87.94 (20.65)	16	0.27
SP	99.56 (1.72)	93.58 (11.54)	6	0.78
T	97.51 (4.20)	75.21 (26.86)	—	–0.78
V	91.20 (12.68)	64.98 (14.95)	30	0.015
M (*SD*)	93.27 (5.70)	68.64 (17.56)	16.52 (7.02)	0.43 (0.84)

Note. Intelligibility is given as mean percentage, with standard deviation in parentheses. An em dash indicates incomplete/nonreported data. VHI-10 = Voice Handicap Index–10; CPIB = Communicative Participation Item Bank.

[a]Speakers whose intelligibility changed less than 5% from quiet to noise. [b]The outlier speaker whose data were removed from analyses.

Figure 1. Relationship between VHI-10 total scores and (left) intelligibility in quiet and (right) intelligibility in noise. Each dot represents a single speaker. A line of best fit and variance scores also are reported on the graphs.



and weak ($r = -.16$, $p > .05$). The relationship between intelligibility in noise and the VHI-10 totals was statistically significant, but only moderate in strength ($r = -.40$; $p < .05$). Relationships were in the predicted directions for the sample of 23 TEP speakers; that is, as intelligibility increased, voice handicap decreased (see Figure 1). Lines of best fit and predicted variance (R^2) values are also shown in Figure 1.

Relationships Between Intelligibility and the CPIB Short Form

Relationships between intelligibility in quiet and in noise and CPIB short form theta scores were also weak and nonsignificant ($r = -.11$, $p > .05$ for intelligibility in quiet; $r = .24$, $p > .05$ for intelligibility in noise). Using a cutoff score of $z = 3$ (Osborne & Overbay, 2004), the outlier speaker Q was subsequently removed. The relationship between intelligibility in quiet and CPIB short form theta scores for the remaining 23 TEP speakers also was nonsignificant and weak ($r = .10$, $p > .05$). Intelligibility in noise showed a relatively stronger relationship with CPIB short form theta scores, yet the association was also nonsignificant, and only weak to moderate in strength ($r = .34$, $p > .05$). As with the VHI-10, the relationships were in the hypothesized directions; as intelligibility increased, CPIB short form theta scores also increased, although this was a particularly weak (almost nil) relationship when speech intelligibility was measured in quiet (see Figure 2). Lines of best fit and predicted variance (R^2) values are shown in Figure 2.

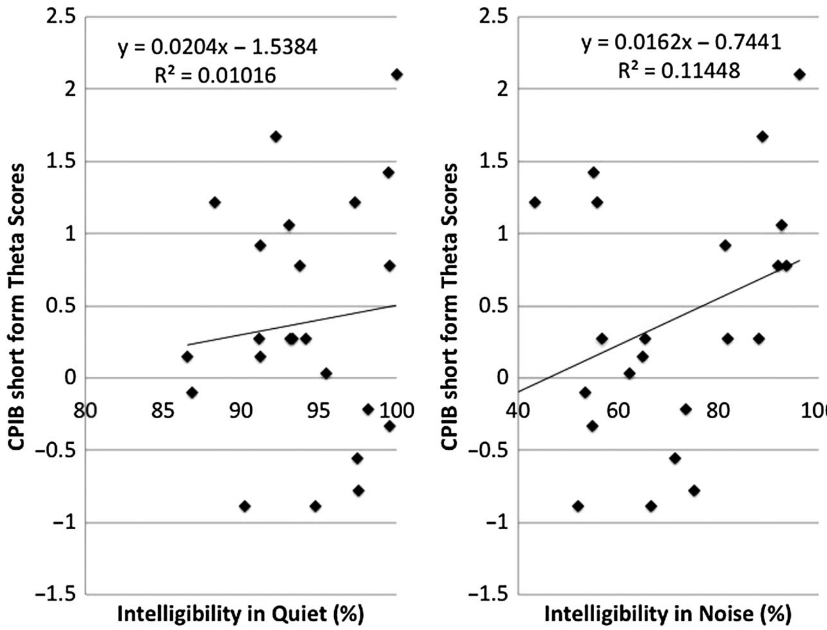
Post Hoc Analysis

Per recommended procedures for correlational analyses (Portney & Watkins, 2000), the data were visually inspected. The right side of the scatter plots presented in Figures 1 and 2 revealed a notable pattern: Above 65% intelligibility in noise, the plots for both the VHI-10 and the CPIB short form data show linear relationships. In consequence, the data were reanalyzed at or below 65% (criterion) and above 65% to better describe these different patterns of relationships between intelligibility in noise and the self-reported measures. The relationships are shown in Figures 3 and 4, along with the lines of best fit and predicted variance (R^2).

