applied sciences

MDPI

Article

Assessing Gender Bias in Auditory-Perceptual Ratings of Tracheoesophageal Speakers

Jenna L. Bucci [1,2], Nedeljko Jovanovic [1] and Philip C. Doyle [3,*]

[1] Voice Production and Perception Laboratory, Department of Otolaryngology Head and Neck Surgery, University of Western Ontario, London, ON N6A 3K7, Canada; nedjovanovic93@gmail.com (N.J.)
[2] Royal Victoria Regional Health Centre, Barrie, ON L4M 6M2, Canada
[3] Department of Otolaryngology Head and Neck Surgery, Division of Laryngology, Stanford University School of Medicine, Stanford, CA 94305, USA
* Correspondence: pdoyle2@stanford.edu

Abstract: Objective: This study examined the relationship between gender and auditory-perceptual evaluation of tracheoesophageal (TE) speech. **Method:** We collected auditory-perceptual judgments of two features, *speech acceptability* and *listener comfort*, from normal-hearing young adult listeners (n = 16) who were naïve to TE speech. Auditory-perceptual judgments were made for 12 TE speakers (6 men and 6 women) on two occasions separated by between 7 and 14 days. During the first session, listeners were deceived about the gender of the voice samples presented, and in the second session, listeners were informed of the true gender of the voice samples. **Results:** The findings 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 when gender is known. **Conclusions:** These data provide insights into the potential influence of speaker gender on listener judgments of TE speech and the impact that such factors may have on communication. Our data indicate that listeners rate female TE speaker samples as less acceptable and less comfortable to listen to when the samples are known to be female speakers.

Keywords: tracheoesophageal speech; auditory-perceptual evaluation; gender; laryngectomy; perceptual psychophysics

Citation: Bucci, J.L.; Jovanovic, N.; Doyle, P.C. Assessing Gender Bias in Auditory-Perceptual Ratings of Tracheoesophageal Speakers. *Appl. Sci.* **2024**, *14*, 3447. https://doi.org/10.3390/app14083447

Academic Editor: Douglas O'Shaughnessy

Received: 13 March 2024
Revised: 8 April 2024
Accepted: 10 April 2024
Published: 19 April 2024

1. Introduction

Following a diagnosis of laryngeal cancer, an individual may undergo surgical removal of the entire larynx, in what is termed "total laryngectomy," to eradicate the malignancy. Following total laryngectomy, the individual is left without a natural voice, which necessitates postsurgical speech and voice rehabilitation in order to regain verbal communication. Fortunately, several options exist for the acquisition of postlaryngectomy "alaryngeal" voice and speech. In many instances today, a surgical–prosthetic method of voice rehabilitation referred to as tracheoesophageal (TE) voice restoration [1] is employed for voice restoration. The TE voice restoration method has been a widely used and well recognized postlaryngectomy voice and speech rehabilitation option for more than 40 years.

1.1. Tracheoesophageal Puncture Voice Restoration

Briefly, TE voice restoration requires creating a small, controlled puncture in the midline between the common tracheal and esophageal walls; following this procedure, a one-way valved voice prosthesis is inserted into the puncture. The TE voice prosthesis does not function as a new voicing source; rather, the prosthesis allows air from the lungs to be diverted into the esophagus during exhalation. When this air is expelled, the tissues that serve as the new voicing source are comprised of the musculature of the upper esophagus and the lower pharynx [1]. The TE voice source that is generated then moves upwards

into the vocal tract (i.e., the pharynx, oral, and nasal cavities), where it is modulated by the articulators to produce speech.

Despite the success of restoring voice and speech to those who have undergone total laryngectomy, TE speech is nevertheless distinctly different from the normal laryngeal voice; these changes have been documented in several acoustic studies [2–4]. Despite the restoration of voice, TE speech remains altered in its overall perceptual voice quality to the listener. TE speech is characterized by variable levels of substantial aperiodicity (noise) secondary to the aerodynamic influence on the new voice source that is comprised of muscular tissues of the pharynx and esophagus [5]. While much individual variation in the auditory-perceptual quality of TE voices exists, it is clear when listening to TE speakers that one's voice is quite unlike that of a normal laryngeal speaker [6]. Because judgments of a speaker's voice quality are based on a construct that fundamentally evolves from the listener's perception, auditory-perceptual ratings are considered the gold standard for voice description in the areas of both clinical and experimental research [7,8].

1.2. Gender Considerations

Historically, studies of TE speakers have focused on men due to the relative predominance of males in those diagnosed with laryngeal cancer. However, it is increasingly important to consider the female perspective, given the increasing proportional number of women who today are diagnosed with laryngeal cancer and may subsequently require total laryngectomy. In fact, recent data suggest that the ratio of men to women who are diagnosed with laryngeal cancer has now fallen to less than 4:1 [9,10]. Additionally, because TE speech is widely employed as a primary method of speech rehabilitation in North America, the increase in female TE speakers will also increase [10–12].

Considering the dramatic change in voice quality following laryngectomy and TE puncture (i.e., lowered pitch, increased aperiodicity in voice signal, etc.), speaker gender becomes of considerable importance. In fact, published data indicate that women who use TE speech experience a considerable deviation from their prelaryngectomy voice [13]. Specifically, female TE speakers experience a significantly lower vocal pitch that that of their prelaryngectomy voice, a variably rough voice quality and, consequently, the loss of a feminine-sounding voice [3]. Further, female TE speakers commonly report being misidentified as male when speaking on the telephone [13] or where other visual information is unavailable to a communication partner. Therefore, it is possible that, when compared to men, women may experience more negative judgments of their TE speech because of its more significant deviation from the expected qualities of a normal female voice [14–18]. Reductions in voice quality for women who have undergone laryngectomy have been shown to be socially penalizing and stigmatizing [19].

Although the relative impact of changes specific to females who undergo laryngectomy and use TE speech have been documented in part, at present, little is known about how a listener's awareness of a TE speaker's gender potentially affects their judgments of a TE speaker's voice [4,13]. Previous studies have confirmed that auditory-perceptual data serve as a valuable complement to patient-reported outcomes (PROs) following laryngectomy [19]. However, at present, there have been no explorations of potential gender bias in any group of postlaryngectomy speakers. Given the popularity of TE puncture voice restoration today, such questions have clear merit relative to understanding rehabilitation outcomes. Therefore, auditory-perceptual evaluation can be valuable in helping to understand the potentially different communication experiences of both male and female TE speakers in terms of their desire to share, interact socially, and participate in life activities [20–25]. When such considerations are coupled with the potential for social penalty, issues specific to voice quality become of even greater importance [25–27]. Clinically, information of this type would also appear important as it may serve to shape the counseling that is provided during both pre- and postlaryngectomy speech/voice rehabilitation efforts. Consequently, it was hoped that information presented in the present study could empirically identify the presence or absence of a gender bias in perceptions of

TE speakers. Ultimately, the goal was to determine how perceptions of TE speech may be influenced by the listener's knowledge of a speaker's gender.

