

A Comparison Between Self-Rated and Listener-Rated Outcomes in Tracheoesophageal Speech

Jordan R. Johnson

A thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2012

Committee:

Kathryn Yorkston

Martin Nevdahl

Tanya Eadie

Program Authorized to Offer Degree:

Speech & Hearing Sciences

TABLE OF CONTENTS

List of Figures	ii
List of Tables	iii
Acknowledgements	iv
Introduction	1
Outcome Measures	3
Quality of Life after Total Laryngectomy	3
Intelligibility and Acceptability	5
Vocal Effort Perceptions	7
Effort in TE Speech	9
Experimental Questions	9
Methods	11
Overview and Design	11
Data Collection – TE Speakers	12
Preparation of Speech Samples	14
Reliability of Listener Ratings	14
Data Analysis	15
Results	16
Speaker Demographics	16
Difference Between Speakers’ and Listeners’ Auditory-Perceptual Judgments	16
Relationships Between Speakers’ and Listeners’ Auditory-Perceptual Judgments ...	18
Relationship Between Speakers’ Judgments of Acceptability and Vocal Effort	20
Relationship Between Listeners’ Judgments of Acceptability and Vocal Effort	21
Relationship Between Auditory-Perceptual Dimensions with Voice-Related QOL ..	22
Discussion	27
Speaker Demographics	27
Difference Between Speakers’ and Listeners’ Auditory-Perceptual Judgments	29
Relationships Between Speakers’ and Listeners’ Auditory-Perceptual Judgments ...	31
Relationship Between Judgments of Acceptability and Vocal Effort	33
Relationship Between Auditory-Perceptual Dimensions with Voice-Related QOL ..	35
Future Directions and Clinical Implications	38
References	42

LIST OF FIGURES

Figure Number	Page
1. Correlation Between Listener and Speaker Rated Acceptability.....	18
2. Correlation Between Listener and Speaker Rated Vocal Effort.....	19
3. Correlation Between Speaker Rated Vocal Effort and Acceptability.....	20
4. Correlation Between Listener Rated Vocal Effort and Acceptability.....	21
5. Correlation Between Speaker Rated Acceptability and Voice-Related QOL.....	22
6. Correlation Between Listener Rated Acceptability and Voice-Related QOL.....	23
7. Correlation Between Speaker Rated Vocal Effort and Voice-Related QOL.....	24
8. Correlation Between Listener Rated Vocal Effort and Voice-Related QOL.....	25

LIST OF TABLES

Table Number	Page
1. Demographic Characteristics of TE Speakers	16
2. Summary of Correlations Between Speakers' and Listeners' Judgments of Acceptability, Vocal Effort, and VHI-10 Scores	26

ACKNOWLEDGEMENTS

I would like to sincerely thank Dr. Tanya Eadie for her guidance and support throughout this endeavor. It has been a wonderful experience to learn from and work with such a knowledgeable and kind supervisor.

I would also like to thank my committee members, Dr. Kathryn Yorkston and Martin Nevdahl, for sharing their considerable knowledge and time in this project.

A special thanks to the members of the Vocal Function Lab for their continued encouragement of me while completing this project, but specifically Devon Sawin for her help with scheduling and data collection.

Finally, thank you to my wonderful parents and friends. Without their love and support, none of this would have been possible.

INTRODUCTION

Total laryngectomy (TL) plus radiation therapy (RT) remains a primary treatment for advanced laryngeal carcinoma relative to local control and survival (Hartl et al., 2011). In TL, the larynx is surgically removed and the trachea is sutured at the base of the neck, creating a tracheostoma that is a direct pathway to the lungs (Eadie, 2003). The presence of a tracheostoma results in loss of the ability to inhale and exhale through the nasal and oral cavities (Eadie, 2003). The altered anatomy also necessitates the use of a new, alaryngeal voice mechanism for verbal communication.

Three common alaryngeal speech methods include the use of an electrolarynx (EL) as an external sound source, as well as two methods that are produced “intrinsically”: esophageal speech (ES), and tracheoesophageal (TE) speech (Kazi, Sayed, & Dwivedi, 2010). EL speech is produced by placing an artificial laryngeal device against the neck, cheek, or oral cavity and articulating the sound that has been transmitted through the tissues. In contrast, both ES and TE speech are produced using sound sources within the body. To use ES, air must be injected below the pharyngoesophageal (PE) segment and then released back up, causing the upper esophageal sphincter tissue to vibrate and become the sound source. The individual then articulates (as in the normal, healthy system) to produce speech. Finally, TE speech is achieved by a surgical procedure that places a puncture between the posterior tracheal wall into the anterior esophageal wall. Later, a one-way valve voice prosthesis is placed in the puncture, allowing air to pass through from the trachea into the esophagus when the tracheostoma is occluded, causing the upper esophageal sphincter to vibrate. Similar to ES, the sound emitted from the upper esophageal sphincter is modified by the articulators to produce speech. However, unlike ES, the

respiratory support for the sound source in TE speech is provided by the lungs, which offers advantages in terms of air flow and pressures needed for longer, fluent speech (Andrews, 1999).

In the past decade, TE speech has been considered the “gold standard” in post-laryngectomy voice rehabilitation (Kazi et al., 2010). Advantages of using a voice prosthesis include: nearly immediate voice production following surgery, the possibility of sustained speech (i.e., beyond the limits of ES), the ability to reverse the surgery and use another speech method, and complication rates that are relatively low (Kazi et al., 2010). In addition, TE speech has been shown to be more similar to laryngeal speech than ES on parameters such as fundamental frequency, speaking rate, and maximum phonation time (Kazi et al., 2010). On intelligibility measures and the auditory-perceptual measure of “pleasantness,” TE speakers are typically ranked the highest among all alaryngeal speech methods (Eadie & Doyle, 2004).

Although TE speech is the preferred alaryngeal speech method for listeners, a study by Finizia et al. (1998) found that TE speech is judged as having worse voice quality and is less acceptable than both normal speakers and speakers treated with radical RT. In addition, TE speakers often report more “effortful” speech, which may be linked to the increased driving pressures required for alaryngeal voice production (Bohnenkamp, Forrest, Klaben, & Stager, 2011). For individuals who have undergone TL, these results may directly influence social interactions and a communication partner’s reactions. In addition, effortful speech might also affect the speaker’s satisfaction with voice and functional communication, as well as his or her adjustments to daily living (Eadie & Doyle, 2004). The overall purpose of this study is to determine how self-reported effort in speech production, as well as other listeners’ judgments of effort may or may not differ from traditional perceptual outcomes, such as speech acceptability. A second broad purpose is to determine whether these auditory-perceptual measures relate to

functional outcomes such as voice-related quality of life in TE speakers. Before these relationships can be understood, a review of typical outcomes after TL needs to be summarized.

Outcome Measures

Although the primary outcome of interest in laryngeal cancer is curative treatment (Eadie & Doyle, 2004), different outcome measures reveal information about other important dimensions related to TE speech. For example, patient-reported outcomes have become a focus in the past decade, with quality of life (QOL) measures typically being used in the head and neck cancer population. Understanding how other factors (e.g., perceived voice quality or speech acceptability, speech effort, stage of disease, coping) predict QOL is important for directing rehabilitation. A summary of QOL outcomes will be provided before these relationships may be understood.

Quality of Life after Total Laryngectomy

QOL measures are used to determine a person's functional success with speech, as well as other aspects of daily functioning. According to the World Health Organization (WHO, 1997), QOL is defined as:

“individuals' perceptions of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept affected in a complex way by the person's physical health, psychological state, level of independence, social relationships, personal beliefs and their relationship to salient features of their environment” (p. 1).

This definition reflects the multifaceted nature of QOL, and the wide variety of factors that influence it. By standard, QOL measures may be disease-specific (e.g., relating to all symptoms of head and neck cancer) or symptom-related (e.g. voice-related QOL: measuring voice-related

symptoms of various voice disorders). Because disease-specific QOL measures may only ask one or two specific questions regarding certain functions, a voice- or speech-related QOL measure is better suited to capture QOL as it relates to the voice symptoms experienced by a person using alaryngeal speech.

One voice-related QOL measure that is commonly used in clinical practice is the *Voice Handicap Index* (VHI; Jacobson et al., 1997). The VHI assesses the impact of a voice disorder on an individual by means of a self-administered questionnaire. Individuals rate their perceived voice handicap across three dimensions: functional, emotional, and physical. This measure was validated using participants exhibiting a variety of voice disorders, including those having undergone laryngectomy. Therefore, it serves as a viable measurement of functional outcomes in this population (Jacobson et al., 1997).

In a study by Evans, Carding, and Drinnan (2009), the VHI was administered to a group of male patients (n=53) who had undergone TL. The mean score of the men using surgical voice rehabilitation (i.e., TE speech) (n=26) was 44.7, which indicated a moderate voice handicap. Similarly, Azevedo, Montoni, Fliho, Kowalski, and Angelis (2011) showed that patients using TE speech after TL (n=17) reported scores that reflected a moderate voice handicap (mean=31). More recent studies have used the short form of the VHI, the VHI-10, to assess similar outcome domains (Rosen et al., 2004). For example, Oridate et al. (2009) found laryngectomees (n=27) to have a mean VHI-10¹ score of 11.26 (SD=7.17), which also represents a moderate handicap.

The results of all of these studies suggest that voice-related QOL measures, such as the VHI, reveal the negative impact of voice-related symptoms on QOL in people who have undergone TL and specifically those using TE speech. One hallmark of these studies is the wide range of scores reported by the participants. For example in the study by Evans et al. (2009), the

¹ Scores on the VHI-10 range from 0 (normal) to 40 (most severely limited) voice handicap

VHI scores ranged from 4 to 106. These results show that while the average score was consistent with a moderate voice handicap, there were also individuals who were within the range of “normal”, as well as those who were severely impacted by their TE voice. If one could determine what predicts a “successful” outcome (and perhaps more importantly, a “poor” outcome), then clinicians could plan more appropriately during rehabilitation. While a variety of factors affect voice-related QOL, two potential predictors include a) other listeners’ perceptions of the speech sample (i.e., reflective of the communication partner’s reactions), or b) the speaker’s self-rated perception of quality, acceptability, or effort when speaking. These outcomes will be examined next.

Intelligibility and Acceptability

Two common measurements of alaryngeal speech with a long history of use in the clinical setting include ratings of intelligibility and acceptability by both clinicians and naïve listeners (Bennett & Weinberg, 1973; Law, Ma, & Yiu, 2009; Miralles & Cervera, 1995; Tardy-Mitzell, Andrews, & Bowman, 1985). Speech intelligibility has been defined as the percentage of speech items that are correctly identified by the listener (Hillman, Walsh, & Heaton, 2005), while speech acceptability is defined as the degree to which a speaker is fluent, uses an expected rate and intonation, and has a pleasant quality to his or her voice (Tardy-Mitzell, Andrews, & Bowman, 1985). The acceptability of one’s speech has been shown to be highly interrelated with severity, naturalness, and pleasantness of the speaker’s voice (Eadie & Doyle, 2004). The extent to which a listener deems one’s speech to be acceptable may have implications about the social penalty one could experience in everyday life, with some demonstrated relationship with QOL measures (Eadie & Doyle, 2004). Similarly, lower speech intelligibility may influence communication abilities of an alaryngeal speaker and have a negative impact on conversation

(Law, Ma, & Yiu, 2009). For these reasons, measures of acceptability and intelligibility are helpful in gauging the success of one's alaryngeal voice rehabilitation.

