

A Critical Review and Study:

Determining the Presence of a Gender Bias in Perceptual Ratings of Tracheoesophageal Speakers

Jenna L. Wilkinson

M.Cl.Sc (SLP) Candidate

University of Western Ontario: School of Communication Sciences and Disorders

This report presents findings from a critical review of literature and the results of a study examining the relationship between gender and perceptions of tracheoesophageal (TE) speech. The critical review included evaluations of two randomized block designs, a between groups case-control study and one between groups post test design. The study involved collecting perceptual judgments on scales of acceptability and listener comfort from listeners ($n=16$) who were naïve to TE speech. These judgments were made on two occasions: one during which listeners were deceived about the gender of voice samples presented, and the second during which listeners were told the true gender of voice samples. Findings from both the critical review and pilot study suggest that a gender bias exists in perceptions of TE speech, and that female TE speakers tend to be disproportionately penalized when compared to their male counterparts.

Introduction

Upon a diagnosis of laryngeal cancer, an individual may undergo surgical removal of the larynx in what is termed “total laryngectomy” to eradicate the malignancy. However, following laryngectomy, clients are left without a natural voice, and must undergo therapy to learn how to produce a postsurgical “alaryngeal” voice. In many instances today, a method of surgical-prosthetic voice rehabilitation is used. This alaryngeal method is referred to as tracheoesophageal (TE) voice restoration (Singer & Blom, 1980). In this procedure, a small controlled puncture is created between the common tracheal and esophageal wall and a one-way voice prosthesis is inserted. This voice prosthesis allows air from the lungs to be diverted into the esophagus where the alaryngeal voice source occurs. This vibrated TE sound source then flows upwards to the oral cavity, where it is manipulated by the articulators to produce speech. Despite the success of restoring “voice” to those who have undergone laryngectomy, TE speech is very different from the normal laryngeal voice in terms of its fundamental frequency (pitch), the level of “noise” in the signal, and its overall voice quality (Robbins, Fisher, Blom, & Singer, 1984). While much variation exists among TE speakers, it is clear when listening to a TE speaker that his or her voice is quite unlike that of a normal laryngeal speaker (Doyle & Eadie, 2005). As voice quality is a construct fundamentally based on perception, auditory-perceptual ratings are considered the gold standard in the area of voice research (Kreiman, Gerratt, Kempster, Erman & Berke, 1993).

Historically, studies of TE speakers have focused on males, however, it is important to consider the female perspective, as an increasing number of women today

require total laryngectomy and become TE speakers (Canadian Cancer Society, 2013). Considering the change in voice quality in the context of gender is important, as women who use TE speech experience a considerable deviation from their pre-laryngectomy voice (Eadie, Doyle, Hansen & Beaudin, 2008). Specifically, female TE speakers experience a significantly lower pitch, a rough voice quality, and the loss of a feminine sounding voice (Pindzola & Cain, 1988). In fact, female TE speakers commonly report being misidentified as male when speaking on the telephone (Eadie et al., 2008). Therefore, it is possible that when compared to men, women experience more negative judgments of their TE speech because of its more significant deviation from the expected qualities of a female voice. Additionally, reductions in voice quality for women have been shown to be socially penalizing (Eadie, Doyle, Beaudin, & Day, 2012).

Currently, little is known about how a listener’s awareness of a TE speaker’s gender affects the judgments that he or she makes about a TE speaker’s voice. Previous studies in this field have confirmed that auditory-perceptual data are valuable as they generally complement patient-reported outcomes following laryngectomy (Eadie, Day, Sawin, Lamvik, & Doyle, 2012). Perceptual ratings can be valuable in helping to understand the different experiences of male and female TE speakers in terms of their desire to share, communicate, and participate in life activities, as the potential for social penalty exists. Clinically, this information is important as it may shape the counseling that is provided pre- and postlaryngectomy. Additionally, it is important to know whether or not there is a need for a more feminine sounding method of alaryngeal voice. It is hoped that the information presented in this study

will empirically identify the presence or absence of a gender bias in perceptions of TE speech, in an effort to provide clinically relevant data to the field of Speech-Language Pathology.