As observed in Figure 3, the relationships between intelligibility in noise (at or below 65%) and the self-reported measures were weak and nonsignificant ($r = .21$, $p > .05$ for the VHI-10; $r = -.11$, $p > .05$ for the CPIB short form). These results are consistent with those found in the total sample of speakers (see Figure 2). For this subset of 11 speakers, relationships were not in the predicted directions; that is, as speech intelligibility increased, voice handicap increased and communicative participation decreased. However, these relationships were extremely weak, predicting only 1% to 4% of the variance (R^2) in the self-reported scores.

As observed in Figure 4, the relationships between intelligibility in noise (above 65%) and the self-reported measures were much stronger than relationships for all speakers and for the subset of speakers who were at or less than 65% intelligible in noise (see Figures 1–3). There was a strong and statistically significant relationship shown between intelligibility in noise and the VHI-10 total scores ($r = -.89$, $p < .05$) for 12 speakers who were more than 65% intelligible in noise. As intelligibility in noise increased, VHI-10 scores significantly decreased (i.e., decreased voice handicap). The relationship between intelligibility in noise and the CPIB

Figure 2. Relationship between CPIB short form theta scores and (left) intelligibility in quiet and (right) intelligibility in noise. Each dot represents a single speaker. A line of best fit and variance scores also are reported on the graphs.



short form scores also was strong and statistically significant ($r = .87$, $p < .05$) for the same 12 speakers who were more than 65% intelligible. As intelligibility in noise increased, communicative participation also strongly and significantly increased for these 12 TEP speakers.

Discussion

This study investigated the relationship between speech intelligibility and self-reported voice handicap and communicative participation. The main question was whether these

Figure 3. Relationship between intelligibility in noise and (left) VHI-10 total scores and (right) CPIB short form theta scores in speakers with 65% or less intelligibility (n = 11) . Each dot represents a single speaker. A line of best fit and variance scores also are reported on the graphs.

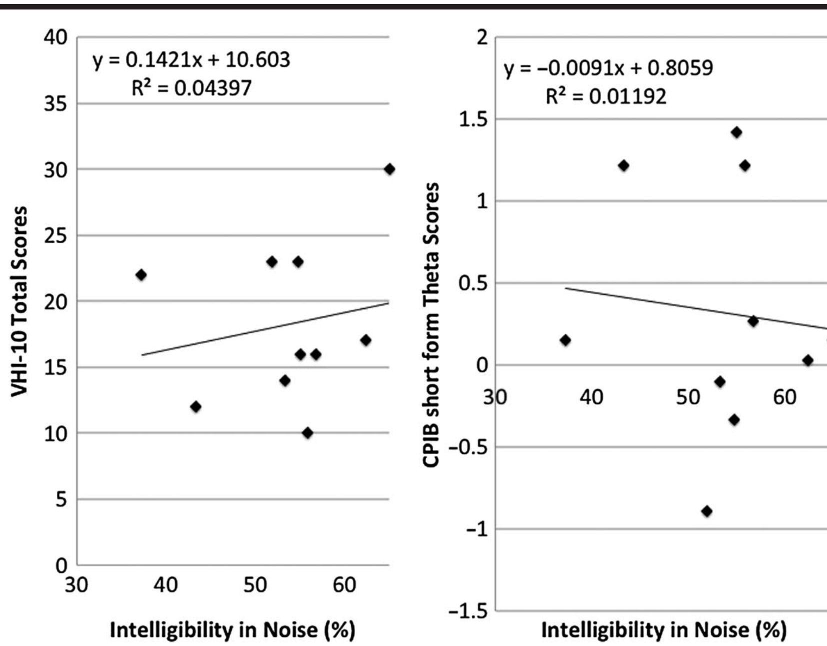
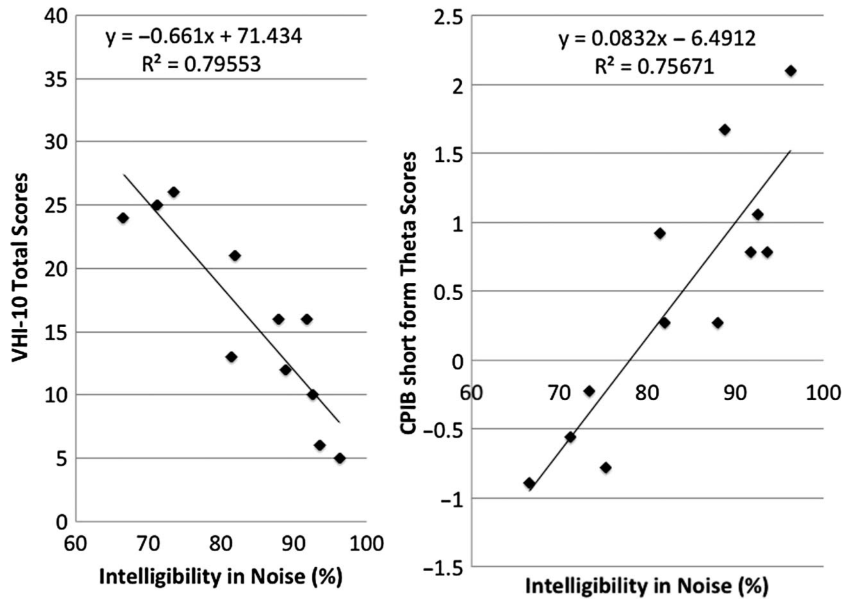