TE speech has been shown to be less acceptable than laryngeal speech and female TE speakers receive significantly lower acceptability ratings than men (Eadie & Doyle, 2004). This finding, that women are rated to have less pleasant, less acceptable, and more severe voices than men, may suggest a higher social penalty for women who are TE speakers. Reduced speech acceptability, in turn, may relate to a differential penalty for women in terms of their QOL outcomes (Eadie & Doyle, 2004). The intelligibility of TE speakers has been shown to be equal to ES and better than EL speech, implying that TE speech is one of the best options for voice rehabilitation for people who have had a laryngectomy (Tardy-Mitzell, Andrews, & Bowman, 1985). In the previous study, naïve listeners understood the TE speakers on average 93% of the time (Tardy-Mitzell, Andrews, & Bowman, 1985). This result indicates that TE speakers are considered to be highly intelligible to a naïve listener, the type of listener they would be most likely to encounter on a day-to-day basis (Tardy-Mitzell, Andrews, & Bowman, 1985). A study by Law, Ma, and Yiu also revealed that age could be a factor in judgments of alaryngeal speech (2009). For example, results showed that while younger judges had an easier time understanding alaryngeal speech, they also found alaryngeal speech less acceptable than older judges. Thus, factors such as the listener's age as well as the speaker's sex must be considered when designing perceptual studies.

Both listener-rated acceptability and intelligibility are important dimensions to consider when measuring post-laryngectomy outcomes. Both factors may influence the way a communication partner interacts with a TE speaker, resulting in positive or negative changes to the speaker's QOL. Finizia et al. (1998) used a comparison of listener and speaker ratings to

examine voice quality, acceptability, and intelligibility of TE speakers. Results of this study showed that speakers judged their own voices to be significantly better than the listeners in the domains of voice quality and acceptability. These findings suggest that TE speakers appear to be more accepting of their new voice than the people who are listening to them. It also might suggest that self-rated perceptions of voice might more strongly predict voice-related QOL than perceptions by other listeners. Literature supporting this hypothesis is examined next.

Vocal Effort Perceptions

Although voice acceptability might be a good indicator of the social acceptance of an alaryngeal speaker's voice, one recent study indicates that an unfamiliar listener's perceptions of either acceptability or intelligibility might not be a strong predictor of the patient's own report of his or her voice-related QOL (Eadie et al., 2011). In that study, 20 speakers who had undergone TL and who used various types of alaryngeal speech were audio recorded and their speech samples were presented to 21 inexperienced listeners. Listeners transcribed sentences to derive intelligibility scores, and they also judged speech acceptability of the speech samples. Eadie et al. (2011) found weak relationships between listener perceptions of acceptability and intelligibility with self-reported voice related QOL, as measured by the VHI-10 ($r = -.295$, acceptability vs VHI-10; $r = .025$, intelligibility vs. VHI-10). These results indicate that listeners' judgments of acceptability and intelligibility are not necessarily predictive of the speakers' voice-related QOL and necessitate the exploration of other auditory-perceptual characteristics that may be more predictive.

One voice-related symptom that has been studied in other voice populations, but not TL, is speaker and listener perception of speaking effort. Studies have examined both naïve and experienced listener perceptions of vocal effort (VE) in individuals with neurological, organic,

and functional types of dysphonia (Eadie et al, 2007; Eadie et al, 2010). VE has been defined as “the perceived effort in producing voice” (Eadie et al, 2007). Eadie et al. (2007) asked individuals with adductor spasmodic dysphonia (ADSD) to complete the VHI and to rate their perceived VE immediately following their recording of a standardized reading passage, the Rainbow Passage (Fairbanks, 1960). Individuals rated their VE on a 100-mm visual analog scale. The speakers’ ratings were then compared to similar judgments by both naïve and experienced listeners. The results of this study showed that naïve and experienced listeners’ judgments of speaker effort were strongly correlated ($r= 0.917$). However, naïve and experienced listeners’ judgments were only moderately correlated with speakers’ effort judgments. This suggests that speakers use a different strategy for perceiving VE than listeners. Interestingly, this study also showed that the speakers’ VE ratings were moderately correlated to their VHI scores ($r = .608$), and were stronger predictors of voice-related QOL than voice quality. These results might suggest that vocal effort is a factor that should be examined when trying to predict voice-related QOL in other populations.

A similar study examined speaker and other listeners’ perceptions of VE in individuals with various types of laryngeal-based dysphonia (Eadie et al., 2010). The results of this study supported the earlier findings that inexperienced and experienced listeners’ perceptions of VE were strongly correlated, but when compared with speaker perceptions, the relationship was only moderate in strength. Once again, self-rated VE as well as other listeners’ judgments of VE were more strongly related to VHI scores than voice quality severity, with the self-rated scores being most predictive of voice-related QOL. These results indicate that the experience of the listener does not impact his or her ability to rate VE, but that speakers are possibly using different strategies in their evaluation of perceived VE than the listeners. Both studies reveal that a

speaker's VE is correlated with voice-related QOL, indicating that VE is a salient dimension that should be taken into consideration during evaluations of individuals with dysphonia (Eadie et al., 2007; Eadie et al., 2010).

Effort in TE Speech

Though there is a paucity of data related to TE speakers' perception of VE, TE speakers commonly complain of the effort involved in producing speech. This complaint is validated by research that suggests that phonation created by the PE segment requires greater pressures to achieve vibration than phonation using the normal laryngeal system (Bohnenkamp et al., 2011). Thus, speaker effort might be different in an alaryngeal population than individuals with laryngeal-based dysphonia because TE speakers are using a different system of speech production. Aerodynamic studies of TE speakers have shown that the pressure required to vibrate the PE segment is nearly double that of a typical larynx, and as individuals are asked to increase their speaking effort, the amount of pressure also increases (Kotby, Hegazi, Kamal, Gamal el Dien, & Nassar, 2009; Moon & Weinberg, 1987). The results of these aerodynamic studies support the contention that TE speech is more effortful than laryngeal speech. How this increase in effort might relate to the speaker's and other listeners' perceptions of that effort or a speaker's satisfaction with his or her voice is yet to be investigated.

Experimental Questions

In summary, both self-reported and other listener-rated outcomes are important to consider after total laryngectomy. Comparisons between speakers' and other listeners' perceptions of dimensions such as speech acceptability and perceived VE have revealed differences in strategies used by individuals with various types of voice disorders. However, at this time there is limited research investigating the relationship between these measures, and

whether they can be differentiated in their use by speakers or other listeners. Results from previous studies also have shown that although speech acceptability is reduced in TE speech (compared to laryngeal speech), it may not relate strongly to a TE speaker's voice-related QOL (Eadie et al., 2011). Thus, one must consider other factors that could predict voice-related QOL, such as perceived VE (Eadie et al., 2007; Eadie et al., 2010). VE is important to investigate because the amount of effort (and air pressure) required to produce TE speech is much greater than the normal laryngeal system (Bohnenkamp et al., 2011), and because TE speakers frequently report increased effort in speaking. How this dimension predicts voice-related QOL must therefore be examined in TE speakers. Consequently, this study was designed to answer the following experimental questions:

(a) Are there differences in judgments of speech acceptability or perceived speaking effort when these dimensions are rated by TE speakers and unfamiliar listeners? (b) What are the relationships among judgments of speech acceptability and perceived speaking effort when these dimensions are rated by TE speakers and unfamiliar listeners? and (c) How do self-rated and listener-rated speech acceptability and vocal effort relate to voice-related QOL in TE speakers?

Experimental examination of these specific conditions will add to the existing literature regarding alaryngeal speech methods, help to inform clinical management, and suggest future research of people who have undergone TL.

METHODS

Overview and Design

This study included two groups of participants: (1) 16 individuals who had undergone TL and who use TE speech as their primary method of communication provided speech samples and completed a standardized voice-related QOL measure as well as speech acceptability and effort ratings, and (2) 20 naïve listeners who performed perceived speech effort and acceptability ratings of the TE speakers. All participants were native speakers of English and did not report any other significant health conditions that might affect their speech production (with the exception of TL for the first group of participants). The University of Washington Human Subjects Committee approved the procedures used in this study; all participants were paid for their participation.

This study used a combined comparative and correlational design to investigate the following in TE speech: a) the possible differences between speakers' and listeners' judgments of acceptability/vocal effort in TE speech; b) the relationship between the speakers' perceived acceptability/vocal effort and the naïve listeners' perceived acceptability/vocal effort of the speaker; and, c) the relationship between acceptability/vocal effort from both the speaker and listener (predictor variables) with voice-related QOL (predicted variable).

Speakers: Sixteen individuals who had undergone a TL secondary to laryngeal cancer and who used TE speech as their primary method of communication served as speakers for the present study. Of the 16 speakers, 14 were male and 2 were female. The mean age of the speakers was 63 years (SD = 9.6). In order to allow for time to adjust to a new communication method, participants were required to be at least one-year status post TL. On average, participants were 6.6 years post TL surgery (SD = 4.6). Participants were recruited from support groups in Washington State, including the Seattle, Tri-Cities, and Spokane areas, as well as from

an international conference for individuals who have undergone TL, the International Association of Laryngectomees, held in Kansas City, MO. Exclusion criteria included individuals with speech impairments unrelated to laryngeal cancer, neurologic disorders, and any previously altered anatomy related to the upper aerodigestive tract.

Naïve Listeners: Twenty adult naïve listeners, 18 years or older, were recruited from the Seattle community, including the student population at the University of Washington. Of the 20 listeners, 15 were male and 5 were female. The mean age of the listeners was 24 years (SD = 6.01). A naïve listener was defined as an individual with no prior experience or strictly undergraduate coursework related to alaryngeal speech or TL. Listeners were required to pass a hearing screening (25 dB at octave frequencies of 250-8000 Hz) and be a native speaker of English.

Data Collection – TE Speakers

First, speakers completed a consent form and provided demographic information (e.g. age, gender, native language, date of diagnosis, and date of TL). Information related to speech and history of speech-language pathology services also was obtained.