Objectives

This purpose of this paper sought to critically review the existing background literature relating to gender differences in alaryngeal speech, and how these perceptions of alaryngeal speech are formed. With this research as a foundation, the next objective was to perform an experimental study where naïve listeners would judge TE speech samples based on their identification as a given gender. Ultimately, the goal was to determine both how perceptions of TE speech may be influenced by knowledge of the speaker's gender, and how these perceptions may be clinically relevant in the practice of Speech-Language Pathology with those who have undergone laryngectomy.

Study 1: Critical Review

Methods

Search Strategy

Computerized databases including Scholars Portal, Scitation, and ASHA publications were searched using the following search strategy: [(tracheoesophageal) AND (gender) OR (female), AND (speech) OR (speakers)]. Reference lists accompanying several articles were also used to obtain relevant articles.

Selection Criteria

For this review, studies were required to include data on female and male tracheoesophageal speakers for comparisons based on gender. Perceptual and/or acoustic measures were required, either through others' ratings (eg. naïve listeners) or self-ratings.

Data Collection

The literature search resulted in the identification and selection of four articles. These articles included two randomized block designs (Eadie & Doyle, 2004; Eadie, Doyle, Hansen, & Beaudin, 2008), a between groups case-control study (Kazi, Kiverniti, Prasad, Venkitaraman, Nutting, Clarke, Rhys-Evans, & Harrington, 2006), and a between groups post-test design (Searl & Small, 2002).

Results

Eadie and Doyle (2004) employed a randomized block design to study 28 tracheoesophageal speakers

(22 men; 6 women) in order to determine the relationship between listener judgements of TE speech and speakers' self-ratings of quality of life. Speakers completed recordings of a reading passage and a quality of life instrument, while 15 naïve listeners with no history of hearing, speech, voice, or language difficulties rated the samples on overall speech severity, naturalness, acceptability, and pleasantness. Results were analyzed using appropriate *t*-test correlations and revealed that male TE speakers were more highly rated on all perceptual features, though quality of life scores revealed no difference by gender. Moderate correlations were found between quality of life scores and the auditory-perceptual ratings of speakers.

Strengths of this study included the high level of validity accounted for in a variety of domains. This study included a large speaker sample size (28) relative to the number of laryngectomees in the population at large. Additionally, direct magnitude estimation scaling with a modulus was used appropriately to determine select speech samples to be included. Speaker factors were controlled including native language, time post-laryngectomy, and speech stimuli through the use of the Fairbanks' Rainbow Passage (Fairbanks, 1960). These speech samples were carefully recorded with controls for sampling rate and background noise. Controls were also ensured for both speaker and dimension order effects so that a relative equality across samples was presented to listeners. The detailed description of methods allows for replication of this study. Steps were also taken to ensure that reliability of listeners' perceptions was high, as the modulus was repeated every six stimuli. As well, 25% of the speech samples were rated twice by listeners for reliability. Interrater reliability was acceptable as revealed by an appropriate Cronbach alpha. It is not clear how much information listeners were given about speaker demographics when they made their judgements. Additionally, there were a small number of female compared to male TE speakers recorded, thus it is important to question the generalizability of the gender specific data. Additionally, the gender of those who served as listeners in this study is unclear.

Overall, this study provides highly suggestive evidence that male TE speakers are more preferred by listeners than female TE speakers based on perceptual ratings, and that there is a moderate correlation between these perceptual ratings and self-reported quality of life scores.

Eadie et al. (2008) used a randomized block design to investigate how the gender of TE speakers would

impact listener preference and acceptability ratings, how well listeners could identify speaker gender, and the acoustic characteristics that may have influenced such judgments. For the purpose of this review, the focus will be on how the gender of TE speakers would impact listener preference and acceptability ratings. In this study, audio recordings from 12 TE speakers (6 men; 6 women) were chosen, and 20 naïve listeners (17 females; 3 males) with normal hearing thresholds rated the severity of the samples based on standard reference stimuli. Appropriate methods and data analysis were employed, and results showed that with gender known, female TE speakers were judged to be less acceptable and less natural than male TE speakers.