2. Methods

2.1. Participant Speakers

Twelve adults (six men and six women) who had undergone total laryngectomy and TE puncture voice restoration served as participant speakers for this investigation. The TE speakers who participated in this study ranged in age from 58 to 71 years of age. These participants had used TE speech for a minimum period of one year and were judged by a highly experienced clinician to be excellent examples of this mode of alaryngeal speech. All of the speakers exhibited excellent speech intelligibility, self-reported that they were in good general health, and confirmed that they remained communicatively active in both their vocational and avocational endeavors.

All speakers who provided samples used TE speech as their primary method of communication, had received radiation treatment postoperatively, and were native English speakers. Speaker samples were excluded if the individual reported any history of perioperative complications following laryngectomy and TE puncture or identified other medical conditions that might affect speech, language, or hearing, including oral or pharyngeal resection, cancer recurrence or a second primary cancer, or chronic obstructive pulmonary disease, asthma, persistent swallowing difficulties, or neurological disease.

2.2. Participant Listeners

Sixteen normal-hearing young adults (eight females and eight males), ranging in age from 22 to 27 years of age (mean age = 23; 9), who were enrolled as either undergraduate or graduate students at a single institution, were recruited as listeners in this study. All listeners were considered naïve to voice disorders and alaryngeal speech as they did not have any formal exposure to postlaryngectomy speech options or any education in the area of voice disorders, voice, or speech disorders associated with head and neck cancer. All listeners were native English speakers, and none had reported any 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 (#104645).

2.3. Speech Stimuli

High-quality digital audio recordings of the Rainbow Passage [28] were obtained for all 12 TE speakers from a large archival library of TE speech samples. These samples were judged by two independent professional Speech-Language Pathologists (SLPs) to not be obviously associated with either gender. Thus, these samples were selected to be representative of a larger group of TE speakers. All original speech samples were recorded using a headset microphone (Shure SM10a; Shure Incorporated, Niles, IL, USA) and either a digital minidisk (MD) research-quality recorder (Sony MZ-R55; Sony Corp., New York, NY, USA) or a digital audiotape portable recorder (Sony PCM-M1), with all recordings being obtained in a quiet experimental setting, free of ambient noise. 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 (version 2.0.6, Pittsburgh, PA, USA).* Each sample was edited to extract the second sentence of the Rainbow Passage, *"The rainbow is a division of white light into many beautiful colors".* Aside from the samples being edited to include this sentence exclusively, the only other editing that took place was the addition of 3 s of silence on either side of the sample sentence, to ensure that listeners could easily attend to the entire speech sample during the experimental listening tasks.

2.4. Orientation and Listening Procedure

As an a priori requirement of the experimental procedure, all listeners were required to participate in two listening sessions. Upon arrival for the first listening session, participant listeners were informed that the speech samples to which they would listen were abnormal voice samples where the quality of the voice was reduced from normal expectation. Each listener was then asked to listen to four TE speech samples (two males, two females) that were not part of the experimental stimuli. These four samples had been compiled into a single audio file so that each voice sample would play continuously, one after another. The purpose of this task was to familiarize and orient listeners to the unique types of voices on which they would soon be making judgements.

This exposure task was included to reduce any potential surprise related to the unusual and abnormal voice qualities that often characterize TE speech. By doing so, we believed that potential confounds related to less favorable ratings of early samples due to the unusual acoustic nature and listener adaptation could be reduced. While the gender of each of the four exposure samples was not specified, listeners were told that these samples would include both male and female voices. Listeners were allowed to listen to these familiarization samples as many times as they desired before formally beginning the experimental rating session.

Each listener had control over the computer mouse so that he/she could independently select and repeat the audio file as many times as desired. Listeners completed this task while seated comfortably at a personal computer (Dell, Round Rock, TX, USA), listening to the audio files via headphones (Sony MDRV-150); each listener was able to independently adjust the listening volume to a comfortable loudness level. Once the listener was ready to begin the experimental rating task, he/she was presented with a randomized playlist of the 16 TE voice samples (12 primary samples and 4 duplicates, for reliability) and a series of rating scales that were provided on paper in numerical order, 1 through 16. Two auditory-perceptual dimensions were assessed in the sessions, either "speech acceptability" (SA) or a dimension termed "listener comfort" (LC).

In the first listening session, the gender of each sample was indicated in the margin on the rating form for each sample by the letter "F" for female or "M" for male. However, in this first listening session, the gender indicated was in actuality the opposite of the speaker's true gender. For the purposes of data analysis, this first session was identified as "Gender Opposite". Each listener was asked to read the definition of the feature on which he/she would be making their judgments (either SA or LC) and was then asked to systematically play the list of randomized voice samples and rate each one independent of one another in a sequential manner.

2.5. Auditory-Perceptual Rating Task

For ratings of the auditory-perceptual feature SA, listeners were asked to rate a voice based on "The 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?" [29]. Alternatively, ratings of the auditory-perceptual feature LC required the listener to rate samples based on "How comfortable would you feel listening to the person's speech in a social situation?" [30]. Each audio file was individually labeled as "Sample 1", "Sample 2" . . . "Sample 16", etc., and, accordingly, each rating scale provided was labeled with the same sample number. Upon listening to each consecutive sample, the listener bisected the line of the visual analog rating scale at the point that they believed best represented their judgment of that sample for each dimension; this procedure was followed identically for each of the randomized set of samples presented. Listeners could play each individual sample as many times as desired before making their judgment using the rating scale, but they were instructed that once a rating was made, it could not be altered later, nor could they return to past samples.

Regardless of whether SA or LC was being evaluated, each rating scale was comprised of a solid line measuring 100 mm and listeners were shown examples of how the scale was to be used. The appearance of the rating scales used were consistent with those of the Consensus Auditory-Perceptual Evaluation—Voice (CAPE-V) that is used in clinical studies and research associated with laryngeal-based voice disorders [31]. Below each scale, descriptive indicators of "mild", "moderate", and "profound" were provided at approximately 25 mm, 55 mm, and 85 mm, respectively. Thus, as scores moved from left-to-right on the scale (increased) for both SA and LC, listener judgments became increasingly more favorable for either dimension; that is, higher scores indicated a more positive judgment by a listener.

Once the first auditory-perceptual dimension was rated, listeners were provided with a short break of approximately 15–20 min and then were asked to complete a second rating session, which addressed the second dimension that they had not yet completed (either SA or LC). During the rating task for the second dimension, listeners were provided with the same samples in a new randomized order. The delegation of these scales (SA and LC) was counterbalanced so that 8 of the 16 listeners rated SA first, followed by LC; conversely, the other 8 listeners rated LC first, followed by SA. Thus, controls for potential order effects related to the auditory-perceptual dimensions were considered. Once ratings of both SA and LC were completed in this session, listeners were dismissed and scheduled to return for a follow-up listening session in 7 to 14 days.