VHI-10. Speakers completed a voice-related QOL measure, the VHI-10 (Rosen, Lee, Osborne, Zullo, & Murry, 2004). This measure was either finished the day of the recordings or accepted by mail up to two months post-recording. This criterion was chosen to control for changes in QOL that may occur after the speech samples were recorded. The VHI-10 was used as a subjective measure of a patient's overall perception of handicap due to his or her vocal function (VHI-10; Rosen et al., 2004). The VHI-10 is a shortened version of the original 30-item VHI, synthesizing three subscales measuring functional, physical, and emotional domains into one total scale (Rosen et al., 2004). The VHI-10 consists of 10 items measuring self-rated voice

limitations by asking patients how often they experience a condition described by a set of statements (e.g. “My voice makes it difficult for people to hear me”). These 10 statements are summed to derive a composite VHI-10 score that ranges from “0” (no voice handicap) to “40” (severe voice handicap). The VHI-10 has concurrent validity ($r = .90$) with the original 30-item VHI (Rosen et al., 2004), and has been used in the head and neck cancer literature to evaluate outcomes after treatment such as total laryngectomy (Evans et al., 2009; Lundstrom et al., 2009; Moerman et al., 2004).

Speech recordings and self-ratings of effort and acceptability. Voice samples were recorded in a sound-treated room onto a digital audiorecorder (Sony PCM-R500) using a headset microphone (AKG-C20) with a three-inch mouth-to-microphone distance at an angle of about 30 degrees above horizontal to minimize stoma noise (Bellandese, Lerman, & Gilbert, 2001). Speakers were asked to record the “Rainbow Passage” (Fairbanks, 1960), a standard reading passage used in measuring voice and alaryngeal voice outcomes (Bellandese, 2009; Bellandese et al., 2001).

Following the recording of the speech sample, speakers judged their perceived speaking effort using a 100 mm visual analog scale (VAS) with the endpoints labeled as “no effort” and “extremely effortful.” They were familiarized with the rating scale and provided a definition of vocal effort. For the purposes of this study, vocal effort (VE) was defined as “the perceived effort in producing voice” (Eadie et al., 2008). Similarly, speakers judged their perceived speech acceptability using a 100 mm VAS, with the endpoints labeled “very unacceptable” to “very acceptable.” To rate speech acceptability, speakers were asked to give careful consideration to the attributes of pitch, rate, and understandability (Bennett & Weinberg, 1973).

Preparation of Speech Samples

Speech samples were transferred to a computer and converted into WAV files using acoustic software (Sony Soundforge) at a sampling rate of 44,100 kHz. The second sentence of the Rainbow Passage, “The rainbow is a division of white light into many beautiful colors,” (Fairbanks, 1960) was edited using acoustic software (Sony Soundforge). The WAV files were entered into a custom-made software program (Ruby on Rails) that randomized the presentation order of the speech samples and recorded the listeners’ responses (speech acceptability or vocal effort) on a rating scale.

Listener procedure. Naïve listeners completed a consent form, demographic questionnaire (e.g. age, native language, and personal experience with alaryngeal speech and TL), and a hearing screening prior to listening to speech samples. Following completion of these items, listeners judged the effort they perceived the TE speaker to have used and the acceptability of each voice, using the same definitions as the speakers. After presentation of the second sentence of the Rainbow Passage (Fairbanks, 1960) over headphones (Samson RH600), listeners judged the TE speech for VE and acceptability using the same 100 mm VAS and instructions used by the speakers. Listeners were allowed to listen to each recorded sentence one time (N = 16 speech samples + 5 repeated for reliability = 21 speech samples for each dimension, VE and acceptability). Speech samples were presented in a random order and counterbalanced for rating dimension.

Reliability of Listener Ratings

To evaluate intrarater reliability of the listener’s speech acceptability and perceived VE ratings, 30 percent (n=5) of the speaker samples were repeated. To evaluate the relationship between the original and repeated samples for the ratings of each listener, Pearson correlation

coefficients were calculated. The mean reliability value (r) was 0.85 (SD = 0.18) for speech acceptability, and 0.85 (SD = 0.17) for perceived VE. These results are consistent with other perceptual studies (Eadie et al., 2008), and indicate adequate intrarater reliability for both perceptual dimensions.

Interrater reliability was calculated by using intraclass correlation coefficients (ICCs) to determine the consistency of speech acceptability/VE ratings between listeners (Shrout & Fleiss, 1979). The group mean average ICC for speech acceptability was 0.975. For VE, the group mean average ICC was 0.970. These results suggest that listeners were able to consistently judge both perceptual dimensions as a group.

Data Analysis

This study investigated possible differences and relationships between speech acceptability as well as perceived VE rated by both TE speakers and other listeners. To determine whether there were differences between the rater groups for each of the dimensions, 2 independent t -tests were performed. Specifically, group averages for speaker-rated and listener-rated measures of speech acceptability and perceived VE were compared. In order to examine the relationship between the speakers' and listeners' ratings of speech acceptability and VE, the average of all listener ratings for each speaker were calculated and this average was compared to the speaker's own ratings using multiple Pearson's correlation coefficients. To examine relationships between listener-rated and QOL outcomes, multiple correlations were performed. The predictor variables in this study were perceived VE and speech acceptability (as rated by speakers and naïve listeners), and the predicted variable was a measure of voice-related QOL (VHI-10). To measure the relationships between the variables, four Pearson's correlations were calculated using each group's average for speech acceptability/VE with the total VHI-10 scores.

RESULTS

Speaker Demographics

Demographic characteristics of the 16 TE speakers who participated in the present study are presented in Table 1. Eight participants (50%) reported receiving radiation before their TL surgery, while 6 (37.5%) reported receiving it afterwards; 2 (12.5%) did not receive radiation. The mean VHI-10 scores for all speakers, except for one who did not complete the paperwork, was 17.7 (SD=7.1) (moderate voice handicap). Scores ranged from 4 (i.e., essentially no voice handicap) to 30 (i.e., severe voice handicap, out of a maximum score of 40).

Table 1. Demographic Characteristics of TE Speakers (N=16), Including Participant Gender, Age (y), Years Post TL, Radiation History, and VHI-10 Totals

Gender	Age (y)	Years Post TL	Radiation History	VHI-10 Total
M	70	14	Before TL	23
F	57	4	Before TL	23
F	39	2	After TL	24
M	66	5	Before TL	10
M	72	6	Before TL	16
M	56	15	After TL	12
M	67	6	Not Reported	16
M	50	11	Before TL	21
M	72	2	Before TL	10
M	72	8	Before TL	4
M	60	1	After TL	22
M	66	12	After TL	30
M	70	Not Reported	Not Reported	Not Reported
M	68	6	Before TL	13
M	73	1	After TL	25
M	56	6	After TL	16

Difference Between Speakers' and Listeners' Auditory-Perceptual Judgments

Acceptability. An independent samples *t*-test comparing the mean scores of speaker and listener judgments of acceptability found a significant difference between the means of the two groups ($t(29) = -4.41, p < .01$). The mean of the speakers' judgments of acceptability was significantly higher ($m = 65.07, sd = 24.10$) than the mean of the listeners' judgments ($m = 30.24$,

$sd = 19.45$), indicating that speakers judged their speech as significantly more acceptable than the listeners did.

Vocal Effort. An independent samples t -test comparing the mean scores of speaker and listener judgments of vocal effort was performed to determine whether there were differences between the groups. Results showed a significant difference between the two groups ($t(29) = 4.762, p < .01$). Specifically, listeners' judgments of vocal effort were significantly higher ($m = 50.89, sd = 20.69$) than the speakers' judgments ($m = 19.93, sd = 14.80$). In other words, listeners as a group perceived the speakers as using significantly more effort to produce TE speech than the speakers' own perceptions of their effort.

Relationships Between Speakers' and Listeners' Auditory-Perceptual Judgments

Acceptability. A Pearson correlation coefficient was calculated to examine the relationship between speaker and listener judgments of acceptability. A weak positive correlation was found ($r(13) = .138, p = .625$), indicating a very limited linear relationship between the two variables (see figure 1). Thus, speaker and listener judgments of acceptability do not appear to relate to one another (or at least do so very weakly).

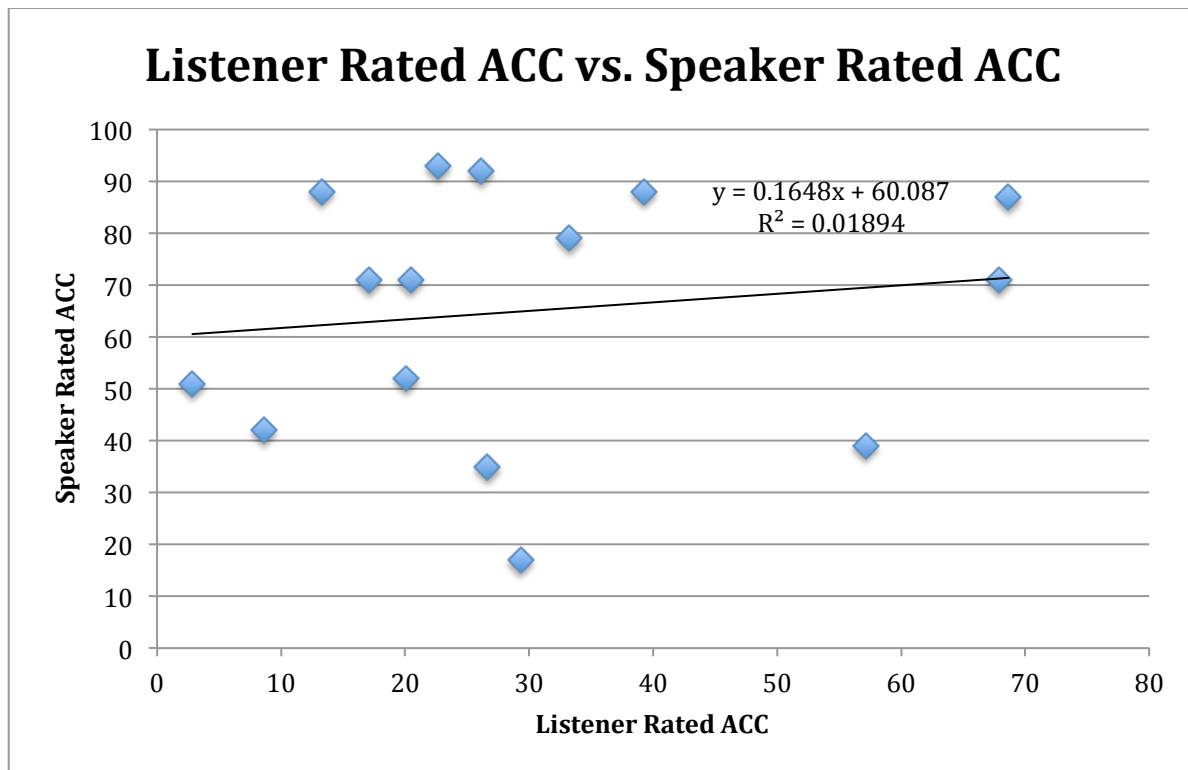


Figure 1: Correlation between Listener and Speaker Rated Acceptability (ACC)

Vocal Effort. A Pearson correlation coefficient was calculated to determine the relationship between speaker and listener judgments of vocal effort. A moderate positive correlation was found ($r(13) = .476, p = .073$), indicating a linear relationship between the two variables (see figure 2). Speaker and listener judgments of vocal effort appear to be moderately related to one another.