Figure 4. Relationship between intelligibility in noise and (left) VHI-10 total scores and (right) CPIB short form theta scores in speakers with greater than 65% intelligibility (n = 12). Each dot represents a single speaker. A line of best fit and variance scores also are reported on the graphs.



relationships were stronger when intelligibility was measured in quiet, as is standard practice, or in noise, which might better reflect adverse conditions in everyday life. Results revealed weak, nonsignificant relationships between intelligibility in quiet and the VHI-10 and the CPIB short form scores. Slightly stronger (but still weak to moderate, nonsignificant) relationships were observed between measures of intelligibility in noise and both self-reported measures. However, half the speakers' data revealed a strong linear pattern. Scores from 12 speakers who were more than 65% intelligible in noise strongly and significantly related to both the VHI-10 and CPIB short form scores, with approximately 76%–79% of the variance predicted. These results suggest that with the addition of background noise, intelligibility scores may be a strong predictor of voice handicap and communicative participation for a select group of TEP speakers. How these results compare to previous findings, and what this means for future research and clinical practice, is discussed next.

Summary Scores

Overall, results from this study revealed moderate voice-handicap scores among the TEP speakers (M = 16.52, SD = 7.02). These values are comparable to those reported in the literature for similar types of alaryngeal speakers (Azevedo et al., 2012; Evans et al., 2009; Lundström, Hammarberg, & Munck-Wikland, 2009). The CPIB short form theta scores for the 24 TEP speakers (M = 0.43, SD = 0.84) were also relatively consistent with those of a larger independent group of 66 TEP speakers who completed the entire item bank from the CPIB (M = 0.20, SD = 0.82; Eadie et al., 2014). These results strengthen the external validity of our findings.

In this study, speech intelligibility in quiet ranged from 74.13% to 100.00% (M = 93.27%) across the speakers, which is consistent with previous research (Eadie et al., 2013; Tardy-Mitzell et al., 1985). As hypothesized, speech intelligibility in noise was lower than in quiet, similar to a previous study of esophageal and electrolaryngeal speakers (Holley, Lerman, & Randolph, 1983). The reduction in intelligibility (from 93% to 69%) in a +6 dB SNR condition is a substantive decrease that may reveal the sensitivity of disordered speech to adverse conditions. McAuliffe et al. (2009) showed that across similar noise conditions and with similar stimuli, three adult male speakers in a control group did not show a decrease in intelligibility. Instead, their averages remained consistent (above 90%) when speakers were presented in a no-noise condition and a +6 dB SNR condition. However, similar to the majority of TEP speakers in the present study, the three dysarthric speakers in that study were significantly affected in relatively low-noise conditions (+6 dB SNR) and showed a large, significant decrease in intelligibility. These results support the contention that speakers with communication disorders, including those using TEP speech, may be differentially penalized in difficult listening environments. Future studies should examine these effects in different types and levels of noise (i.e., beyond +6 dB SNR and multitalker babble). In addition, in lieu of controlling the SNR, speech samples could be presented at their naturally varying sound pressure levels in the presence of a constant noise source, which might better reflect some everyday environments.

Consistent with the present study, McColl et al. (1998) also found that their one superior TEP speaker was adversely affected in noise. They noted large differences in performance between the TEP speaker and a typical laryngeal speaker when noise was presented at a +6 dB SNR. It is interesting to note that for four out of 24 speakers in this study (M, N, P, SC; see Table 2), intelligibility appeared to remain relatively constant from quiet to noise conditions. This result is similar to how the (nondysarthric) speakers in the control group performed in the study by McAuliffe et al. (2009). These findings suggest that there may be some TEP speakers who have speech characteristics more similar to those of typical laryngeal speakers that may help listeners overcome a masker such as multitalker babble. Future studies should include acoustic analyses of these speech samples to determine whether there are overall characteristics that optimize intelligibility in noisy backgrounds. Results from these studies could have implications for future speech rehabilitation approaches.