In the second experimental listening session, which occurred between 7 and 14 days after completion of the first, listeners were presented stimuli and asked to assess either SA or LC in the reverse order of that completed in the first listening session. In this follow-up listening session, however, the true gender of each speaker sample was now indicated in the margin of the rating form. For the purposes of data analysis, this second session was identified as "Gender Known". However, it should be noted that, regardless of the session (i.e., Gender Opposite or Gender Known), 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 (Gender Opposite) was to ascertain whether listeners would rate speaker samples differently based on a prescribed gender. In both conditions, only the examiners were aware of the true gender of the voice sample being presented.

2.6. Reliability

At the end of all listening sessions and to ensure measures of internal validity for listener ratings, 4 additional voice samples (25%) were duplicated from the 12 original voice samples. Thus, Samples 13, 14, 15, and 16 represented duplicate samples selected from those played in the first 12 randomized samples (two males and two females). Listeners were asked to provide these reliability ratings for both SA and LC judgments for all experimental sessions; thus, reliability was gathered in all four rating sessions. Ratings obtained from these duplicate measures were then compared to the first rating of each sample to evaluate the consistency of ratings. When raw data from judgments of reliability samples were compared to initial judgments of the same sample, analysis revealed that 75% of all listener reliability judgments fell within $+/-10$ scaled points of the original rating, indicating judgments that were highly consistent. Less than 3% of all ratings exceeded $+/-15$ scaled points between judgments.

2.7. Data Analysis

The statistical relationship between measures of SA and LC were determined using Pearson correlation coefficients. The relationships between the Gender Opposite and Gender Known conditions were also calculated using Pearson correlation coefficients and independent *t*-tests. These analyses were first completed for all listener scores combined for the entire group of speakers and then separately for male and female listener groups. A predetermined level of statistical significance ($p < 0.05$) was used for all analyses.

3. Results

3.1. Measures of Central Tendency

Analyses of the raw auditory-perceptual data for all TE speakers, as well as that for male and female speakers by dimension (LC and SA) and across listening sessions, are provided in Table 1. As can be seen, the central tendency measures for all speakers revealed good consistency in listener scores for LC across both the Gender Opposite and Gender Known listening conditions. While similar results occurred for SA, a reduction in mean values of approximately 4% was observed in the Gender Known condition. Thus, when all speakers are considered, the auditory-perceptual scores were lower (less favorable) when the correct gender was provided to listeners. However, when these data were segmented by speaker gender, greater variability in scores for both SA and LC across the two conditions were noted.

Table 1. Measures of central tendency: mean, median, and standard deviation (SD) for listener comfort and speech acceptability stratified by session (Gender Opposite and Gender Known) and speaker gender (male and female).

Group		Overall			Male			Female		
Listening Session		**Session 1** *Gender Opposite*	**Session 2** *Gender Known*	***p*-Value**	**Session 1** *Gender Opposite*	**Session 2** *Gender Known*	***p*-Value**	**Session 1** *Gender Opposite*	**Session 2** *Gender Known*	***p*-Value**
Listener Comfort	Mean	50.3	49.9	0.878	51.8	56.7	0.200	48.7	43.1	0.081
	Median	50.2	49.0		53.6	62.8		41.1	37.0	
	SD	21.2	23.5		24.8	26.5		19.2	20.1	
Speech Acceptability	Mean	54.0	49.7	0.051	55.1	55.0	0.973	53.0	44.3	0.022
	Median	50.7	47.9		61.6	64.9		44.9	38.3	
	SD	19.9	20.4		23.7	25.5		17.6	14.1	

p-values determined using dependent samples *t*-test (significance set at $p = 0.05$).

Based on measures of central tendency, and as it can be seen in Table 1, female TE speakers were rated less favorably for both SA and LC when compared to their male counterparts. Yet it is also important to acknowledge that considerable reductions in scores were identified for female TE speakers when gender was known to the listener (i.e., Session 2 scores). In fact, in the Gender Known condition, females had scores that were lower than male scores by 13.6% and 8.7% for LC and SA, respectively.

3.2. Normality Testing

All statistical analyses were executed using SPSS statistical software (IBM, version 28.0) using two-tailed testing at an a priori probability level of 0.05 to reflect a 95% confidence interval. Moreover, outcome variables were analyzed to determine the normality of distribution and guide the selection of statistical tests. A Shapiro–Wilk test was performed to determine the normality of distribution of the two dependent variables (i.e., SA and LC). Despite some variation in the distributions, the results of the Shapiro–Wilk test indicated that data for ratings of SA ($p = 0.255$) and LC ($p = 0.142$) were normally distributed. Consequently, further statistical analyses were conducted based on parametric assumptions.

3.3. Dependent t-Tests

A dependent-samples *t*-test was performed to compare ratings of LC and SA across listening Sessions 1 (Gender Opposite) and 2 (Gender Known). This initial analysis was performed for the combined group of speakers. For LC, the results indicated that there was not a statistically significant difference ($t(11) = 0.157$, $p = 0.878$) between ratings at Session 1 (mean ± standard deviation [SD]: 50.3 ± 21.2) and Session 2 (mean ± SD: 49.7 ± 23.5). Based on the results for SA (Figure 1), the differences in mean ratings also were not found to be statistically significant across Session 1 (mean ± SD: 54.0 ± 19.9) and Session 2 (mean ± SD: 49.7 ± 20.4). The differences in mean ratings of SA did, however, closely approach statistical significance ($t(11) = 2.192$, $p = 0.051$).

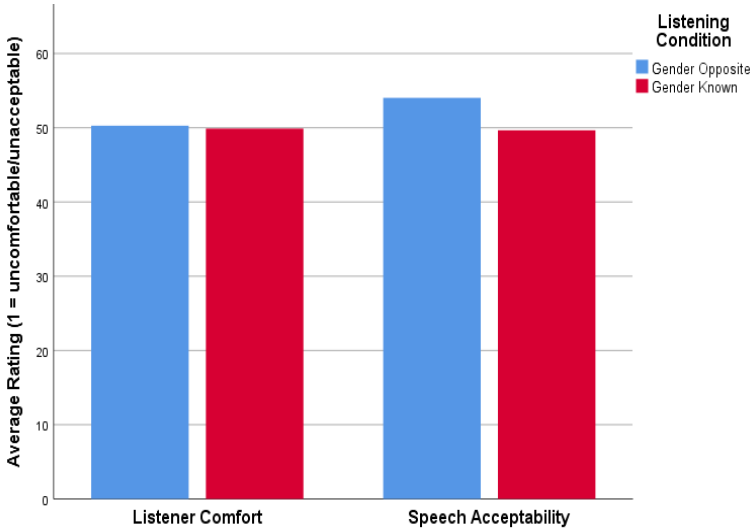

Figure 1. The mean ratings of all TE speakers across Session 1 (Gender Opposite) and Session 2 (Gender Known) for listener comfort and speech acceptability.