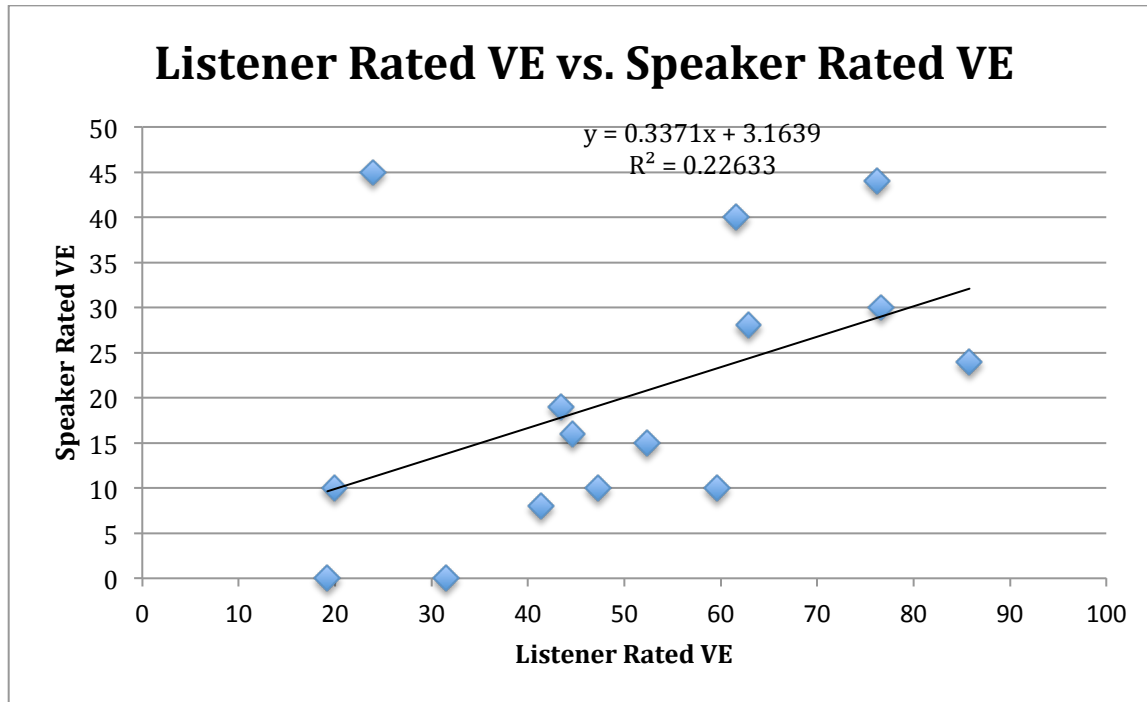


Figure 2: Correlation between Listener and Speaker Rated Vocal Effort (VE)

Relationship Between Speakers' Judgments of Acceptability and Vocal Effort

A Pearson correlation coefficient was calculated to determine the relationship between speaker judgments of acceptability and vocal effort. A strong, significant negative correlation was found, ($r(13) = -.860, p < .001$), indicating a linear relationship between the two variables (see figure 3). Speakers who rated their own acceptability as high tended to rate their perceived vocal effort as low, and vice versa.

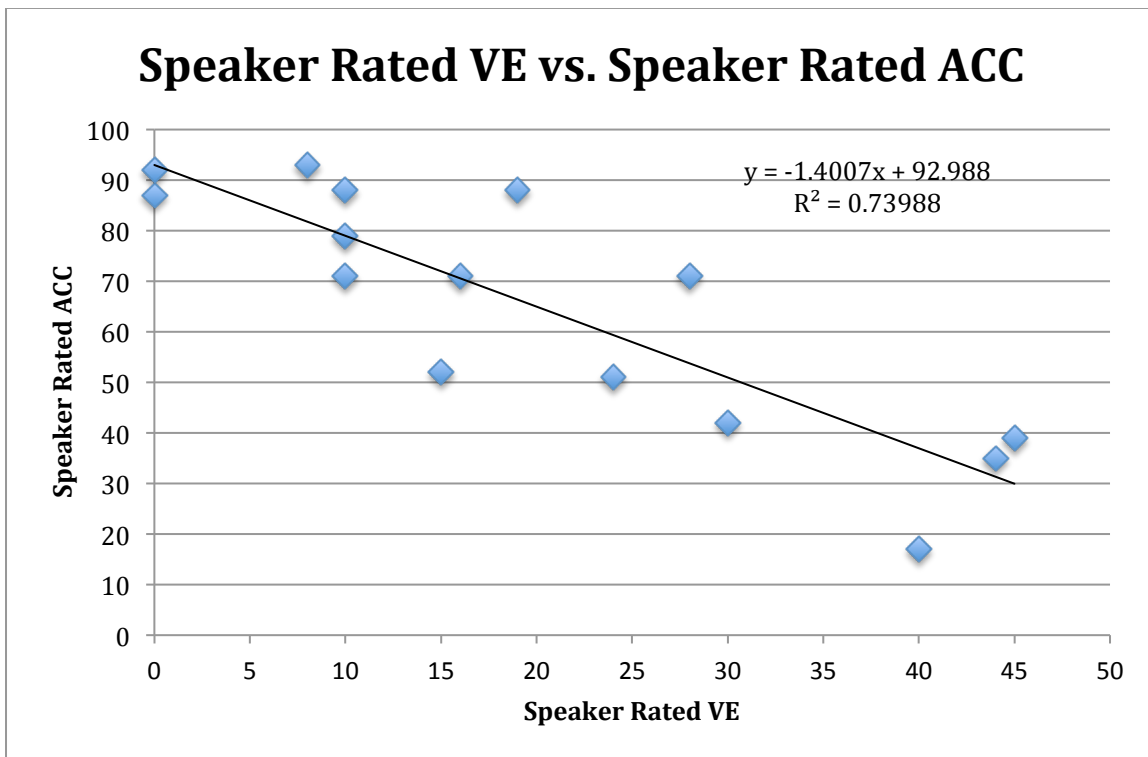


Figure 3: Correlation Between Speaker Rated Vocal Effort (VE) and Acceptability (ACC)

Relationship Between Listeners' Judgments of Acceptability and Vocal Effort

A Pearson correlation coefficient was calculated to determine the relationship between listener judgments of acceptability and vocal effort. A strong, significant negative correlation was found ($r(14) = -.802, p < .001$), indicating a linear relationship between the two variables (see figure 4). Listeners who rated speech as highly acceptable tended to rate its perceived vocal effort as low, and vice versa.

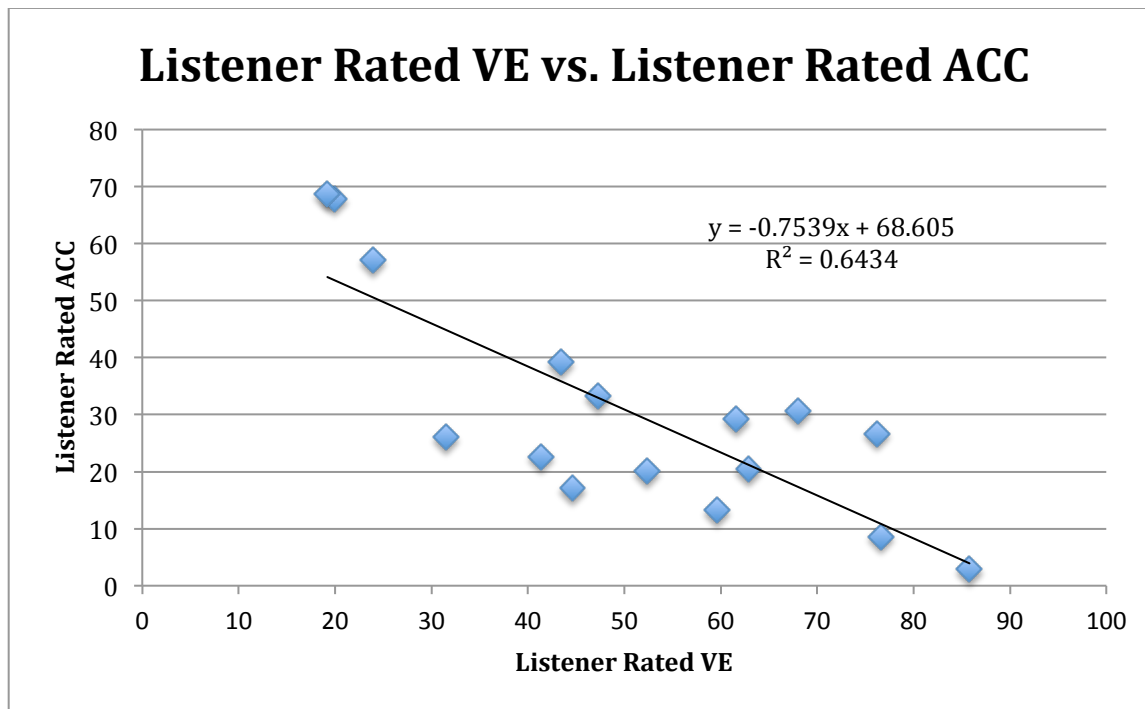


Figure 4: Correlation Between Listener Rated Vocal Effort (VE) and Acceptability (ACC)

Relationship Between Auditory-Perceptual Dimensions with Voice-Related QOL

Acceptability. A Pearson correlation coefficient was calculated to examine the relationship between speaker judgments of acceptability and voice-related QOL. A strong, significant negative correlation was found ($r(12) = -.702, p = .005$) between judgments and VHI-10 scores, indicating a linear relationship between the two variables (see figure 5). Speakers who rated their own speech acceptability as high, tended to have lower VHI-10 scores, and vice versa.