Although it might be assumed that highly intelligible speakers in quiet would also show better intelligibility in noise, some speakers in this study showed a different pattern. To be specific, some speakers who were more than 95% intelligible in quiet dropped more significantly in noise (i.e., larger change scores) than speakers who were less than 95% intelligible. As with speakers who may be able to overcome noise, future studies should investigate characteristics of speakers who are particularly vulnerable in adverse conditions. For example, McColl (2006) found that activation of a Lombard effect may actually detract from overall speech intelligibility for some TEP speakers: Whereas typical laryngeal speakers may modulate the changes associated with the Lombard effect and maintain speech intelligibility in noise, three out of four TEP speakers in that study were negatively affected by speaking in noise. Recording the TEP speakers in noise (as opposed to adding the noise experimentally) could therefore be an important avenue for future research. This would allow further study of respiratory, neoglottal sound source, and articulatory contributions to improved (or reduced) intelligibility in noise, and thereby identify and tailor targets for intervention. For example, some researchers (McColl, 2006; Searl, 2007) have suggested that adaptations such as those found in "clear speech" could help promote intelligibility among alaryngeal speakers, because they have been shown to be effective for both typical speakers and those with dysarthria (Tjaden et al., 2014). However, Searl (2007) cautioned that it is unknown whether there may also be negative effects for TEP speakers who already expend greater physical effort during speech production, with increased respiratory and oral pressures generated during articulation.

In addition to speaker factors, it is important to consider how intelligibility measures are obtained. Factors related to the listener must be considered, including short-term memory, listener reliability, listener experience, and predictability of the sentences. For example, in this study, sentences were selected at random from the SIT (Yorkston et al., 1996); but it is possible that some sentences were more predictable than others, which could allow listeners to use more top-down strategies to increase performance (Beverly et al., 2010). We must also consider the fact that speech intelligibility is not synonymous with listener comprehension (Hustad, 2008), nor with the amount of effort a listener expends to decode a message (Nagle & Eadie, 2012). These variables could affect the social interaction between a speaker and his or her communication partner, and might be better predictors of self-reported outcomes than intelligibility alone.

Listener selection must also be considered when performing intelligibility studies. For example, Doyle, Swift, and Haaf (1989) found that experienced clinicians judged four TEP speakers as significantly more intelligible than inexperienced listeners did when they evaluated speech samples in quiet. This factor must be considered when experienced listeners are often the ones who conduct these types of assessments in clinical environments. As a final matter, we must also consider how intelligibility is measured. In this study, we used a transcription procedure involving sentences from a standard clinical instrument (Yorkston et al., 1996). However, Sussman and Tjaden (2012) found that scaled estimates for speech severity are also needed to characterize speech impairment. In addition, other researchers have used stimuli that permit analyses of phonemic and phonetic errors that might clarify the effect of multitalker babble on TEP speech intelligibility (Doyle et al, 1989; Miralles & Cervera, 1995). This type of analysis might also ease comparison with the hearing literature, which uses different assessment tools and criteria for determining speech recognition and intelligibility levels (Van Engen & Bradlow, 2007). All of these factors need consideration in designing future studies and examining relationships between intelligibility and self-reported measures.

Relationships Between Intelligibility and Self-Reported Measures

This study investigated the relationship between listener-rated intelligibility in quiet and in noise and self-reported voice and communication outcomes in TEP speakers. The weak relationships found between intelligibility in quiet and the self-reported outcomes for 23 TEP speakers ($r = -.16$ for the VHI-10; $r = .10$ for the CPIB short form) were expected. Meyer et al. (2004) similarly found no significant association between sentence intelligibility and speech items on QOL scales in their group of people who had received laryngectomies. Likewise, Eadie et al. (2013) found weak correlations between speech intelligibility and the VHI-10 among 25 alaryngeal speakers ($r = .042$). Donovan et al. (2008) also report a weak correlation ($r = .35$) between speech intelligibility measured by unfamiliar listeners and self-reported communicative effectiveness in individuals with Parkinson's disease.

Results from the present study also revealed only weak to moderate relationships between speech intelligibility in noise and self-rated outcomes for 23 speakers ($r = -.40$, $p < .05$ for the VHI-10; $r = .34$, $p > .05$ for the CPIB short form). One possible reason for this relative increased strength in the relationship between measures is that speaking with noise in the background is more representative of daily communication environments than speaking in quiet. In particular, both the CPIB short form and the VHI-10 ask specific questions about voice and communication in background noise. However, as a whole, results from this study suggest that an unfamiliar communication partner's ability to understand a speaker in quiet or noise may not be strongly predictive of how a person with a communication disorder perceives everyday voice or communication function.