The methods used in this study were appropriate to the objectives of the study, thus demonstrating construct validity. Speech recordings were made and controls for reading passage, background noise and sampling rate were considered. Speaker demographics and time post-laryngectomy were both specified. The methods in this study were described in detail such that replication would be possible. Intra- and interrater reliability were both calculated using appropriate Pearson correlation coefficients, with adequate coefficient values obtained. An appropriate 2 x 2 mixed-model analysis of variance was used to determine how gender impacted ratings of naturalness and acceptability.

A possible limitation of this study is that listeners were not given a definition of “preference” in the first paired comparison, so it is unknown what basis on which their judgements were made. In term of nuisance variables, it is possible that the large range of months postlaryngectomy may have played a role in the competency of speakers’ communication with the TE prosthesis, thereby influencing the quality of speaker samples. Additionally, it is impossible to know how the imbalance of gender across listeners impacted the results.

Relative to the question at hand, this study presents a compelling argument that the listener’s knowledge of a TE speaker’s gender penalizes the ratings of only female TE speakers.

Kazi et al. (2006) assessed 10 female TE speakers aged 41-81, 10 male TE speakers aged 40-79, and a control group of 10 female laryngeal speakers aged 35-72 in a between-groups case control study designed to analyze female TE speech in a multidimensional way. Acoustic data were collected and perceptual evaluations were completed by expert raters through general impressions and the use of a

voice quality scale. Each speaker completed quality of life and voice handicap surveys, both of which are commonly used in voice studies. Relevant findings to the research question at hand were that expert raters did not judge male and female TE speakers to be significantly different, however female TE speakers self-reported a significant-to-severe voice handicap when compared to the self-reports of male TE speakers.

Strengths of this study include its clear method, where protocols were clarified such that a replication of the procedures could be completed. Blinding procedures were used such that the raters did not know the identity of the speakers. The authors also used appropriate measures of inter- and intra-rater reliability.

Still, there are several concerns with this study. Firstly, the authors used a control group that was not representative of the TE speaker groups as they were poorly matched for age and sex. The reason for this decision was unclear. Additionally, the male and female TE speaker samples’ respective time since laryngectomy was unclear. Secondly, many of this study’s methods and data analyses were problematic. In the acoustic analyses, female TE speakers showed a more significant deviation from the control group of female laryngeal speakers when compared to male TE speakers; however the authors do not explain their methods of data analysis in this domain, making it difficult to credit. In the study’s perceptual ratings, the use of *only* two, *experienced* listeners (otolaryngologists) makes it impossible to know how the general population would perceive these samples, reducing the generalizability of these results. While quality of life questionnaires showed no significant differences between male and female laryngectomees, the authors note that “speech” and “appearance” were considered the most important issues for females. Unfortunately, there were no comparable data provided for male laryngectomees.

This article provides an equivocal statement that female TE speakers’ voice handicap self-ratings are generally worse when compared to men. The article suggests that there may be some degree of social penalty for female TE speakers.

Searl & Small (2002) designed a between groups post-test study to examine how naïve listeners would rate male (n=6) and female (n=6) TE speech samples from experienced speakers judged to be above average to excellent TE speech. A group of 25 naïve female listeners judged the samples using a gender rating scale (masculinity-femininity). Appropriate

chi-square and Spearman rank order correlations revealed that listeners were able to reliably identify speaker gender, and that while results were statistically significant for both male and female TE speakers, the correlation was higher for male TE speakers, indicating that female TE speakers' gender was more often identified as masculine or ambiguous.

This study was strong in its internal validity, controlling for the intelligibility of speakers, type of TE prosthesis, and method of stoma occlusion. As well, sufficient details about the recordings of the speech samples and the reading passage used were provided for possible replication. Adequate inter- and intra-rater reliability measures were reported. The scale used for judgements was well described and justified, and the presentation of speaker samples and rating scales were meticulously controlled and randomized to eliminate order effects.