When our data were further analyzed by gender, there was only a statistically significant difference between mean ratings of SA for our female speakers ($t(5) = 3.291$, $p = 0.022$) between Session 1 (mean ± SD: 53.0 ± 17.6) and Session 2 (mean ± SD: 44.3 ± 14.1). Although not statistically significant ($t(5) = 2.181$, $p = 0.081$), a similar effect was observed for the differences in mean ratings of LC for females between Session 1 (mean ± SD: 48.7 ± 19.2) and Session 2 (mean ± SD: 43.1 ± 20.1)—again, a finding that approached statistical significance. Based on these findings, female speakers were judged to be more unacceptable (i.e., SA) and, to a lesser degree, more uncomfortable to listen to (i.e., LC) when known to be female (Session 2) as compared to when thought to be male (Session 1).

When male speakers were analyzed independent of their female counterparts, there were no statistically significant differences in mean ratings of LC ($t(5) = -1.475$, $p = 0.200$) between Sessions 1 (mean ± SD: 51.8 ± 24.8) and 2 (mean ± SD: 56.7 ± 26.5) or for mean ratings of SA ($t(5) = 0.035$, $p = 0.973$) across Session 1 (mean ± SD: 55.1 ± 23.7) and Session 2 (mean ± SD: 55.0 ± 25.5). A graphic representation of gender-based differences in judgments of auditory-perceptual features of LC and SA is shown in Figure 2 for female speakers and in Figure 3 for male speakers.

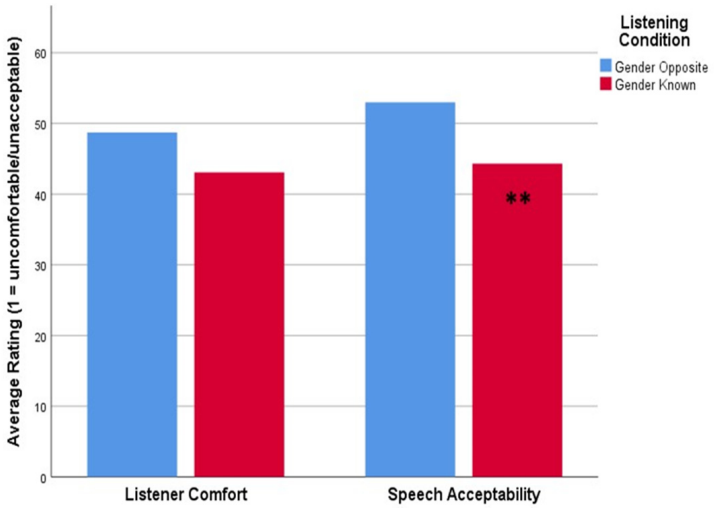

Figure 2. The mean ratings for female TE speakers across Sessions 1 (Gender Opposite) and 2 (Gender Known) for listener comfort and speech acceptability. Note: ** indicates statistically significant differences in scores between Session 1 and Session 2.

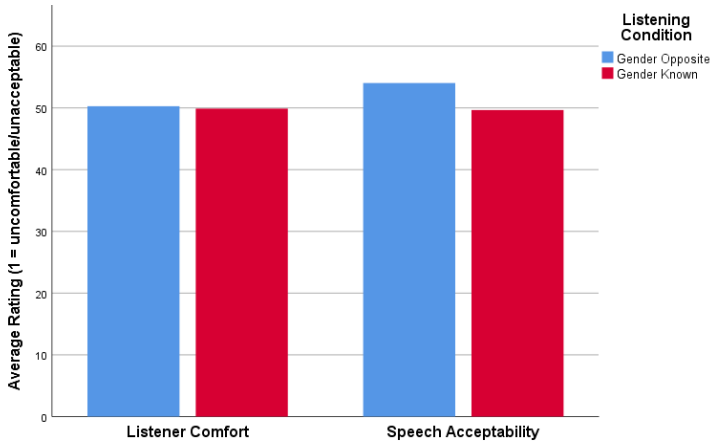

Figure 3. The mean ratings for male TE speakers across Sessions 1 (Gender Opposite) and 2 (Gender Known) for listener comfort and speech acceptability.

3.4. Correlation Analysis

To determine the relationship between ratings of LC and SA over the two listening sessions (Gender Opposite vs. Gender Known), we performed a Pearson product-moment correlation analysis. The results indicate a strong, positive correlation between ratings for both LC and SA in Session 1 (r = 0.972, n = 12, $p < 0.001$) and Session 2 (r = 0.969, n = 12, $p < 0.001$) in Figures 4 and 5, respectively.

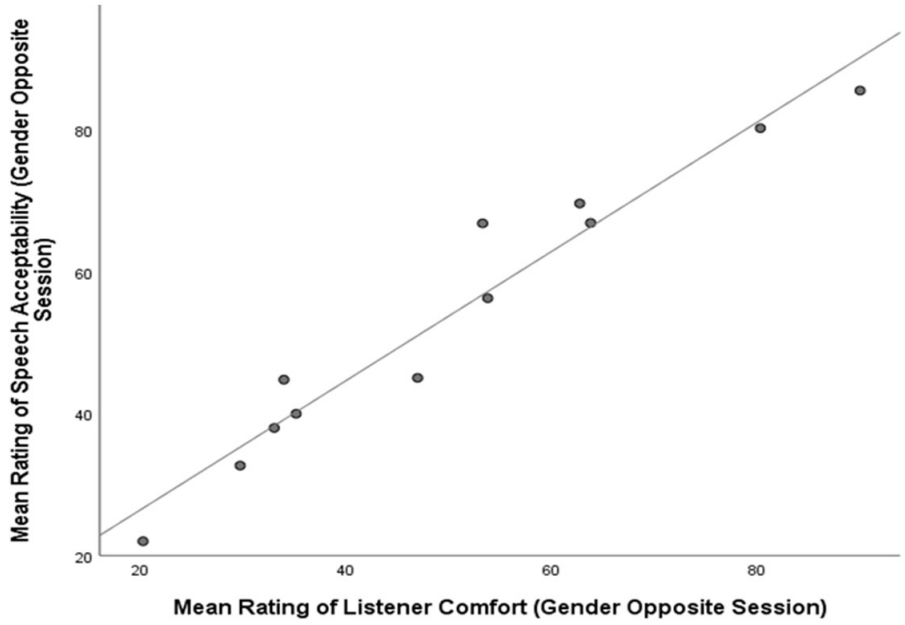

Figure 4. Relationship between mean ratings of listener comfort and speech acceptability for Session 1 (Gender Opposite).

In addition, ratings of LC were found to have a strong and positive correlation across both sessions (r = 0.927, n = 12, $p < 0.001$), with similar results obtained for ratings of SA (r = 0.942, n = 12, $p < 0.001$). A graphical representation of the relationship between ratings of LC and SA across Sessions 1 and 2 is presented in Figures 6 and 7, respectively.