One might question whether relationships reported in these studies were weak because the researchers did not control for ceiling effects among measures, such as intelligibility, that may be unevenly distributed across the range of the scale. To address these possible concerns, data from the present study also were analyzed using log transformed scores (Keene, 1995). The transformation stretches out the upper and lower ends of the scale, thereby allowing for valid comparison of scores across the entire range. Results of the analysis did not reveal any differences related to log transformation: Relationships remained unchanged in strength. Results from this study thus did not appear to be confounded by ceiling effects, lending support to their validity.

An interesting pattern was noted in examining relationships between intelligibility in noise and self-reported measures: Scores from 12 speakers who were more than 65% intelligible in noise strongly and significantly predicted self-reported outcomes ($r = -.89$ for the VHI-10; $r = .87$ for the CPIB short form; see Figure 4). These relationships greatly contrasted with those from 11 speakers who were 65% or less intelligible in noise (see Figure 3). One possible reason for the increased strength of relationships in the subgroup of 12 speakers who were more than 65% intelligible in noise may relate to the reliability of the measures. For example, Beukelman et al. (2011) have suggested that listeners may not expend as much effort listening to speakers who are severely unintelligible, because it might be too frustrating or unproductive to try to decode their messages. If that were the case in this study, we would assume that listeners would also show weaker reliability for the 11 speakers who were 65% or less intelligible in noise. The large variation in scores among this speaker group would then result in an unpredictable relationship with self-reported scores. However, an examination of interrater reliability values in this study did not reveal this to be true; no discernible pattern between intelligibility scores and reliability was observed.

Results from this study collectively highlight the multidimensional nature of self-reported outcomes. Among half the TEP speakers, a measure of speech impairment (intelligibility in noise) was a strong predictor of communication success. Yet for the other 11 speakers, a similar level of self-rated communication success was reported despite a significant drop in speech intelligibility in noise. A post hoc analysis showed that these two groups (those who were more than 65% intelligible in noise and those who were 65% or less intelligible in noise) were not otherwise differentiated by sex, age, education, radiation treatment, or time since laryngectomy. Other factors clearly affected the TEP speakers' ability to participate in conversations in everyday settings. Baylor, Yorkston, Bamer, Britton, and Amtmann (2010) investigated factors that affected communicative participation in a group of 498 community-dwelling adults with multiple sclerosis. Using regression modeling, their results revealed that communicative participation could be predicted with six out of 13 variables: self-reported fatigue, slurred speech, depression, problems with thinking, employment status, and social support. These variables accounted for about half the variance in the self-reported outcomes. The authors stated that one of the most "notable finding(s) is that communicative participation is not associated solely with communication disorder characteristics but that other variables, particularly fatigue, depression, and social support are also significantly related to participation" (p. 149).

Similar factors likely play a role in communicative participation in TEP speakers. From the present study, it appears that these dimensions may play an especially significant role in the communicative participation of speakers with poorer speech intelligibility in noise. Perhaps speakers with lower intelligibility performance are more reliant on other coping strategies to manage their communication success in adverse environments. Investigating how factors that go beyond speech and voice impairments—such as social support, depression, coping, and fatigue—affect self-reported outcomes after total laryngectomy should therefore be a focus of future study.

The sample of TEP speakers who participated in this study must be considered when interpreting the results. Demographics from this group were consistent with those reported in other research studies (Eadie et al., 2013; Vilaseca, Chen, & Backscheider, 2006). However, some investigators have raised questions about how representative research groups are when compared to clinical populations (Blood, Luther, & Stemple, 1992). Most subjects who participated were older White men; the majority had received radiation (88%), lived with family (79%), and reported some college education (75%). They were all individuals who were at least 1 year postlaryngectomy, and many were recruited through support groups. All of these factors could have affected communication and voice outcomes. For example, those who are involved in support groups or have more education or family support may be coping better and/or may be in better health than those who do not choose to participate in research (Blood et al., 1992). Consequently, scores may be elevated among this type of recruited sample, and results should be interpreted with these potential biases in mind.

Results from the present study suggest that noise is an important factor to consider when assessing speech intelligibility in individuals with communication disorders, such as those who use TEP speech. Similar to those with hearing loss, who may be more affected by noise than those with normal hearing, most TEP speakers are differentially affected by noise in comparison to typical speakers. This is particularly relevant when designing future research studies, documenting treatment outcomes, and counseling TEP speakers. However, noise is only one variable among many that may affect a person's voice-related QOL and communicative participation (Baylor et al., 2011, 2013; Op de Coul et al., 2005).