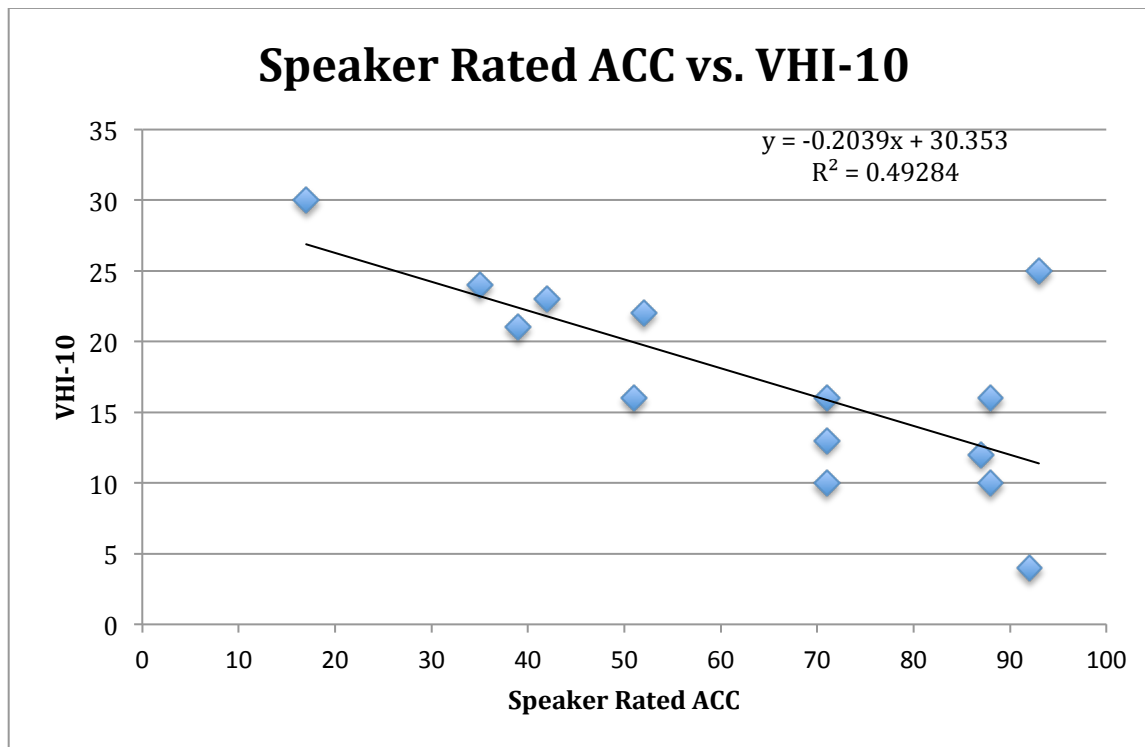


Figure 5: Correlation Between Speaker Rated Acceptability (ACC) and Voice-Related QOL (VHI-10)

A Pearson correlation coefficient also was calculated to examine the relationship between listener judgments of acceptability and voice-related QOL. A weak negative correlation was found, ($r(13) = -.109, p = .700$), indicating a very limited linear relationship between the two variables (see figure 6). Thus, listeners' ratings of acceptability do not appear to relate to speaker-rated voice-related QOL.

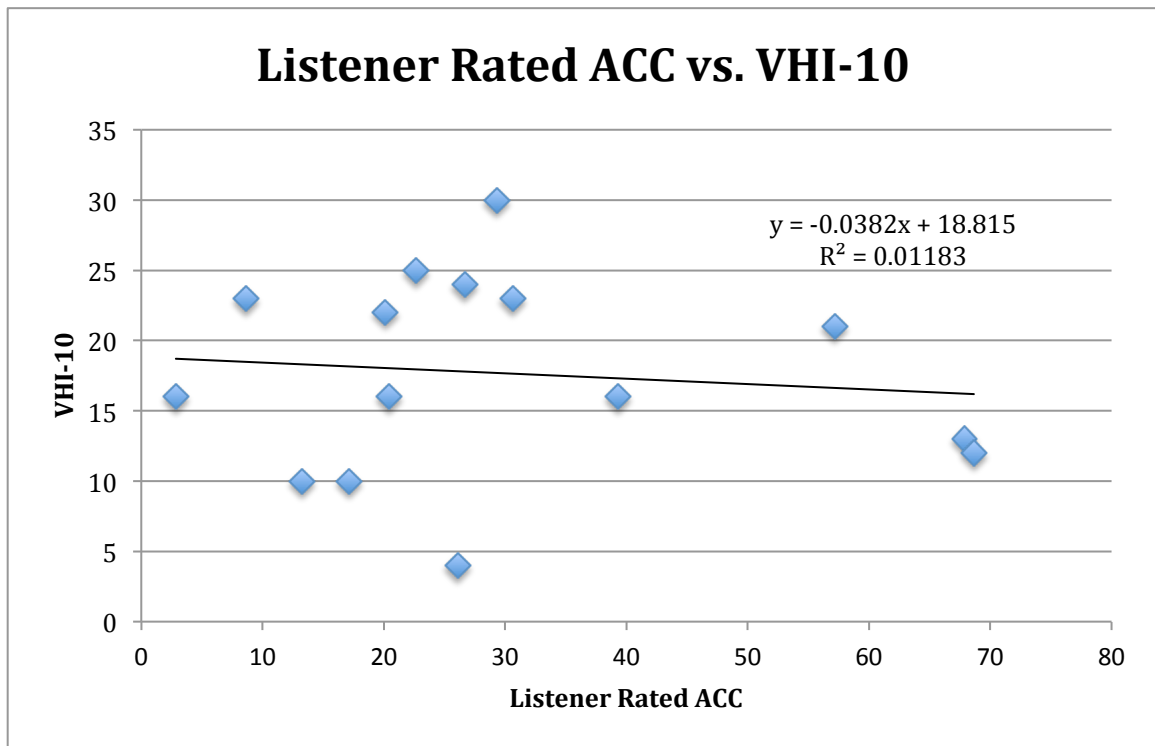


Figure 6: Correlation Between Listener Rated Acceptability (ACC) and Voice-Related QOL (VHI-10)

Vocal Effort. A Pearson correlation coefficient was calculated to examine the relationship between speaker judgments of vocal effort and voice-related QOL. A moderate to strong significant positive correlation was found ($r(12) = .676, p = .008$), indicating a linear relationship between the two variables (see figure 7). Speakers who rated their own vocal effort as high, tended to have high VHI-10 scores, and vice versa.

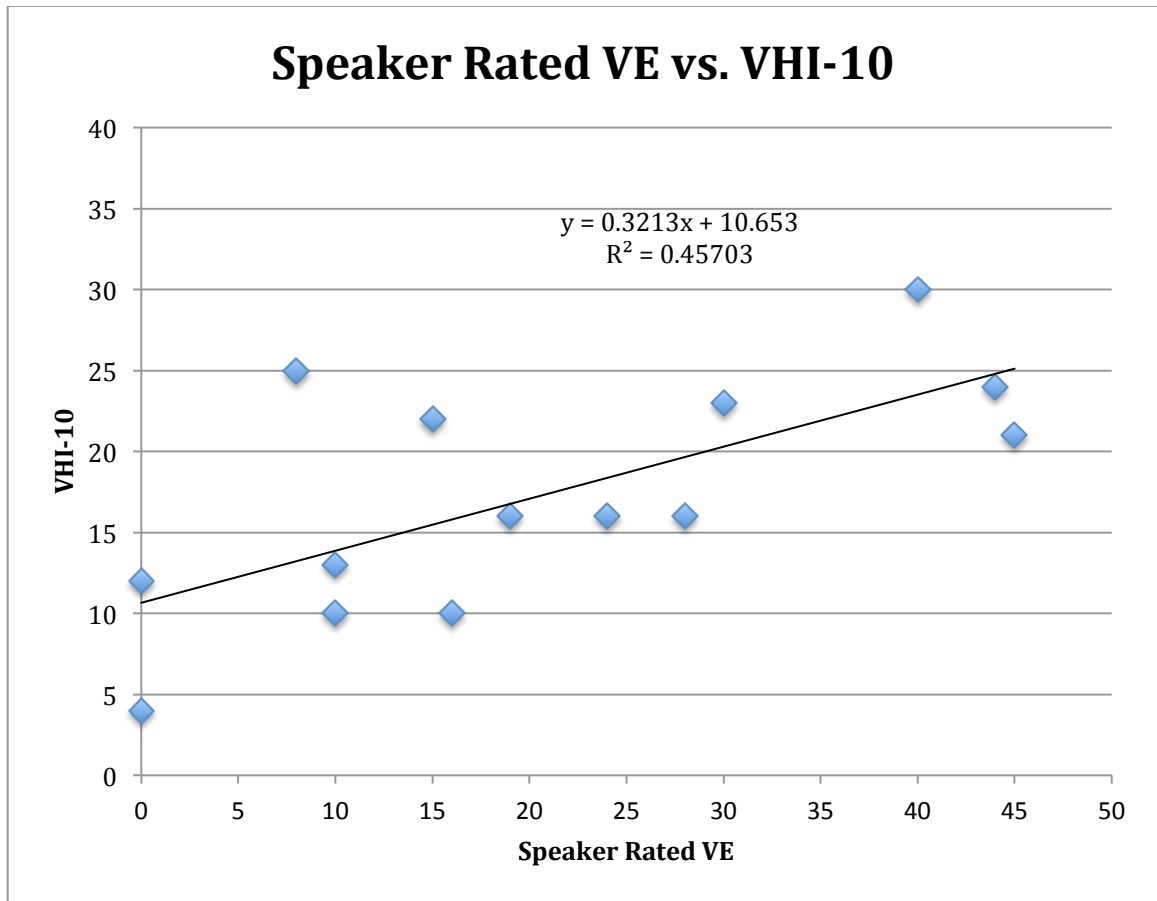


Figure 7: Correlation Between Speaker Rated Vocal Effort (VE) and Voice-Related QOL (VHI-10)

Finally, a Pearson correlation coefficient was calculated to determine the relationship between listener judgments of vocal effort and voice-related QOL. A moderate, but not statistically significant, positive correlation was found ($r(13) = .413, p = .126$), indicating a linear relationship between the two variables (see figure 8). In samples where listeners rated vocal effort as high, VHI-10 scores related to those samples were high, and vice versa.

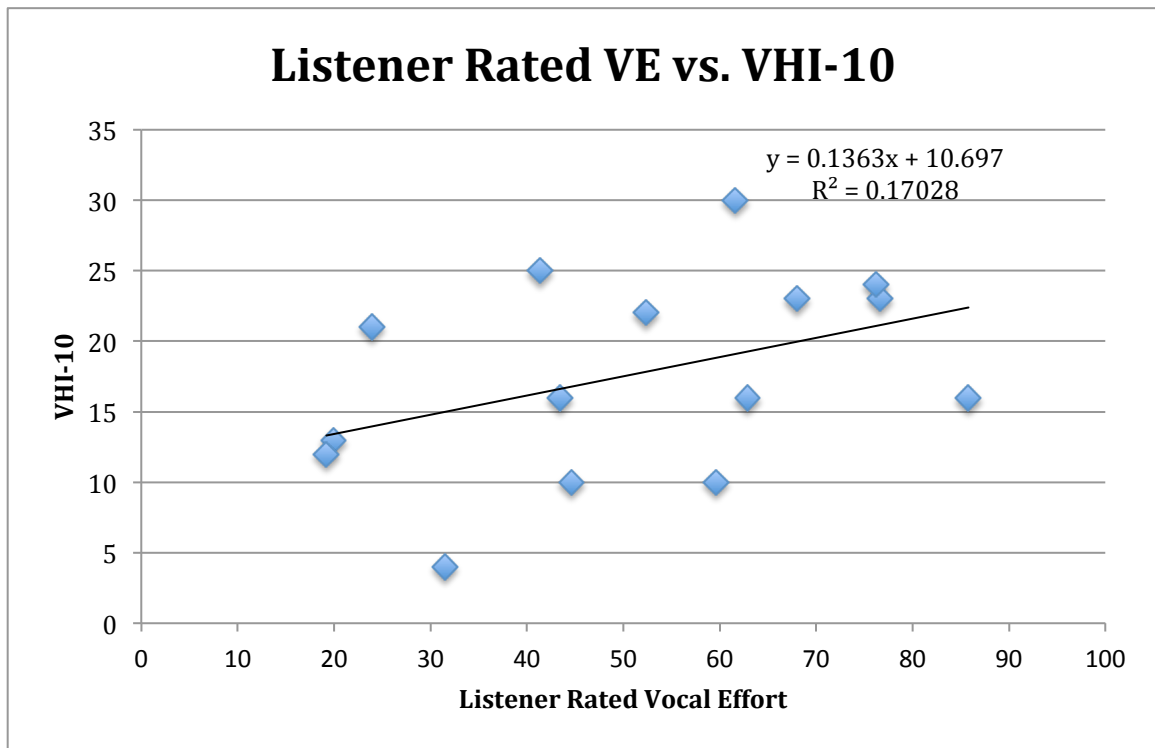


Figure 8: Correlation Between Listener Rated Vocal Effort (VE) and Voice-Related QOL (VHI-10)

A summary of the correlations between speakers' and listeners' judgments of acceptability, vocal effort, and voice-related quality of life scores is provided in Table 2.