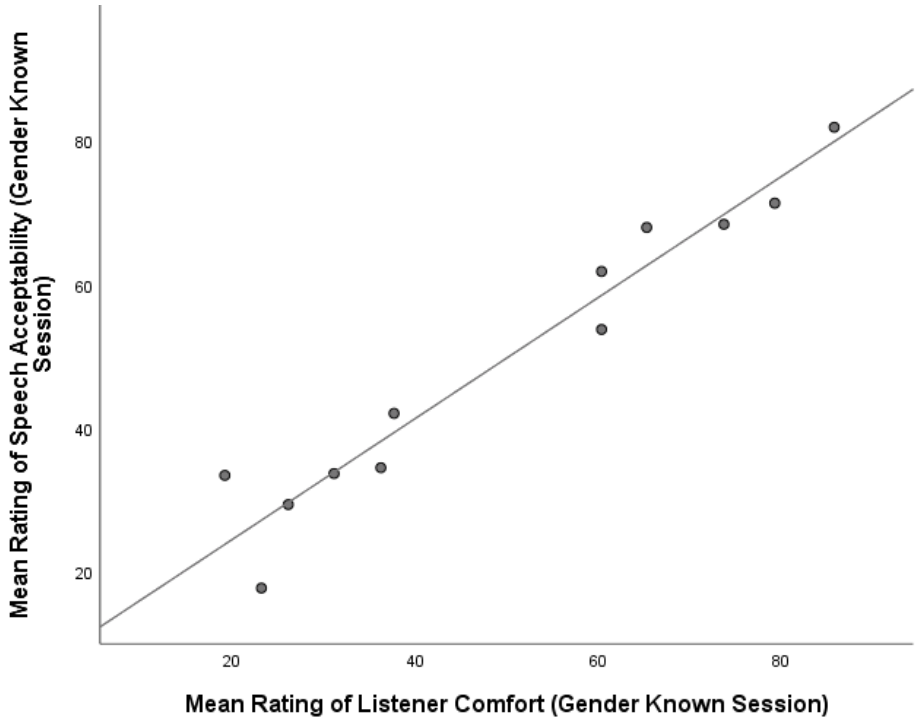

Figure 5. Relationship between mean ratings of listener comfort and speech acceptability for Session 2 (Gender Known).

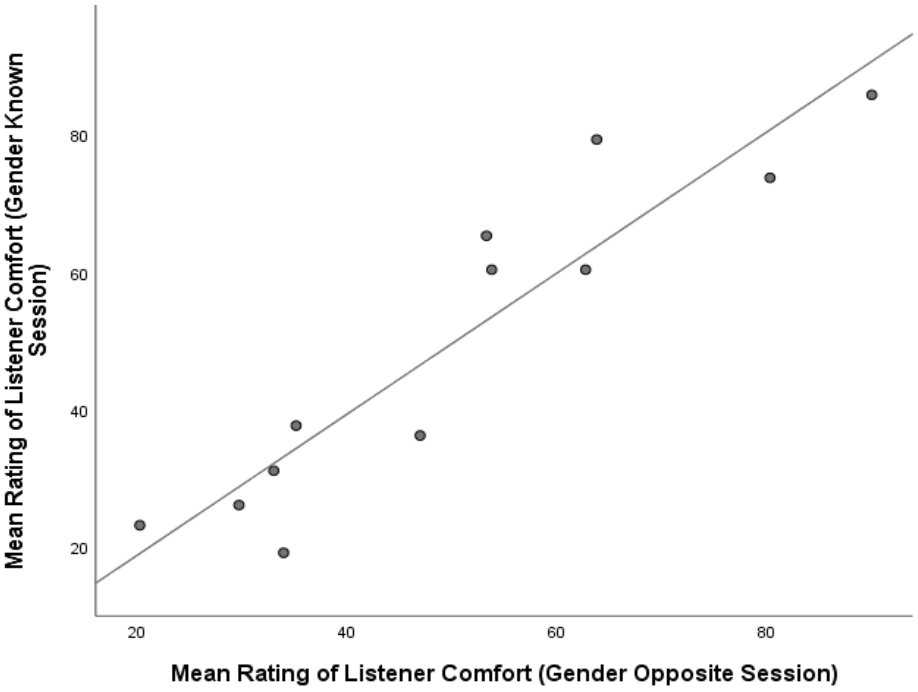

Figure 6. Relationship between mean ratings of listener comfort across session 1 (Gender Opposite) and session 2 (Gender Known).

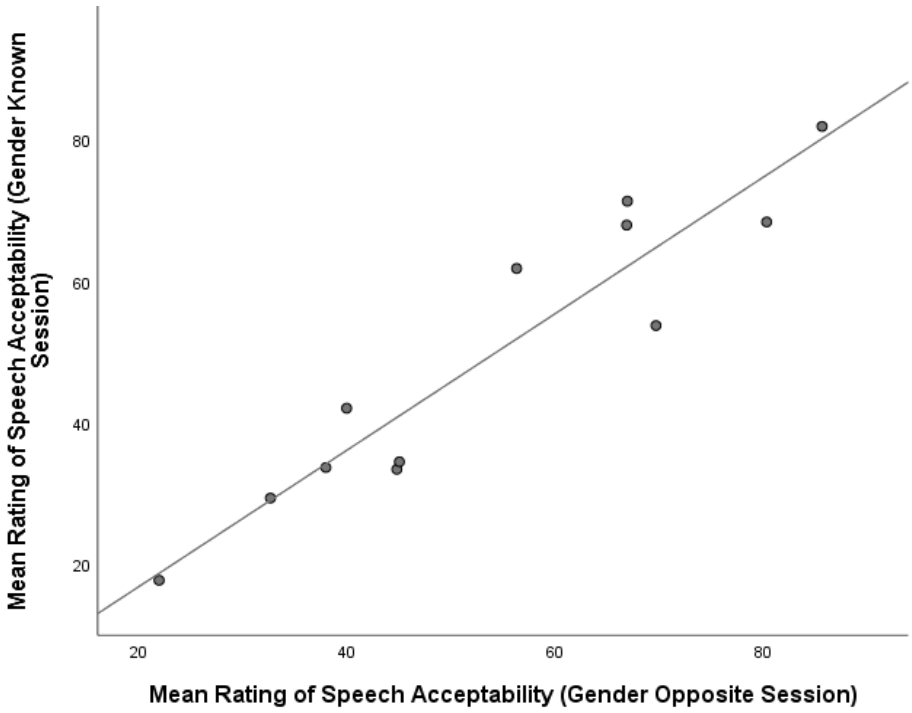

Figure 7. Relationship between mean ratings of speech acceptability across session 1 (Gender Opposite) and session 2 (Gender Known).

Finally, in order to visually present the shifts in listener scores for both SA and LC across the two listening conditions, these comparative data are presented in Figures 8 and 9. As can be seen, the mean auditory-perceptual scores for the female TE speakers was always less than that of their male counterparts. This finding was true regardless of the dimension being evaluated, or whether gender was known. Inspection of these two figures clearly indicates that when females were correctly identified and that was known to the listeners in the follow up session, they were rated less favorably. However, although scores for SA remained unchanged for males across the Gender Opposite and Gender Known listening sessions (Figure 8), they were rated more favorably for LC when gender was known (Figure 9). In comparison, our female TE speakers were consistently judged less favorably for both SA and LC when gender was known.