The multidimensional nature of self-reported outcomes was apparent among TEP speakers in this study who were less than 65% intelligible in noise but still reported good functional outcomes. These individuals clearly had adopted coping strategies for dealing with adverse environmental conditions. Other variables that may affect outcomes include supplemental communication strategies (e.g., nonverbal cues), familiarity of the communication partner and conversational topic, level of social support, daily communication demands, and time postlaryngectomy (Baylor et al., 2011). How these factors interact with outcomes needs further study. In the end, however, these results continue to highlight the importance of complementary and comprehensive assessment tools. Measures of speech intelligibility and self-reported outcomes are both necessary for capturing communication-related difficulties among populations with communication disorders, including those who have undergone total laryngectomy (Eadie, 2007).

Acknowledgments

This work is based on a thesis by the second author, submitted in partial fulfillment of the requirements for the Master of Science degree in medical speech-language pathology from the Department of Speech and Hearing Sciences, University of Washington. We would like to thank Kristie Spencer and Kathryn Yorkston for their valuable insight and feedback. We also would like to thank many members of the International Association of Laryngectomees for their participation in this study. Portions of this research were supported by National Cancer Institute Grant R01CA177635, awarded to the first author.

References

Azevedo, E. H. M., Montoni, N., Filho, J. G., Kowalski, L. P., & Carrara-de Angelis, E. (2012). Vocal handicap and quality of life after treatment of advanced squamous carcinoma of the larynx and/or hypopharynx. *Journal of Voice, 26,* e63–e71.

Baylor, C., Burns, M., Eadie, T., Britton, D., & Yorkston, K. (2011). A qualitative study of interference with communicative participation across communication disorders in adults. *American Journal of Speech-Language Pathology, 20,* 269–287.

Baylor, C., Yorkston, K., Bamer, A., Britton, D., & Amtmann, D. (2010). Variables associated with communicative participation in people with multiple sclerosis: A regression analysis. *American Journal of Speech-Language Pathology, 19,* 143–153.

Baylor, C., Yorkston, K., Eadie, T., Kim, J., Chung, H., & Amtmann, D. (2013). The Communicative Participation Item Bank (CPIB): Item bank calibration and development of a disorder-generic short form. *Journal of Speech, Language, and Hearing Research, 56,* 1190–1208. doi:10.1044/1092-4388(2012/12-0140)

Baylor, C. R., Yorkston, K. M., Eadie, T. L., Miller, R. M., & Amtmann, D. (2009). Developing the Communicative Participation Item Bank: Rasch analysis results from a spasmodic dysphonia sample. *Journal of Speech, Language, and Hearing Research, 52,* 1302–1320.

Bennett, S., & Weinberg, B. (1973). Acceptability ratings of normal, esophageal, and artificial larynx speech. *Journal of Speech and Hearing Research, 16,* 608–615.

Beukelman, D. R., Childes, J., Carrell, T., Funk, T., Ball, L. J., & Pattee, G. L. (2011). Perceived attention allocation of listeners who transcribe the speech of speakers with amyotrophic lateral sclerosis. *Speech Communication, 53,* 801–806.

Beverly, D., Cannito, M. P., Chorna, L., Wolf, T., Suiter, D. M., & Bene, E. R. (2010). Influence of stimulus sentence characteristics on speech intelligibility scores in hypokinetic dysarthria. *Journal of Medical Speech-Language Pathology, 18*(4), 9–13.

Blood, G. W., Luther, A. R., & Stemple, J. C. (1992). Coping and adjustment in alaryngeal speakers. *American Journal of Speech-Language Pathology, 1*(1), 63–69.

Bunton, K. (2006). Fundamental frequency as a perceptual cue for vowel identification in speakers with Parkinson's disease. *Folia Phoniatrica et Logopaedica, 58,* 323–339.

Campbell, B. H., Marbella, A., & Layde, P. M. (2000). Candidate's thesis: Quality of life and recurrence concern in survivors of head and neck cancer. *The Laryngoscope, 110,* 895–906.

Chen, S.-C., Yu, P.-J., Hong, M.-Y., Chen, M.-H., Chu, P.-Y., Chen, Y.-J., . . . Lai, Y.-H. (2015). Communication dysfunction, body image, and symptom severity in postoperative head and neck cancer patients: Factors associated with the amount of speaking after treatment. *Supportive Care in Cancer, 23,* 2375–2382. doi:10.1007/s00520-014-2587-3

D'Alatri, L., Bussu, F., Scarano, E., Paludetti, G., & Marchese, M. R. (2012). Objective and subjective assessment of tracheoesophageal prosthesis voice outcome. *Journal of Voice, 26,* 607–613. doi:10.1016/j.jvoice.2011.08.013

Donovan, N. J., Kendall, D. L., Young, M. E., & Rosenbek, J. C. (2008). The Communicative Effectiveness Survey: Preliminary evidence of construct validity. *American Journal of Speech-Language Pathology, 17,* 335–347.