Table 2: Summary of Correlations between Speakers' and Listeners' Judgments of ACC, VE, and VHI-10 scores

Correlation	<i>r</i>	Strength of Relationship
Listener Rated ACC vs. Speaker Rated ACC	.138	Weak
Listener Rated VE vs. Speaker Rated VE	.476	Moderate
Speaker Rated VE vs. Speaker Rated ACC	-.860**	Strong
Listener Rated VE vs. Listener Rated ACC	-.802**	Strong
Speaker Rated ACC vs. VHI-10	-.702**	Strong
Listener Rated ACC vs. VHI-10	-.109	Weak
Speaker Rated VE vs. VHI-10	.676**	Moderate to Strong
Listener Rated VE vs. VHI-10	.413	Moderate

**=Correlation is significant at the 0.01 level (2-tailed).

DISCUSSION

The purposes of the present study were threefold. The first objective was to determine if there were differences in judgments of speech acceptability or perceived VE when TE speakers and unfamiliar listeners rated these dimensions. The second purpose was to determine if relationships existed among judgments of speech acceptability and perceived VE when TE speakers and unfamiliar listeners rated these dimensions. The third objective was to determine how self-rated and listener-rated speech acceptability and VE related to voice-related QOL in TE speakers. In general, the results revealed that listeners judge TE speech as less acceptable and more effortful than speakers themselves, and that when relationships are examined, listener- and speaker-rated outcomes are only weakly to moderately related. In particular, judgments of vocal effort tended to be more strongly related between listeners and speakers than traditional ratings of speech acceptability. Interestingly, relationships between these two characteristics were strongest when both rated by the same group, either speakers or listeners. Finally, relationships between measures of speech acceptability and effort with voice-related QOL were only strong when the speakers judged these dimensions.

These findings are first discussed relative to speaker characteristics in the section to follow. Specific differences and relationships between auditory-perceptual measures, and the relationship between auditory-perceptual measures and voice-related QOL are presented next. Finally, implications of the present findings for both clinical practice and future research are summarized.

Speaker Demographics

Demographic characteristics of the TE speakers (gender, age, years post TL, radiation history, and VHI-10 totals) in the present study are summarized in Table 1. The average age (63

y) of the speakers in the present study was similar to that reported in other recent studies (Eadie et al., 2007; Eadie et al., 2008; Eadie & Doyle, 2004; Law, Ma, & Yiu, 2009). The proportion of men ($n = 14$) to women ($n = 2$) (i.e., 7:1) is relatively high when compared to the population diagnosed with laryngeal cancer (i.e., 2.44:1) (American Cancer Society, 2012). The small sample of women in this study makes conclusions difficult to generalize when considering any possible gender differences. However, when the data from the two women who participated were analyzed, their results seemed to be comparable to their male counterparts. How gender affects measures of vocal effort, acceptability, or voice-related QOL should be considered in future studies that include larger samples of women.

This study included people who had undergone TL at minimum 1 year prior to their participation, with a mean period after laryngectomy of 6.6 years. The fact that speakers were one or more years post-laryngectomy could mean that they have had more time to adjust to their new voice and the changes associated with TL. These results may not generalize to TE speakers in the acute phase after their TL, as they may not have had sufficient time to cope with the number of changes in their lives. However, post-hoc analyses from this study revealed only weak correlations between time post TL with all measures (voice-related QOL, speaker-rated auditory-perceptual measures, and listener-rated measures), indicating that at least within this group of TE speakers, time did not seem to play a large factor in outcomes. How time post TL affects the measures of vocal effort or acceptability should be considered in future longitudinal studies.

The majority of participants in this study had undergone some form of radiation therapy (87.5%), which is consistent with most other studies (Eadie & Doyle, 2004; Finizia et al., 1998). Finally, the mean VHI-10 score (17.7) (moderate handicap) is similar to scores reported in other recent studies (Evans, Carding, and Drinnan, 2009; Oridate et al., 2009), with a large range of

scores (from relatively no voice-related handicap to severe voice handicap). Collectively, the demographic information in the present study is comparable to data presented by other authors, suggesting that these results may directly relate to those found in the literature.

One factor that could have affected the results of this study pertained to the method of recruitment for the speakers. Specifically, the majority of participants were recruited at the national conference of the International Association of Laryngectomees (IAL). Recruiting participants who have the ability to attend a national conference may be a potential limitation as they may be healthier, more financially stable, and adjusted to their TL, allowing them to seek out others with experiences similar to theirs. Participants' ability to travel would depend greatly on their health and financial status. These factors could bias the sample towards including those who have a higher voice-related QOL because those in greater health and higher socioeconomic status may have better health-related outcomes (Blood et al., 1992; Groome et al., 2006). Similarly, attending an international conference that addresses life after TL may suggest that participants are better adjusted to their condition, or seeking a means to become better adjusted, than their counterparts who are not interested in attending this conference. All of these factors would influence QOL measures in a positive way. Conversely, recruiting participants who are in attendance at a conference, such as the IAL, may have affected scores negatively. For example, attending a conference may be an indication that these speakers are actually worse off than their counterparts because they feel the need to seek out support, whereas others may not feel the need to attend a support group or conference. The results of this study must be interpreted with these potential biases in mind.

Difference Between Speakers' and Listeners' Auditory-Perceptual Judgments

In the present study, a significant difference was found between the speakers' and listeners' judgments of acceptability, with speakers rating acceptability significantly higher than listeners. This difference indicates that speakers find their voices to be significantly more acceptable than listeners. These data are consistent with results found by Finizia et al. (1998), who reported that patients who had undergone a laryngectomy judged their own voices to be significantly more acceptable than listeners rating the speakers' voices. This finding suggests that speakers and listeners may be using different internal standards to judge the auditory-perceptual characteristic of acceptability. Specifically, it might be suggested that speakers with alaryngeal voices adapt over time to how their voice sounds when compared to what would be considered "normally acceptable" (Kreiman et al., 1993), but that unfamiliar listeners continue to penalize alaryngeal speakers for the "differentness" of their speech. This result also supports the continued use of speech acceptability as a dimension that measures the social penalty of alaryngeal speech, since TE speakers were on the whole judged more "unacceptable" than "acceptable" (mean listener-rated judgment = 30.24 on a 100mm VAS) (Eadie & Doyle, 2004; Finizia et al., 1998, Law, Ma, & Yiu, 2009).

Similar to acceptability ratings, results showed that listeners rated VE significantly higher than speakers. This difference suggests that listeners perceive TE speech to be significantly more effortful than speakers. These differences have not yet been investigated in the alaryngeal literature, but have been examined in those with laryngeal-based voice disorders (Eadie et al., 2007; Eadie et al., 2010). In contrast to the present study, previous results showed no significant group differences between speaker-rated VE and listener-rated VE in individuals with hyperfunctional voice disorders or spasmodic dysphonia. Though these populations may use increased air pressures for voice production, they may not be as obvious as the increases in air

pressures used in TE speech (Bohnenkamp et al., 2011). One possible explanation for the present results may relate to an adaptation response. Because listeners are not accustomed to listening to the different voice quality accompanying TE speech (as well as the increased air pressures required for this intrinsic method of speech production), they may be more sensitive to the differences they hear that require more effort by the speaker. Similarly, as this new method of speech becomes more familiar and practiced to the speaker, he or she might begin to associate less effort with voice production. How these perceptual judgments relate to the actual air pressures used by the TE speakers should be a subject of future investigations. Specifically, a longitudinal study might look at speakers' judgments of vocal effort in the acute phase post TL or post TE puncture, and continue with follow up judgments of vocal effort every few months to determine if, and how quickly, adaptation to their new voice occurs. One could also examine the effect of different types of TE prostheses (indwelling or non-indwelling) and different sizes on both judgments of VE as well as air pressures. However, differences between speaker- and listener-rated outcomes first must be reconciled with how these measures relate to one another.

Relationships Between Speakers' and Listeners' Auditory-Perceptual Judgments

The results of the present study suggest that speaker and listener judgments of acceptability do not appear to relate to one another as evidenced by a weak positive correlation ($r = .138$). This finding further supports the idea that speakers and listeners use different strategies and standards to evaluate auditory-perceptual characteristics of speech. These results differ from those who have reported moderate relationships between auditory-perceptual judgments made by speakers and listeners in other voice-disordered populations (Eadie et al., 2007; Eadie et al., 2008).

In contrast to the weak relationship found between speakers' and listeners' judgments of acceptability, relationships between speakers' and listeners' VE ratings were moderate ($r = .476$). This moderate positive correlation between the two variables suggests that VE, when judged by a listener, is a stronger predictor of the speaker's own judgment of VE, than the perceived acceptability of speech. The results from this study are consistent with those reported by Eadie et al. (2007), who also found moderate relationships between VE judgments made by speakers with spasmodic dysphonia and other listeners.

One reason why results may differ between acceptability and VE may relate to what is actually being judged. For example, speech acceptability, as defined in this study, asks a person to judge each voice based on pitch, rate, and understandability (Bennett & Weinberg, 1973). In judging acceptability, listeners are rating multiple qualities at a time while giving these qualities one overall score. This may be a potential limitation because different listeners may consider one aspect of acceptability (i.e., pitch, rate, or understandability) as more influential in their judgment than the other two, making it difficult to compare their rating to the ratings of others. Similarly, this dimension is heavily influenced by social norms and ideals that bias each rater. Before rating a voice for acceptability, listeners have already determined how a typical voice should sound. Because TE speech is different than laryngeal speech, listeners may be biased and automatically rate acceptability as lower due to the fact that it does not fit their idea of the wide range of socially acceptable voices (Eadie & Doyle, 2005; Eadie, Doyle, Hansen, & Beaudin, 2008). Despite the potential variability in using this dimension, unfamiliar listeners do appear to make this judgment consistently (as demonstrated by strong interrater reliability and by the strong relationship demonstrated between this variable and VE). These data support the use of speech acceptability to gauge the impact of TE speech on unfamiliar communication partners.