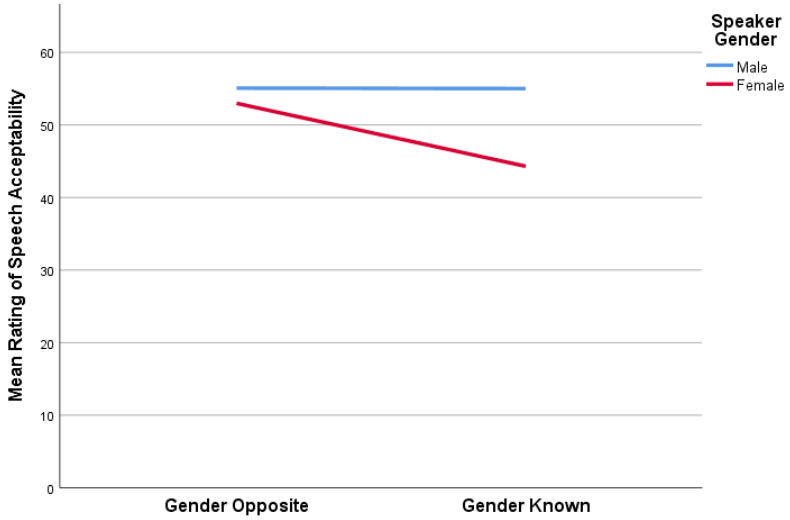

Figure 8. The mean ratings of speech acceptability for TE speakers across Sessions 1 (Gender Opposite) and 2 (Gender Known) based on gender.

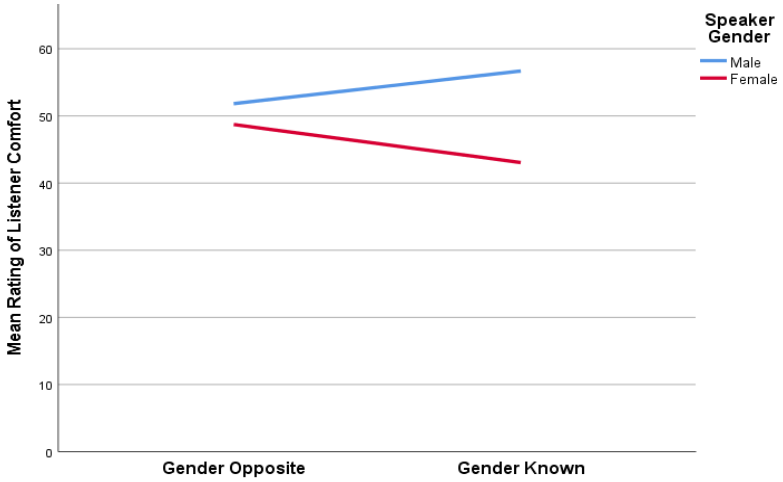

Figure 9. The mean ratings of listener comfort for TE speakers across Sessions 1 (Gender Opposite) and 2 (Gender Known) based on gender.

4. Discussion

The objective of this study was directed at the question of how a listener's knowledge of a TE speaker's gender would influence their judgments of the auditory-perceptual features of SA and LC. While concerns have evolved from the past literature around the potential for differential social penalty specific to voice quality in those who are laryngectomized, these investigations sought to assess gender differences associated with TE speech. In seeking to address this question, several variables were purposely controlled. In anticipation of the deception component of this experiment (Session 1), speaker samples were chosen based on the experimenter's evaluation that the samples included were ambiguous and not obviously associated with either gender. Had the samples clearly sounded male or female, the deception aspect of the experiment may have become apparent to the listener, and we believe that listener ratings of female speakers would have deteriorated further. However, consideration should be given to the fact that TE samples are characterized by lower fundamental frequency and considerable noise [2], which inherently decrease the overall feminine characteristics of the voices [13]. Therefore, to control for variables that might influence perceptual judgements, several steps were taken as part of the experimental design, namely, the deception in Session 1.

First, four randomized lists of speaker samples were developed so that listeners were always presented with a uniquely ordered list of samples on which to complete their ratings. Secondly, the auditory-perceptual feature rated in each session was counterbalanced (i.e., if ratings of SA or LC were carried out first, the opposite feature would be rated 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 identification of each of the speaker samples prior to each listening session; this was true whether judgments were being made for either the true gender identification or deception condition.

In evaluating the present data, it is important to consider what the SA ratings obtained in this study suggest. As described in the previous literature by Eadie and colleagues [13], judgments of SA encourage the listener to identify "acceptability" as it relates to their own personal beliefs about deviation from a normal signal and potential disability [26]. Thus, it is not unreasonable to assume that gender is a critical factor in determining a level of acceptability regardless of speaker gender. That is, listeners likely have preconceived templates of how men and women should sound, and, when intrinsic perceptual standards are violated, or thresholds challenged, the associated ratings of a sample will be altered [8,13]. However, in this regard and specific to the present data, it is interesting to note that penalizing judgements were more apparent for female TE speakers for SA as compared to LC.

By strict definition, there is some degree of "comfort" inherently considered in "acceptability" ratings; thus, the features may not be entirely mutually exclusive. Yet, there were differences in how these two auditory-perceptual constructs (i.e., SA and LC) were rated across listening sessions. It appears that when listeners are requested to make overall judgments about a speaker's voice, LC may comprise one component of the rating, but more quantitative aspects of the voice, such as its pitch, noise level, etc., may further influence such overall judgments of SA [29]. This difference between the SA 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 as being slightly more acceptable than when the samples were known to be female.

The same trend was seen for the construct of LC; however, this difference was not found to be significantly different. When male TE speakers were analyzed independently from their female counterparts, there were no significant differences in ratings for either SA or LC. Consequently, it may be concluded that female TE speakers may face greater penalty for the unusual characteristics of their alaryngeal voices than male TE speakers when it comes to listener judgments; this finding was more prominent for SA in the present study [32].

When *listener* data were analyzed by gender, findings revealed that neither male nor female listeners rated a particular speaker gender significantly worse than the other. This suggests that auditory-perceptual ratings of male and female TE speakers are fairly consistent across a variety of listeners and cannot necessarily be predicted based on the gender of the rater. Further, it may suggest that male and female listeners have similar ideals when it comes to the SA and LC associated with listening to TE speech [23,24]. In this context, it must again be noted that TE voices are inherently and considerably different from that of a normal speaker. Thus, the abnormality of the voice/speech sample will be easily recognized, and a listener may work to adjust their perceptual template in order to provide a rating that may better represent the sample population being assessed—in the present case, that of TE speakers [7,8].

One final area that deserves comment is that related to potential acoustic markers that may distinguish TE speakers. Because the pharyngoesophageal voice source in TE speakers is not capable of adductor–abductor control, the signal generated is highly aperiodic. Thus, the noise quality of TE voices, in addition to the lowered fundamental frequency [14,15], may lead to misperceptions of gender that may often default to that of a male speaker. In assessing data from the present study that are reflected in Figures 5–7, it can be seen that there appear to be two "clusters" of speaker data represented in those graphics. This raised the question of whether those clusters might correspond to distinct groups of our male and female TE speakers. Consequently, we assessed those data in a post-hoc manner to determine where individual speakers were represented on those figures.