Doyle, P. C., & Keith, R. L. (Eds.). (2005). *Contemporary considerations in the treatment and rehabilitation of head and neck cancer: Voice, speech, and swallowing*. Austin, TX: Pro-Ed.

Doyle, P. C., Swift, E. R., & Haaf, R. G. (1989). Effects of listener sophistication on judgments of tracheoesophageal talker intelligibility. *Journal of Communication Disorders, 22,* 105–113.

Eadie, T. L. (2007). Application of the ICF in communication after total laryngectomy. *Seminars in Speech and Language, 28,* 291–300. doi:10.1055/s-2007-986526

Eadie, T. L., & Bowker, B. C. (2012). Coping and quality of life after total laryngectomy. *Otolaryngology—Head & Neck Surgery, 146,* 959–965.

Eadie, T. L., Day, A. M. B., Sawin, D. E., Lamvik, K., & Doyle, P. C. (2013). Auditory-perceptual speech outcomes and quality of life after total laryngectomy. *Otolaryngology—Head & Neck Surgery, 148,* 82–88.

Eadie, T. L., & Doyle, P.C. (2004). Auditory-perceptual scaling and quality of life in tracheoesophageal speakers. *The Laryngoscope, 114,* 753–759.

Eadie, T. L., Lamvik, K., Baylor, C. R., Yorkston, K. M., Kim, J., & Amtmann, D. (2014). Communicative participation and quality of life in head and neck cancer. *Annals of Otology, Rhinology & Laryngology, 123,* 257–264.

Eadie, T. L., Yorkston, K. M., Klasner, E. R., Dudgeon, B. J., Deitz, J., Baylor, C. R., … Amtmann, D. (2006). Measuring communicative participation: A review of self-report instruments in speech-language pathology. *American Journal of Speech-Language Pathology, 15,* 307–320.

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of PRESTO: A new high-variability