Making judgments of speech acceptability are somewhat different than making judgments of voice quality, or another physically related dimension of voice, such as VE. In rating a physical dimension such as VE, listeners appear to use the acoustic signal and compare it with an internal anchor of their own experience with effortful speech production to determine whether or not a voice sounds effortful (Brandt et al., 1969). One explanation for the moderate relationship between speakers' and listeners' judgments of VE is that speakers and listeners have both experienced effortful voice production at one or more points in their lives, which could help to create an internal anchor related to VE that is similar in both groups. However, speaker-rated VE is different in that the speaker has access to kinesthetic cues as well as acoustic cues (Brandt et al., 1969), whereas the listener is limited to acoustic cues only for making these judgments. This is a possible reason why the relationships between speaker-rated and listener-rated VE remains only moderate. The physiological and acoustic bases of these judgments should be studied in the future.

Relationships Between Judgments of Acceptability and Vocal Effort

Results from the present study showed a strong negative correlation between speakers' judgments of acceptability and perceived VE ($r = -.860$). Specifically, speakers who rated their voice as less acceptable tended to have higher ratings of perceived VE. This finding suggests that, in speakers, self-rated acceptability is a strong predictor of perceived VE, and vice versa. It also might indicate that speakers are using similar strategies to evaluate these two perceptual characteristics. However, whether one outcome (e.g., speech acceptability) is causally dependent on another variable (e.g., VE) is a subject for future research.

Similar to the speaker findings, listeners' judgments of acceptability and VE appear to be strong predictors of one another, as evidenced by a significant negative correlation ($r = -.802$)

between the two variables. This study found that listeners who rated a speaker's voice as highly acceptable tended to perceive the same voice as requiring less VE. Conversely, listeners who rated a speaker's voice as less acceptable tended to perceive the same voice as requiring more VE. This may be an indication that listeners use similar strategies for making these judgments. Yet, because VE and acceptability were only moderately (VE) or weakly (acceptability) related to judgments made by other listeners (speakers vs. listeners), it indicates that while each type of rater appeared to use these dimensions similarly, they appeared to be used differentially by each type of rater. These results again support the notion that speakers and other listeners may be using different cues for making these types of judgments.

Previous studies have compared relationships between different types of speech dimensions (e.g., naturalness vs. acceptability in alaryngeal speech, vocal effort vs. overall severity in other populations) (Eadie et al., 2008; Eadie & Doyle, 2004). Eadie and Doyle (2004) compared ratings of naturalness and acceptability in TE speakers as judged by naïve listeners. They found that these two dimensions were strongly related to one another ($r = .883$). This suggests that listeners used these dimensions similarly when rating each voice, which introduces the question of whether or not there is a use for rating both dimensions. The authors also concluded that the listeners were capable of partitioning these two dimensions, as well as the dimensions of severity and pleasantness. They argued that use of multidimensional constructs help to evaluate the global character of TE speech, which supports the use of rating both naturalness and acceptability, even though they are strongly related dimensions.

Eadie et al. (2010) compared the dimensions of overall severity (OS) and vocal effort as judged by speakers with dysphonia (laryngeal-based voice disorders), naïve listeners, and experienced listeners. No significant differences were found across the groups' judgments of OS

and VE, which indicated that the two appeared to be relatively equivalent dimensions. Therefore it may not be necessary to rate speakers on both dimensions due to this suggested equivalency. However, when the relationships between the groups' judgments of OS and VE were examined, it was found that naïve and experienced listeners' judgments were strongly and significantly correlated, but that listener-rated judgments were only weakly to moderately related to speakers' ratings. Again, this suggests that listeners, no matter their level of experience, are using similar strategies to rate both OS and VE. However, it also supports the contention that speakers are using different strategies for making their judgments. Collectively, these results may support the use of OS or VE as clinician-rated dimensions, particularly for populations in which VE is the most salient dimension (e.g., such as hyperfunctional-based voice disorders).

In the present study, acceptability and VE were shown to strongly correlate to one another within groups (listeners-listeners and speakers-speakers). Because of this result, it would be expected that they would also be similarly related across the listener and speaker groups. However, in this study, this was not the case. VE was the only dimension shown to be moderately related across the two groups ($r = 0.476$), while acceptability showed a weak relationship across groups ($r = 0.138$). This result suggests that the two dimensions are somewhat different and may indicate that both dimensions should be considered when evaluating TE speech. It also implies that if listeners are limited to judging only one dimension, judgments of VE may be a better predictor of the speakers' perception of their own voice than judgments of acceptability.

Relationships Between Auditory-Perceptual Dimensions with Voice-Related QOL

In the present study, when speakers' ratings of acceptability were compared to their own VHI-10 scores, a strong negative correlation ($r = -.702$) was found. Thus, the higher a speaker

rated his or her own speech acceptability, the better his or her voice-related QOL (with low VHI-10 scores representing better voice-related QOL). Similar to speech acceptability, data also revealed a moderate to strong relationship between speaker-rated judgments of VE with voice-related QOL ($r = .676$).

One possible explanation for these strong observed relationships between measures might relate to the fact that both are derived from the same person (speaker-rated acceptability/VE and VHI-10 scores). Because the person who has undergone the TL best understands how the treatment and its consequences have affected speech and voice, including its impact on well-being, it was hypothesized that these measures would more strongly relate to one another than dimensions rated by others (including clinicians and naïve listeners). This finding is supported by previous studies that have shown naïve and experienced listeners' judgments of spasmodic dysphonia to be only weakly ($r = 0.30 - 0.38$) to moderately ($r = 0.59 - 0.63$) related to the speaker's voice-related QOL (Eadie et al., 2007; Eadie et al., 2010). Though experienced listeners were not included in this study, previous literature has revealed that their judgments of voice quality or perceived VE may more strongly predict voice-related QOL than judgments made by naïve listeners. These results have been found for speakers with benign vocal fold lesions; however, it must be noted that even experienced listeners' judgments are only moderate ($r = 0.44 - 0.48$) predictors (Behrman, Sulica, & He, 2004; Murry, Medrado, Hogikyan, & Aviv, 2004).

In contrast to the strong relationships between with speaker-rated speech dimensions and voice-related QOL, there was only a weak (trivial) relationship ($r = -.109$) between listeners' judgments of acceptability and the speaker's VHI-10 scores. This result indicates that others' perceptions of the acceptability of a speaker's voice do not relate to that speaker's voice-related

QOL. Therefore, it is not appropriate to use unfamiliar listeners' judgments of acceptability as a predictor of voice-related QOL in TE speakers. Because unfamiliar listeners do not have the knowledge of how TL affects one's life, it is understandable that their subjective ratings of the acceptability of speakers' voices would not relate to the speakers' own ratings of their voice-related QOL. This finding is consistent with Eadie et al. (2012), who also reported a weak ($r = -0.295$) relationship between listeners' ratings of acceptability of TE speakers' voices and the respective TE speakers' VHI-10 scores.

Contrary to the weak relationship found between listener ratings of acceptability and voice-related QOL, this study found a moderate ($r = .413$) relationship between listener ratings of VE and voice-related QOL. This finding is supported by previous research that also found a moderate relationship ($r = 0.59$) between naïve listener judgments of VE and patient-rated VHI scores in speakers with dysphonia (Eadie et al., 2010). Results from this study are somewhat stronger than an earlier study that investigated relationships between naïve listener judgments of VE and voice-related QOL in spasmodic dysphonia ($r = 0.320$) (Eadie et al., 2007).

This finding is interesting because, even though as a whole outcome measures are strongest when judged by the same raters, only listener-rated VE was found to be a moderate predictor of TE speakers' own assessment of voice-related QOL. Because listener-rated VE was shown to have a strong relationship with listener-rated acceptability, one might hypothesize that listener-rated acceptability should also show a moderate correlation to voice-related QOL. However, this was not observed to be the case, as the results of this study found only a weak relationship. One possible explanation for this finding is that VE is a more physical, concrete measure (i.e., potentially related to underlying air pressures), while acceptability is inherently related to social concepts of what is considered "normal" (Doyle & Keith, 2005). Though the

precise similarities and differences between acceptability and VE are unclear, future studies should attempt to replicate the finding of the current study that listener-rated VE serves as a predictor of speaker-rated QOL, while listener rated acceptability does not appear to predict speaker-rated QOL in TE speech. These studies may consider limiting speakers to only those using TE (or esophageal) speech, as VE is most likely only a sensitive measure for those who use intrinsic methods of alaryngeal speech.

Future Directions and Clinical Implications

In the present investigation, speakers were asked to report demographic information, including primary method of alaryngeal speech, years post TL, and history of radiation therapy. While this is all pertinent information, one limitation of the present study is that more specific information regarding the speakers' smoking status/lung health and type of heat/moisture exchanger (HME) cassette used to cover their stoma was not collected. In the future, it would be important to know a speaker's smoking status/lung health because, if a person's baseline already indicates compromised respiratory function, their rating of VE may be negatively impacted. Bohnenkamp et al. (2011) reported that there were no significant differences in lung volumes used for speech breathing between TE speakers with Chronic Obstructive Pulmonary Disease (COPD) and those without. However, it was noted that this finding could have been limited by small sample size and by the possibility that some speakers in their study potentially did not report symptoms of COPD, even though they in fact had the disease.

Similarly, the type of HME cassette a speaker uses, or if he does not use one, should be noted. Literature surrounding HME use and respiratory function suggests that alaryngeal speakers who use HMEs report significantly improved respiratory health (e.g., decreases in sputum production, forced expectoration to clean the airway, and stoma cleaning) (Hilgers,

Aaronson, Ackerstaff, Schouwenburg, & van Zandwijk, 1991). Just as decreased lung function may negatively impact VE, it is hypothesized that increased respiratory health may positively impact VE, making this worthwhile information to have when analyzing alaryngeal speech performance. However, not all HMEs are the same, as different companies produce multiple versions of these cassettes. Hence, noting the type of HME used would also be necessary to determine if there are notable differences between types of HMEs and speakers' judgments of VE. Therefore, future studies should include the smoking status/lung health as well as the type of HME cassette used by alaryngeal speakers.

A second possible limitation of the present study is the small sample of TE speakers ($n = 16$). Since the sample of speakers was small, these results should be interpreted with caution when generalized to the larger population of TE speakers due to the inherent variability of the sample. It is possible that correlations between groups were weak to moderate because the sizes of the groups were too small. Yet, the range of scores was variable in this study, suggesting that correlational analysis was appropriate. Similarly, because the current study was limited to analyzing TE speech, it is possible that these results may only apply to TE speakers. Though it is possible that these findings may apply to the other intrinsic alaryngeal speech method, esophageal speech, this study must be replicated within that population to determine whether the results may be generalized.