Concerns of this study include its use of solely female listeners, as well as the fact that all of the listeners were enrolled in communications disorders studies, despite not having any contact with TE speakers. It is impossible to know how the results may have been different had both genders been represented in the listener group, or how the listeners' educational orientation may have influenced the ratings. The authors do acknowledge these limitations.

The study provides a compelling argument that female TE speakers are more at risk than their male counterparts for being perceived as masculine or ambiguous in gender. Still, it does not rule out the possibility that male TE speakers also have the potential to be rated as gender-ambiguous.

Discussion

In sum, reviews of the above studies suggest that a gender bias may exist when perceptions of TE speech are examined. Overall, when the perceptual data are viewed together, it is clear that female TE speech is typically perceived as more masculine, less acceptable, and less preferred than its male counterpart. These results are also reflected through female TE speakers' self reports. Additionally, male TE speakers may be at a mild risk of being identified as ambiguous in gender. To further explore this question, the remainder of this paper will discuss the results of an experimental study on auditory-perception of TE speech in a condition of gender deception as compared to a condition of true gender

identification prior to having listeners make their judgments.

Study 2: Pilot Study

In the following study, "speech acceptability" and "listener comfort" were examined in the context of TE speech perceptions. It was hypothesized that perceptions of TE speech may differ based on the gender assignment associated with each sample.

Methods

Speaker Samples

High quality audio recordings of the Rainbow Passage (Fairbanks, 1960) from 6 male and 6 female TE speakers were selected from a library of TE speech samples. These samples were judged to be highly intelligible and not obviously associated with either gender. All speakers who provided the samples used TE speech as their primary method of communication, had received radiation treatment, and were native English speakers. Speakers were excluded if they reported any history of other medical conditions that might affect speech, language, or hearing.

Listeners

Sixteen (8 females and 8 males) undergraduate and graduate students (mean age of 23 years; 9 months) were recruited as listeners in this study. All of the listeners were considered naïve to voice disorders and alaryngeal speech as they did not have any formal exposure or education in this area. Listeners were native English speakers and had no history of speech, language, or hearing concerns. Permission to conduct this study was formally granted by the Research Ethics Board at the University of Western Ontario.

Speech Stimuli

As noted, speakers provided recordings of the first paragraph of the Rainbow Passage, a passage used for its representativeness of the American English language (Fairbanks, 1960). Speech samples were recorded using a headset microphone (Shure SM10a; Shure Incorporated, Niles, IL) and a digital minidisk (MD) research quality recorder (Sony MZ-R55; Sony Corp., New York, NY) or a digital audiotape portable recorder (Sony PCM-M1) in a quiet setting. All recordings were digital originals recorded at a sampling rate of 48 kHz.

Digital recordings were transferred to a personal computer and saved as WAV files using the acoustic software *Audacity*. Each sample was edited to extract

the second sentence of the abovementioned passage, “The rainbow is a division of white light into many beautiful colours.” Aside from the samples being edited to include this sentence exclusively, the only other editing that took place was the addition of 3 seconds of silence on either side of the sample sentence, to ensure that listeners could easily attend to the entire speech sample during the listening tasks.

Procedures

Naïve listeners were recruited by visiting an undergraduate linguistics class at Western University, wherein the study was described and participants were able to volunteer anonymously. Other listeners were recruited by word of mouth.

When listeners came in individually for their first listening session, they were informed that the samples to which they would listen were abnormal voice samples. As such, each listener was asked to listen to 4 TE speech samples that were not associated with the experiment itself. The purpose of this task was to familiarize and orient listeners with the types of unique voices on which they would soon be making judgements. While the gender of all of these samples was not revealed, listeners were allowed to listen to these samples as many times as they desired before beginning the experiment. Once the listener was ready to begin the rating task, he/she was presented with a randomized list of the 12 TE voice samples and a rating scale (either for “speech acceptability” or “listener comfort”). The listener was asked to read the definition of the feature on which he/she would be making judgements, and was then asked to systematically play the list of voice samples and rate each one independent of one another. The scale of “speech acceptability” asked the listener to rate a voice based on “pitch, rate, understandability, and voice quality. In other words, is the voice pleasing to listen to or does it cause...some discomfort as a listener”. Alternatively, the scale of “listener comfort” asked the listener “how comfortable would you feel listening to the person’s speech in a social situation?”. Listeners could play each individual sample as many times as desired before making a rating on the scale.