sentence recognition test. *Journal of the American Academy of Audioliogy, 24*(1), 26–36. doi:10.3766/jaaa.24.1.4
Evans, E., Carding, P., & Drinnan, M. (2009). The Voice Handicap Index with post-laryngectomy male voices. *International Journal of Language & Communication Disorders, 44,* 575–586.
Hillman, R. E., Walsh, M. J., & Heaton, J. T. (2005). Laryngectomy speech rehabilitation: A review of outcomes. In P. C. Doyle & R. L. Keith (Eds.), *Contemporary considerations in the treatment and rehabilitation of head and neck cancer: Voice, speech, and swallowing* (pp. 75–90). Austin, TX: Pro-Ed.
Holley, S. C., Lerman, J., & Randolph, K. (1983). A comparison of the intelligibility of esophageal, electrolaryngeal, and normal speech in quiet and in noise. *Journal of Communication Disorders, 16,* 143–155.
Hustad, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language, and Hearing Research, 51,* 562–573.
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,* 198–208.
Jacobson, B. H., Johnson, A., Grywalsky, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The Voice Handicap Index (VHI): Development and validation. *American Journal of Speech-Language Pathology, 6,* 66–70.
Kazi, R., De Cordova, J., Singh, A., Venkitaraman, R., Nutting, C. M., Clarke, P., . . . Harrington, K. J. (2007). Voice-related quality of life in laryngectomees: Assessment using the VHI and V-RQOL symptom scales. *Journal of Voice, 21,* 728–734.
Kazi, R., Sayed, S. I., & Dwivedi, R. C. (2010). Post laryngectomy speech and voice rehabilitation: Past, present and future. *ANZ Journal of Surgery, 80,* 770–771.
Keene, O. N. (1995). The log transformation is special. *Statistics in Medicine, 14,* 811–819.
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,* 482–499.
Law, I. K.-Y., Ma, E. P.-M., & Yiu, E. M.-L. (2009). Speech intelligibility, acceptability, and communication-related quality of life in Chinese alaryngeal speakers. *Archives of Otolaryngology—Head & Neck Surgery, 135,* 704–711.
Lundström, E., Hammarberg, B., & Munck-Wikland, E. (2009). Voice handicap and health-related quality of life in laryngectomees: Assessment with the use of the VHI and EORTC questionnaires. *Folia Phoniatrica et Logopaedica, 61*(2), 83–92.
McCauliffe, M. J., Carpenter, S., & Moran, C. (2010). Speech intelligibility and perceptions of communication effectiveness by speakers with dysarthria following traumatic brain injury and their communication partners. *Brain Injury, 24*(10), 1405–1415. doi:10.3109/02699052.2010.511590
McAuliffe, M. J., Schaefer, M., O'Beirne, G. A., & LaPointe, L. L. (2009, November). *Effect of noise upon the perception of speech intelligibility in dysarthria.* Poster session presented at the Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA.
McColl, D. A. (2006). Intelligibility of tracheoesophageal speech in noise. *Journal of Voice, 20,* 605–615. doi:10.1016/j.jvoice.2005.09.005
McColl, D., Fucci, D., Petrosino, L., Martin, D. E., & McCaffrey, P. (1998). Listener ratings of the intelligibility of tracheoesophageal speech in noise. *Journal of Communication Disorders, 31,* 279–289.
Meyer, T. K., Kuhn, J. C., Campbell, B. H., Marbella, A. M., Myers, K. B., & Layde, P. M. (2004). Speech intelligibility and quality of life in head and neck cancer survivors. *The Laryngoscope, 114,* 1977–1981.
Miralles, J. L., & Cervera, T. (1995). Voice intelligibility in patients who have undergone laryngectomies. *Journal of Speech and Hearing Research, 38,* 564–571.
Nagle, K. F., & Eadie, T. L. (2012). Listener effort for highly intelligible tracheoesophageal speech. *Journal of Communication Disorders, 45,* 235–245.
Nagle, K. F., Eadie, T. L., Wright, D. R., & Sumida, Y. A. (2012). Effect of fundamental frequency on judgments of electrolaryngeal speech. *American Journal of Speech-Language Pathology, 21,* 154–166.
Op de Coul, B. M. R., Ackerstaff, A. H., van As, C. J., van den Hoogen, F. J. A., Meeuwis, C. A., Manni, J. J., & Hilgers, F. J. M. (2005). Quality of life assessment in laryngectomized individuals: Do we need additions to standard questionnaires in specific clinical research projects? *Clinical Otolaryngology, 30,* 169–175.
Oridate, N., Homma, A., Suzuki, S., Nakamaru, Y., Suzuki, F., Hatakeyama, H., . . . Fukuda, S. (2009). Voice-related quality of life after treatment of laryngeal cancer. *Archives of Otolaryngology—Head & Neck Surgery, 135,* 363–368.
Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation, 9,* 6. Retrieved from http://pareonline.net/getvn.asp?v=9&n=6
Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
Rosen, C. A., Lee, A. S., Osborne, J., Zullo, T., & Murry, T. (2004). Development and validation of the Voice Handicap Index-10. *The Laryngoscope, 114,* 1549–1556.
Searl, J. (2007). Bilabial contact pressure and oral air pressure during tracheoesophageal speech. *Annals of Otology, Rhinology & Laryngology, 116,* 304–311.
Shrivastav, R., Sapienza, C. M., & Nandur, V. (2005). Application of psychometric theory to the measurement of voice quality using rating scales. *Journal of Speech, Language, and Hearing Research, 48,* 323–335.
Singer, M. I., & Blom, E. D. (1980). An endoscopic technique for restoration of voice after laryngectomy. *Annals of Otology, Rhinology, & Laryngology, 89,* 529–533.
Sperry, J. L., Wiley, T. L., & Chial, M. R. (1997). Word recognition performance in various background competitors. *Journal of the American Academy of Audiology, 8,* 71–80.
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,* 1208–1219.
Tardy-Mitzell, S., Andrews, M. L., & Bowman, S. A. (1985). Acceptability and intelligibility of tracheoesophageal speech. *Archives of Otolaryngology, 111,* 213–215.
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,* 779–792. doi:10.1044/2014_JSLHR-S-12-0372
van As, C. J., Koopmans-van Beinum, F. J., Pols, L. C. W., & Hilgers, F. J. M. (2003). Perceptual evaluation of tracheoesophageal speech by naive and experienced judges through the use of semantic differential scales. *Journal of Speech, Language, and Hearing Research, 46,* 947–959.

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America, 121,* 519–526.

Vilaseca, I., Chen, A. Y., & Backscheider, A. G. (2006). Long-term quality of life after total laryngectomy. *Head & Neck, 28,* 313–320.

Yorkston, K., & Beukelman, D. (1983). The influence of judge familiarization with the speaker on dysarthric speech intelligibility. In W. R. Berry (Ed.), *Clinical dysarthria* (pp. 155–163). San Diego, CA: College-Hill Press.

Yorkston, K., Beukelman, D., & Tice, R. (1996). *Sentence Intelligibility Test (Version 1.0)*. Lincoln, NE: Tice Technology Services.