In viewing Figure 5 as an example, we were able to determine that the six data points shown in the lower quadrant of that figure represent four female and two male TE speakers. In contrast, the upper quadrant of Figure 5 represents four males and the remaining two female speakers. This would suggest that listeners were able to make independent judgments of speakers based on the dimensions assessed; however, the knowledge of speaker gender that is represented in this figure may have biased their ratings to the less favorable side of the scale for at least four of our six female speakers. Therefore, based on these data, future work that seeks to comprehensively describe the acoustic characteristics of any given TE speaker's voice may provide valuable information that potentially guides a listener's assessment of gender.

5. Limitations of the Present Study

Although the findings of our work provide evidence that a gender bias might exist, several limitations to our work must be noted. First, both the speaker (n = 12) and listener (n = 16) groups in this study are relatively small; thus, generalizations from the present findings must be made with a degree of caution. In this regard, future attempts at replication

of this study, or similar approaches to identifying possible gender bias in postlaryngectomy speakers with a larger group of participants in both groups, would be of benefit. As sample sizes of speakers increase, one would need to be mindful of the proficiency of the speaker, as other factors such as intelligibility reductions would need to be carefully considered. Recall that our speakers were judged to be highly intelligible to reduce any potential confound specific to the listener also having challenges in understanding the speaker.

Secondly, our listener group comprised young adults. It is possible that individuals who are older, perhaps a cohort that is similar in age to the speakers in this study, could provide different findings. To our knowledge, there are no published reports that address gender bias with specific respect to the age of either the speaker(s) or listener(s). However, if older listeners are used in the future, it would be important to assess and quantify one's hearing status, as age-related hearing loss could impact the findings. It would, therefore, appear that future studies that consider not only increasing sample size but also the age of listeners could offer findings that further validate the present results. Lastly, depending on the age of listeners, it is possible that judgments of a speaker may be penalized more or less as one sees themselves as being a "peer" relative to age of the speaker.

6. Clinical Implications

The present study sought to determine whether explicit knowledge of a TE speaker's gender would influence the perceptual ratings assigned by naive listeners. In that respect, our study was designed to address the potential for a gender bias associated with TE voice and speech. Based on the data gathered, it is apparent that listeners rate female TE speaker samples to be more unacceptable and less comfortable to listen to when the samples are known (or are assumed) to be female speakers. While the underlying reasons for this finding remain incomplete, it would appear that expectation(s) of what might be termed "gender markers" and the general acoustic characteristics of TE speech were actively considered by the current listeners during their assessment [5]. However, it does seem likely that a multitude of factors are involved when a listener judges a TE speaker's voice, including its collective attributes of voice "quality" and the general comfort level associated with listening to a particular voice in a social setting and the related disability it may pose [33]. It will be important for future research to evaluate acoustic information in conjunction with perceptual data for male and female TE speaker samples, to determine the specific parameters that may affect listener judgements. Additionally, the combined impact of visual and acoustic information is likely critical to understanding such judgments [17].

In terms of clinical relevance, the present results are important to consider in the context of pre- and postoperative counselling for female laryngectomees, to ensure the most informed level of education in the context of postlaryngectomy rehabilitation [34]. While there would be wide acceptance of the notion that the capacity to restore a more feminine-sounding postlaryngectomy voice would be advantageous [35], this desire remains challenging due to the unique postsurgical anatomy that forms the new voicing source. At the very least, however, providing information on the potential limitations of TE voice and speech relative to the speaker's gender would appear to be an important area of discussion in both pre- and postoperative counselling for those who will undergo total laryngectomy for laryngeal cancer [20–22,25,35].

7. Conclusions

This study empirically examined the potential relationship between speaker gender and listener perceptions of tracheoesophageal (TE) speech. The study involved collecting auditory-perceptual judgments of two descriptive dimensions—speech acceptability (SA) and listener comfort (LC)—from a group of 16 listeners who were naïve to TE speech. Auditory-perceptual judgments were made on two occasions in an effort to determine if knowledge of the speaker's gender influenced a listener's ratings. During the first session, listeners were deceived about the gender of the voice samples presented, and, during the second session, listeners were informed of the true gender of the voice samples. Data

revealed that statistically significant differences in listener ratings were observed for SA when a speaker's correct gender was known. Thus, the same voice samples were rated differently when a speaker was believed to be male versus when they were believed to be female. A similar finding was observed for ratings of LC; however, this trend did not reach significance. Females were judged by listeners to be more unacceptable and to some degree more uncomfortable to listen to when they were known by the rater to be female. When judgments of male speakers were analyzed independent of their female counterparts, no significant differences between how they were rated on SA or LC emerged. Collectively, these findings suggest that at least some degree of gender bias exists in perceptions of TE speech, and that female TE speakers may be disproportionately penalized when compared to their male TE-speaking counterparts. Consequently, consideration of gender and voice quality in those who undergo TE puncture voice restoration deserves continued exploration in the context of postlaryngectomy voice and speech rehabilitation and related outcomes.

Author Contributions: J.L.B.: study conception, design, data acquisition and analysis, interpretation, drafting of manuscript, revising, and final approval. N.J.: data acquisition and analysis, data interpretation, drafting of manuscript, revising, and final approval. P.C.D.: study conception, design, data acquisition and analysis, data interpretation, drafting of manuscript, revising, and final approval. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Research Ethics Board of Western University (#104645), June 2017.

Informed Consent Statement: Informed consent was obtained from all participants involved in this study.

Data Availability Statement: Data is contained within the article.

Acknowledgments: The authors would like to thank Sebastiano Failla for his assistance in preparing digital voice stimuli and for measuring the raw listener data, respectively.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Singer, M.I.; Blom, E.D. An endoscopic technique for restoration of voice after laryngectomy. *Ann. Otol. Rhinol. Laryngol.* **1980**, *89*, 529–533. [CrossRef] [PubMed]
2. Robbins, J.; Fisher, H.B.; Blom, E.C.; Singer, M.I. A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production. *J. Speech Hear. Disord.* **1984**, *49*, 202–210. [CrossRef] [PubMed]
3. Pindzola, R.H.; Cain, B.H. Acceptability ratings of tracheoesophageal speech. *Laryngoscope* **1988**, *98*, 394–397. [CrossRef] [PubMed]
4. Belandese, M.H.; Lerman, J.W.; Gilbert, H.R. An acoustic analysis of excellent female esophageal, tracheoesophageal, and laryngeal speakers. *J. Speech Lang. Hear. Res.* **2001**, *44*, 1315–1320. [CrossRef] [PubMed]
5. Drugman, T.; Rijckaert, M.; Janssens, C.; Remacle, M. Tracheoesophageal speech: A dedicated objective acoustic assessment. *Comput. Speech Lang.* **2014**, *30*, 16–31. [CrossRef]
6. Doyle, P.C.; Eadie, T.L. Scaling of pleasantness and acceptability in tracheoesophageal (TE) speakers. *J. Voice* **2005**, *19*, 373–383.
7. Kent, R.D. Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *Am. J. Speech-Lang. Pathol.* **1996**, *5*, 7–23. [CrossRef]
8. Kreiman, J.; Gerratt, B.R.; Kempster, G.B.; Erman, A.; Berke, G.S. Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *J. Speech Hear. Res.* **1993**, *36*, 21–40. [CrossRef]
9. Ganz, P.A. Current US cancer statistics: Alarming trends in young adults? *JNCI J. Natl. Cancer Inst.* **2019**, *111*, 1241–1242. [CrossRef]
10. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer Statistics, 2021. *Ca Cancer J. Clin.* **2021**, *71*, 7–33. [CrossRef]
11. Fowler, N.M.; Kmiecik, J.; Khan, M.J.; Scharpf, J.S.; Lorenz, R.R.; Burkey, B.B. Improved outcomes in primary prosthesis placement during tracheoesophageal puncture. *Otolaryngol. Head Neck Surg.* **2012**, *147*, 161–162. [CrossRef]
12. Pagedar, N.A.; Bayon, R.; Gudgeon, J.; Nelson, R.F.; Van Daele, D.J.; Hoffman, H.T. Tracheoesophageal puncture with immediate prosthesis placement. *Laryngoscope* **2013**, *124*, 466–468. [CrossRef] [PubMed]
13. Eadie, T.L.; Doyle, P.C.; Hansen, K.; Beaudin, P.G. Influence of speaker gender on listener judgements of tracheoesophageal speech. *J. Voice* **2008**, *22*, 43–57. [CrossRef] [PubMed]