In this first listening condition, the gender of each sample was indicated on the rating form, however the gender indicated was, in actuality, the opposite of speaker’s true gender. Listeners were then asked to complete a second rating form (the alternative to the one they had just completed, either “speech acceptability” or “listener comfort”), and were provided with the same samples in a different, randomized order. The delegation of these scales

was counterbalanced such that 8 listeners received the “speech acceptability” scale prior to receiving the “listener comfort” scale, and the other 8 received the “listener comfort” scale prior to the “speech acceptability” scale. In this way, controls were set up for order effects of TE speaker samples, as well as rating scales. The possibility that the completion of one scale might influence the completion of the other was reduced.

In the second listening session, which occurred 7-14 days after the first, listeners were brought in to follow the same procedures as detailed above. Listeners were presented with the reverse order of rating scales that they received in the first session. In this condition, the true gender of each speaker sample was indicated on the rating form. In both conditions, listeners were led to believe that the gender indicated on the rating form was accurate. The purpose of using deception in the first listening condition was to ascertain whether or not listeners would rate speaker samples differently based on a prescribed gender. In both conditions, only the examiner was aware of the true gender of the voice sample being presented.

Reliability

At the end of both listening sessions, in order to ensure measures of internal validity, 4 voice samples (25%) were duplicated from the 12 original voice samples. Listeners were asked to rate them on both “speech acceptability” and “listener comfort” scales, just as they did for the previous 12 samples, and they were not informed that they were duplicate samples. These measures were compared to how the listener previously rated those particular samples, to evaluate the consistency of rating. Pearson correlation coefficients were calculated to determine how the ratings of these 4 voice samples correlated across “speech acceptability” and “listener comfort” scales across the two listening sessions. When these correlation coefficients were averaged, the results showed the following correlation coefficients for each voice sample: $r = 0.48$, $r = 0.61$, $r = 0.52$, and $r = 0.52$, demonstrating strong correlations between voice sample ratings and, thus high intra-rater reliability. When raw data were analyzed, 75% of raters in their second judgement always rated the same sample within 10 scaled points of the original rating.

Data Analysis

Relationships among listener ratings, as well as relationships between measures of “speech acceptability” and “listener comfort” were revealed through the calculation of Pearson correlation

coefficients. The relationships between the gender opposite and gender known conditions were calculated using Pearson correlation coefficients and independent *t*-tests. These analyses were completed for all listeners, and separately for male and female listener groups. A predetermined level of statistical significance ($P < .05$) was used for all analyses.

Results

Pearson correlation coefficients showed high correlations between mean ratings for “listener comfort” and “speech acceptability” within sessions, with $r=0.97$ resulting in both sessions 1 and 2. When “listener comfort” and “speech acceptability” measures were compared across sessions 1 and 2, an independent *t*-test showed a significant difference in “speech acceptability” ratings from session 1 to session 2, where $t=0.05$. When further analyzed by gender, the statistic that was skewing the overall results in one direction was for the female speakers’ “speech acceptability” ratings from session 1 to session 2, where a significant difference was seen between the two sessions. This independent *t*-test resulted in a value of 0.02, suggesting that the female speakers were rated differently from session 1 to session 2, where the differing variable was the indicated gender. A similar effect was seen in the “listener comfort” domain, where the independent *t*-test resulted in a value of 0.08, thus approaching significance. Thus, females were judged to be more unacceptable and to a lesser degree more uncomfortable to listen to when known to be female, as compared to when thought to be male (Figure 1). When male speakers were analyzed independent of their female counterparts, it was clear that there were no significant differences between how they were rated on “speech acceptability” or “listener comfort” scales from session 1 to session 2. When *listeners* were analyzed by gender, there were no significant findings with regard to how one gender *rated* TE speakers when compared to the other gender.