The findings of this study demonstrate important implications for future practice. Currently, the auditory-perceptual dimension of speech acceptability is commonly used to measure the success of voice rehabilitation in alaryngeal speech in clinical and research practice. It is also somewhat common for clinicians to administer a voice-related QOL measure to alaryngeal speakers to determine the impact of their new voice on everyday situations. Though a

common voice-related QOL measure (e.g., VHI-10) is a self-report form, speech acceptability is typically judged by the clinician.

Based on the results of this study, several recommendations could be made. First, because self- and listener-rated judgments of acceptability and VE do not relate strongly to one another, this weak relationship suggests that TE speakers should be asked to make these types of judgments in a standard assessment (in addition to listener-based measures). However, whether the speaker should be asked to make both types of judgments is debatable, since both measures are strongly related when made by the same person, and because they similarly predict voice-related QOL in TE speakers. Since these measures do not take long to administer, clinicians could include both measures, especially since it is unknown whether similar results would be found in those who have just begun using TE speech as their primary method of communication. This is an area for future research.

Second, results from this study support the notion that VE should be added to standard listener-rated outcomes for TE speech. VE is a newer measure of the success of voice rehabilitation in alaryngeal speech and is used less commonly than acceptability. However, this study showed that while others' judgments of VE moderately predicted a speaker's voice-related QOL, acceptability did not show a relationship with voice-related QOL. This finding might imply that, in standard voice evaluations, the characteristic of VE is an important quality that should be measured, and that it is different than speech acceptability. For this reason, these two dimensions could be considered as complementary, and not redundant measures of TE speech performance.

Finally, both self-rated VE and speech acceptability were related to voice-related QOL in this study. While these relationships are not causal, one might predict that by decreasing a

speaker's VE (or by increasing that person's satisfaction with his or her speech acceptability), an increase in voice-related QOL might occur. This would imply that by introducing the speaker to a treatment that targets decreased VE (or enhances speech acceptability), their everyday participation in activities might also improve. For example, this might involve use of decreased air pressure to drive voice production using automatic speaking valves, use of HME cassettes to improve respiratory function, or improvements in phrasing or articulation of TE speech. In addition, rehabilitation that focuses on counseling or education of unfamiliar communication partners could enhance satisfaction with outcomes. Possible development of other measures or adaptations of current measures may be needed to capture the unique speech and voice characteristics in this alaryngeal population. By gaining a better understanding of the influence of perceptual measures on QOL through these future studies, speech-language pathologists will improve their ability to care for this specialized group of individuals.

References

- American Cancer Society. (2012). *Cancer facts and figures 2012*. Atlanta, GA: American Cancer Society.
- Andrews, M. L. (1999). *Manual of voice treatment: Pediatrics through geriatrics* (2nd ed.). San Diego, CA: Singular Publishing Group, Inc.
- Azevedo, A. H. M., Montoni, N., Filho, J. G., Kowalski, L. P., & Angelis, E. C. (2011). Vocal handicap and quality of life after treatment of advanced squamous carcinoma of the larynx and/or hypopharynx. *Journal of Voice*, e-publication ahead of print, 1-9.
- Behrman, A., Sulica, L., & He, T. (2004). Factors predicting patient perception of dysphonia caused by benign vocal fold lesions. *Laryngoscope*, 114, 1693-1700.
- Bellandese, M. H. (2009). Fundamental frequency and gender identification in standard esophageal and tracheoesophageal speakers. *Journal of Communication Disorders*, 42(2), 89-99.
- Bellandese, M. H., Lerman, J. W., & Gilbert, H. R. (2001). An acoustic analysis of excellent female esophageal, tracheoesophageal, and laryngeal speakers. *Journal of Speech, Language, and Hearing Research*, 44, 1315-1320.
- Bennett, S. & Weinberg, B. (1973). Acceptability ratings of normal, esophageal, and artificial larynx speech. *Journal of Speech, Language, and Hearing Research*, 16, 608-615.
- Blood, G. W., Luther, A. R., & Stemple, J. C. (1992). Coping and adjustment in alaryngeal speakers. *American Journal of Clinical Speech Language Pathology*, 1, 63-69.
- Bohnenkamp, T. A., Forrest, K. M., Klaben, B. K., & Stager, J. M. (2011). Lung volumes used during speech breathing in tracheoesophageal speakers. *Annals of Otolaryngology, Rhinology, and Laryngology*, 120(8), 550-558.
- Brandt, J. F., Ruder, K. F., & Ship, T. (1969). Vocal loudness and effort in continuous speech. *Journal of the Acoustical Society of America*, 46, 1543-1548.
- 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, Texas: Pro-ed.
- Eadie, T. L. (2003). The ICF: a proposed framework for comprehensive rehabilitation of individuals who use alaryngeal speech. *American Journal of Speech-Language Pathology*, 12, 189-197.
- Eadie, T. L. & Doyle, P. C. (2005). Scaling of voice pleasantness and acceptability in tracheoesophageal speakers. *Journal of Voice*, 19 (3), 373-383.

- Eadie, T. L. & Doyle, P. C. (2004). Auditory-perceptual scaling and quality of life in tracheoesophageal speakers. *Laryngoscope*, 114, 753-759.
- Eadie, T. L., Doyle, P. C., Hansen, K., & Beaudin, P. G. (2008). Influence of speaker gender on listener judgments of tracheoesophageal speech. *Journal of Voice*, 22 (1), 43-57.
- Eadie, T.L., Kapsner, M., Rosenweig, J., Waugh, P., Hillel, A., & Merati, A. (2010). The role of experience on judgements of dysphonia. *Journal of Voice*, 25(5), 564-573.
- Eadie, T., Nagle, K., Day, A., Sawin, D., Isetti, D., Lamvik, K., & Doyle, P. (2011, November). Auditory-perceptual outcomes and quality of life among alaryngeal speakers. Poster session presented at ASHA Annual Conference, San Diego, CA.
- Eadie, T.L., Nicolici, C., Baylor, C., Almand, K., Waugh, P., & Maronian, N. (2007). Effect of experience on judgements of adductor spasmodic dysphonia. *Annals of Otolology, Rhinology, and Laryngology*, 116(9), 564-573.
- Evans, E., Carding, P., & Drinnan, M. (2009). The Voice Handicap Index with post-laryngectomy male voices. *International Journal of Language and Communication Disorders*, 44(5), 575-586.
- Fairbanks, G. (1960). *Voice and articulation drillbook*. New York: Harper & Bros.
- Franic, D. M., Bramlett, R. E., & Bothe, A. C. (2005). Psychometric evaluation of disease specific quality of life instruments in voice disorders. *J Voice*, 19(2), 300-315.
- Finizia, C., Hammerlid, E., Westin, T., & Lindstrom, J. (1998). Quality of Life and voice in patients with laryngeal carcinoma: a posttreatment comparison of laryngectomy (salvage surgery) versus radiotherapy. *The Layngoscope*, 108, 1566-1573.
- Groome, P. A., Schulze, K. M., Keller, S., Mackillop, W. J., O'Sullivan, B., Irish, J. C., . . . Warde, P. R. (2006). Explaining socioeconomic status effects in laryngeal cancer. *Clinical Oncology*, 18, 283-292.
- Hartl, D. M., Ferlito, A., Brasnu, D. F., Langendijk, J. A., Rinaldo, A., Silver, C. E., & Wolf, G. T. (2011). Evidence-based review of treatment options for patients with glottis cancer. *Head and Neck*, 33(11): 1638-1648.
- Hilgers, F. J., Aaronson, N. K., Ackerstaff, A. H., Schouwenburg, P. F., & van Zandwijk, N. (1991). The influence of a heat and moisture exchanger (HME) on the respiratory symptoms after total laryngectomy. *Clinical Otolaryngology and Allied Sciences*, 16 (2), 152-156.
- 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). Pro-Ed: Austin.

- Jacobson, B. H., Johnson, A., Grywalski, 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., Rhys-Evans, P., & Harrington, K.J. (2007). Voice-related Quality of Life in laryngectomees: assessment using the VHI and V-RQOL symptom scales. *J Voice*, 21(6), 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(11):770-771.
- Kotby, M. N., Hegazi, M. A., Kamal, I., Gamal de Dien, N., & Nassar, J. (2009). Aerodynamics of the pseudo-glottis. *Folia Phoniatica et Logopaedia*, 61, 24-28.
- Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research*, 36(1), 21-40.
- Law, I., Ma, E., & Yiu, E. (2009). Speech intelligibility, acceptability, and communication-related quality of life in chinese alaryngeal speakers. *Archives of Otolaryngology-Head and Neck Surgery*, 135(7), 704-711.
- Lundstrom, E., Hammarberg, B., & Munck-Wikland, E. (2009). Voice handicap and health-related quality of life in laryngectomees: assessments with the use of VHI and EORTC questionnaires. *Folia Phoniatr Logop*, 61(2), 83-92.
- Miralles, J. L. & Cervera, T. (1995). Voice intelligibility in patients who have undergone laryngectomies. *Journal of Speech, Language, and Hearing Research*, 38:564-571.
- Moerman, M., Martens, J.P., & Dejonckere, P. (2004). Application of the Voice Handicap Index in 45 patients with substitution voicing after total laryngectomy. *Eur Arch Otorhinolaryngol*, 261(8), 423-428.
- Moon, J. B. & Weinberg, B. (1987). Aerodynamic and myoelastic contributions to tracheoesophageal voice production. *Journal of Speech and Hearing Research*, 30, 387-385.
- Murry, T., Medrado, R., Hogikyan, N. D., & Aviv, J. E. (2004). The relationship between ratings of voice quality and quality of life measures. *Journal of Voice*, 18, 183-192.
- Oridate, N., Homma, A., Suzuki, S., Nakamaru, Y., Suzuki, F., Hatakeyama, H., Taki, S., Sakashita, T., Nishizawa, N., Furuta, Y., & Fukuda, S. (2009). Voice-related quality of life after treatment of laryngeal cancer. *Archives of Otolaryngology- Head and Neck Surgery*, 135(4), 363-368.

- Pindzola, R. H. & Cain, B. H. (1988). Acceptability ratings of tracheoesophageal speech. *Laryngoscope*, 94(4), 394-397.
- Plant, R. L. (2005). Aerodynamics of the human larynx during vocal fold vibration. *Laryngoscope*, 115(12), 2087-2100.
- Portone, C.R., Hapner, E.R., McGregor, L., Otto, K., & Johns, M.M. (2007). Correlation of the Voice Handicap Index (VHI) and the Voice-Related Quality of Life Measure (V-RQOL). *J Voice*, 21(6), 723-727.
- Rosen, C.A., Lee, A.S., Osborne, J., Zullo, T., & Murry, T. (2004). Development and validation of the voice handicap index-10. *Laryngoscope*, 114(9), 1549-1556.
- Tardy-Mitzell, S., Andrews, M.L., & Bowman, S. A. (1985). Acceptability and intelligibility of tracheoesophageal speech. *Archives of Otolaryngology*, 111, 213-215.
- World Health Organization. (1997). WHOQOL: measuring quality of life. Retrieved from http://www.who.int/mental_health/media/68.pdf.