14. Gelfer, M.P.; Mikos, V.A. The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *J. Voice* **2005**, *19*, 544–554. [CrossRef] [PubMed]
15. Latinus, M.; Taylor, M.J. Discriminating male and female voices: Differentiating pitch and gender. *Brain Topogr.* **2012**, *25*, 194–204. [CrossRef] [PubMed]
16. Murry, T.; Singh, S. Multidimensional analysis of male and female voices. *J. Acoust. Soc. Am.* **1980**, *68*, 1294–1300. [CrossRef] [PubMed]
17. Peynircioğlu, Z.F.; Brent, W.; Tatz, J.R.; Wyatt, J. McGurk effect in gender identification: Vision trumps audition in voice judgments. *J. Gen. Psychol.* **2017**, *144*, 59–68. [CrossRef] [PubMed]
18. Skuk, V.G.; Schweinberger, S.R. Gender differences in familiar voice identification. *Hear. Res.* **2013**, *296*, 131–140. [CrossRef] [PubMed]
19. Eadie, T.L.; Day, A.M.B.; Sawin, D.E.; Lamvik, K.; Doyle, P.C. Auditory-perceptual speech outcomes and quality of life after total laryngectomy. *Otolaryngol. Head Neck Cancer* **2012**, *148*, 82–88. [CrossRef]
20. Bickford, J.; Coveney, J.; Baker, J.; Hersh, D. Living with the altered self: A qualitative study of life after total laryngectomy. *Int. J. Speech-Lang. Pathol.* **2013**, *15*, 324–333. [CrossRef]
21. Dahl, K.L.; Bolognone, R.K.; Childes, J.M.; Pryor, R.L.; Graville, D.J.; Palmer, A.D. Characteristics associated with communicative participation after total laryngectomy. *J. Commun. Disord.* **2022**, *96*, 1061–1084. [CrossRef] [PubMed]
22. Sharpe, G.; Camoes Costa, V.; Doubé, W.; Sita, J.; McCarthy, C.; Carding, P. Communication changes with laryngectomy and impact on quality of life: A review. *Qual. Life Res.* **2019**, *28*, 863–877. [CrossRef] [PubMed]
23. van Sluis, K.E.; Kornman, A.F.; van der Molen, L.; van den Brekel, M.W.; Yaron, G. Women's perspective on life after total laryngectomy: A qualitative study. *Int. J. Lang. Commun. Disord.* **2020**, *55*, 188–199. [CrossRef] [PubMed]
24. Wulff, N.B.; Dalton, S.O.; Wessel, I.; Arenaz Bua, B.; Löfhede, H.; Hammerlid, E.; Homøe, P. Health-related quality of life, dysphagia, voice problems, depression, and anxiety after total laryngectomy. *Laryngoscope* **2022**, *132*, 980–988. [CrossRef] [PubMed]
25. Cox, S.R.; Theurer, J.A.; Spaulding, S.J.; Doyle, P.C. The multidimensional impact of total laryngectomy on women. *J. Commun. Disord.* **2015**, *56*, 59–75. [CrossRef] [PubMed]
26. Doyle, P.C.; Baker, A.M.; Evitts, P.M. Communication competence and disability secondary to laryngectomy and tracheoesophageal puncture voice restoration. *Int. J. Lang. Commun. Disord.* **2023**, *58*, 441–450. [CrossRef] [PubMed]
27. Eadie, T.; Faust, L.; Bolt, S.; Kapsner-Smith, M.; Pompon, R.H.; Baylor, C.; Méndez, E. Role of psychosocial factors on communicative participation among survivors of head and neck cancer. *Otolaryngol. –Head Neck Surg.* **2018**, *159*, 266–273. [CrossRef] [PubMed]
28. Fairbanks, G. *Voice and Articulation Drillbook*, 2nd ed.; Harper & Row: New York, NY, USA, 1960.
29. Bennett, S.; Weinberg, B. Acceptability ratings of normal, esophageal, and artificial larynx speech. *J. Speech Hear. Res.* **1973**, *16*, 608–615. [CrossRef] [PubMed]
30. O'Brian, S.; Packman, A.; Onslow, M.; Cream, A.; O'Brian, N.; Bastock, K. Is listener comfort a viable construct in stuttering research? *J. Speech-Lang. Hear. Res.* **2003**, *46*, 503–509. [CrossRef]
31. Kempster, G.B.; Gerratt, B.R.; Verdolini Abbott, K.; Barkmeier-Kramer, J.; Hillman, R.E. Consensus Auditory-Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *Am. J. Speech-Lang. Pathol.* **2009**, *18*, 124–132. [CrossRef]
32. Searl, J.P.; Small, L.H. Gender and masculinity-femininity ratings of tracheoesophageal speech. *J. Commun. Disord.* **2002**, *35*, 407–420. [CrossRef] [PubMed]
33. World Health Organization. *International Classification of Impairments, Disabilities, and Handicaps*; World Health Organization: Geneva, Switzerland, 1980.
34. Kazi, R.; Kiverniti, E.; Prasad, V.; Venkitaraman, R.; Nutting, C.M.; Clarke, P.; Rhys-Evans, P.; Harrington, K.J. Multidimensional assessment of female tracheoesophageal prosthetic speech. *Clin. Otolaryngol.* **2006**, *31*, 511–517. [CrossRef] [PubMed]
35. Eadie, T.L.; Doyle, P.C. Auditory-perceptual scaling and quality of life in tracheoesophageal speakers. *Laryngoscope* **2004**, *114*, 753–759. [CrossRef] [PubMed]