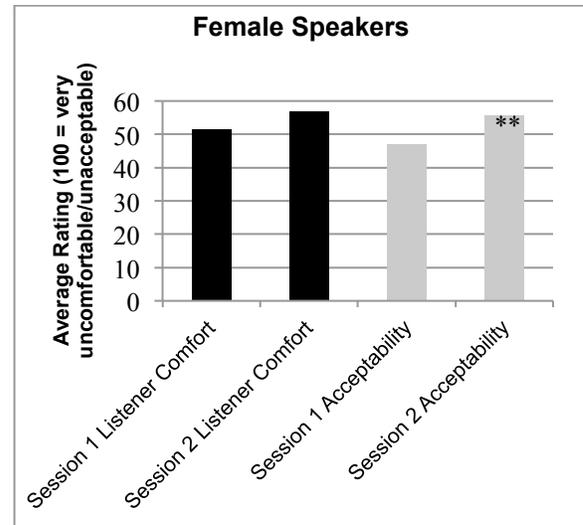


Figure 1. The mean ratings of all female TE speakers across Sessions 1 & 2 for “speech acceptability” and “listener comfort”. In the first session, listeners were led to believe that the speakers were male, while in the second session speakers were correctly identified as female.

**** indicates statistical significance in scores between session 1 and session 2.**

Discussion

This paper sought to answer the question of how a listener’s knowledge of a TE speaker’s gender would influence judgements of the auditory-perceptual features of “listener comfort” and “speech acceptability”.

Several variables were actively controlled in this study. Firstly, in anticipation of the deception component of this experiment, speaker samples were chosen based on the experimenter’s subjectivity that the samples sounded ambiguous, and were not obviously associated with either gender. Had the samples sounded obviously male or female, the deception aspect of the experiment may have become apparent to the listener. In order to control for variables that might influence perceptual judgements, several steps were taken. Firstly, 4 randomized lists of speaker samples were developed so that listeners were always presented with uniquely ordered list of samples on which to complete their ratings. Secondly, upon each listening session, listeners were presented with a unique order of rating scales (for example, if they completed ratings of “listener comfort” prior to “speech acceptability” during the first session, the opposite would occur during the second session). Efforts were made to evaluate listener judgements in the context of speaker gender

and, thus, listeners were directed to the gender of the speaker samples prior to each listening session.

It is important to consider what the “speech acceptability” ratings obtained in this study may suggest. As described in previous literature by Eadie et al., judgements of “speech acceptability” encourage the listener to identify “speech acceptability” as it relates to their own personal beliefs about deviation from a normal signal (2008). Thus, it is possible that gender is a critical factor in determining this, as listeners very likely have preconceived templates of how men and women should each sound (Eadie et al., 2008; Kreiman et al., 1993). It is interesting to note that penalizing judgements were more apparent for female TE speakers in the domain of “speech acceptability” as compared to “listener comfort”. This suggests that female TE speakers are more penalized in the domain of overall voice quality including pitch, rate, understandability, and pleasantness of listening, causing some discomfort to the listener, when compared to “listener comfort”, or how comfortable listeners would be listening to a particular speaker in a social setting. By definition, there is some degree of “comfort” inherently considered in the “speech acceptability” ratings. However, there were differences in how these two constructs of “speech acceptability” and “listener comfort” were rated across listening sessions. It appears that when listeners are making overall judgements about a speaker’s voice, comfort may be one component of the score, but more quantitative aspects of voice may further determine overall scores. This difference between the “speech acceptability” scores across sessions was pronounced for the female TE speaker population, suggesting that when female TE speakers were believed to be male, listeners rated the samples slightly more acceptable than when the samples were known to be female. The same trend was seen for the construct of “listener comfort”, however, this difference was approaching, but not reaching, a level of significance. When male TE speakers were analyzed independent of their female counterparts, there were no significant differences in ratings of “speech acceptability” or “listener comfort” from session 1 to session 2. It can be concluded that female TE speakers may face greater penalization than male TE speakers when it comes to listener judgements, more significantly for the feature of “speech acceptability”.

When listener data were analyzed and broken down by gender, the finding was that neither male nor female listeners rated a particular speaker gender significantly worse than the other. This knowledge is

important as it allows the listener preferences to be generalized to the population at large.

Clinical Implications

The present study sought to determine whether or not knowledge of a TE speaker’s gender would influence the perceptual ratings assigned by naive listeners. Based on the data gathered, it is apparent that listeners rate female TE speaker samples to be more unacceptable and more uncomfortable when the samples are known to be female speakers. While the underlying reasons for this finding are unknown, it does seem likely that a multitude of factors are involved when a listener judges a TE speaker’s voice, including voice quality and comfort level associated with listening to a particular voice. It will be important for future research to look at acoustic information in conjunction with perceptual data for male and female TE speaker samples, to determine the specific parameters that may affect listener judgement.

In terms of clinical application, these results are important to consider in the context of pre- and postoperative counselling for female laryngectomees, to ensure the most optimal rehabilitation. As well, the need for a more feminine-sounding method of voice restoration continues to exist, but is challenging from both anatomical and technological standpoints.

Acknowledgements

The author would like to thank Dr. Philip Doyle, of the University of Western Ontario, for collecting voice samples, providing the measurement templates used, and for supporting and supervising the study; Steven R. Cox, PhD (c) for collecting voice samples; Cassandra Leblanc and Sebastiano Failla for measuring data and preparing voice samples, respectively; and the Voice Production and Perception Laboratory at the University of Western Ontario.

References

- Canadian Cancer Society. (2013). Canadian Cancer Society’s Advisory Committee on Cancer Statistics. In *Canadian Cancer Statistics 2013*. Toronto, Ontario.
- Doyle, PC, and Eadie, TL. (2005). Scaling of pleasantness and acceptability in tracheoesophageal (TE) speakers. *Journal of Voice*, 19(3): 373-383.

- Eadie, T.L., and Doyle, P.C. (2004). Auditory-perceptual scaling and quality of life in tracheoesophageal speakers. *The Laryngoscope*, 114: 753-759.
- Eadie, T.L., Day, A.M.B., Sawin, D.E., Lamvik, K., and Doyle, P.C. (2012). Auditory-perceptual speech outcomes and quality of life after total laryngectomy. *Otolaryngology – Head and Neck Cancer*, 148(1): 82-88.
- Eadie, T.L., Doyle, P.C., Hansen, K., Beaudin, P.G. (2008). Influence of speaker gender on listener judgements of tracheoesophageal speech. *Journal of Voice*, 22(1): 43-57.
- Fairbanks, G. (1960). *Voice and Articulation Drillbook*, 2nd ed. New York: Harper & Row.
- Kazi, R., Kiverniti, E., Prasad, V., Venkitaraman, R., Nutting, C.M., Clarke, P., Rhys-Evans, P., Harrington, K.J. (2006). Multidimensional assessment of female tracheoesophageal prosthetic speech. *Clinical Otolaryngology*, 31: 511-517.
- 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: 21-40.
- Pindzola, R.H., & Cain, B.H. (1988). Acceptability Ratings of Tracheoesophageal Speech. *Laryngoscope*, 98: 394-397.
- Robbins J, Fisher HB, Blom EC, Singer MI. (1984). A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production. *Journal of Speech and Hearing Disorders*, 49(2): 202.
- Searl, J.P., and Small, L.H. (2002). Gender and masculinity-femininity ratings of tracheoesophageal speech. *Journal of Communication Disorders*, 35(2002): 407-420.
- Singer, MI, and Blom, ED. (1980). An endoscopic technique for restoration of voice after laryngectomy. *Ann Otolaryngology*, 89: 529-33.
- World Health Organization. (1980). *International classification of impairments, disabilities, and handicaps*. Geneva: Author.