

A Critical Review and Study:

Situationally-Bound Judgments of “Listener Comfort” for Postlaryngectomy Voice and Speech*

Natalie Smith

M.Cl.Sc. (SLP) Candidate

Western University: School of Communication Sciences and Disorders

This report presents findings from a critical appraisal of the literature as well as the results of a study examining the relationship between listener comfort and a situationally-bound listening scenario for tracheoesophageal (TE), esophageal (ES), and electrolaryngeal (EL) speakers. The critical appraisal included evaluations of one randomized block design, two between groups studies, and three mixed (between and within subjects) design studies. The study involved having naïve listeners (n =12) make auditory-perceptual judgments of listener comfort. Judgments of listener comfort were made based on two separate listening scenarios: one for a suggestive *social communication situation* and the other for a suggestive *telephone conversation*. Findings from the critical review and the study suggest that judgments of listener comfort did not vary for a suggestive listening scenario. However, TE speakers were rated significantly more comfortable to listen to across both listening scenarios compared to the other speaker groups.

Introduction

Disability secondary to voice and speech loss following total laryngectomy has been shown to negatively influence social well-being, as well as postlaryngectomy quality of life (QOL; Doyle, 1999; Eadie & Doyle, 2004; Eadie & Doyle, 2005; Fung & Terrell, 2004). Research in QOL for alaryngeal speakers indicates that they experience a negative impact on daily activities and social participation postlaryngectomy (Doyle, 1999). These activities may include social conversation or speaking on the telephone, therefore creating situationally-bound challenges. By identifying situationally-bound challenges for verbal communication, laryngectomees can make a more informed decision on the method of speech used postlaryngectomy depending on their specific communication needs.

Furthermore, the changes to the acoustic signal postlaryngectomy have the potential to negatively influence listener judgments (Doyle & Eadie, 2005). Although the restoration of postlaryngectomy voice may result in functional levels of communication from the standpoint of intelligibility, decrements in the auditory signal continue to carry a potential penalty for the speaker based on listener judgments. Previous research has shown that tracheoesophageal (TE) speakers are rated significantly better across most auditory judgments compared to esophageal (ES) and electrolaryngeal (EL) speakers, however, TE speakers still differ from laryngeal speakers (Robbins, Fisher, Blom, & Singer, 1984; Doyle &

Eadie, 2005). Studies have also shown that the addition of visual information to the auditory input can increase ratings of intelligibility for alaryngeal speakers, with TE speakers being more intelligible than the other alaryngeal communication methods (Evitts, Portugal, Van Dine, & Holler, 2010).

Based on the potential reductions in listener judgments of postlaryngectomy speech, a study was conducted to compare auditory-perceptual judgments of listeners for all three modes of postlaryngectomy speech. More specifically, the project sought to identify how listener judgments of a perceptual feature termed “listener comfort” (LC) varied depending on a suggested listening scenario. LC is a term used to refer to how comfortable a listener feels listening to a person’s speech and reflects the listener’s feelings about the way the person was speaking, not what the person was saying or their personality (O’Brian et al., 2003).

Objectives

The purpose of this paper sought to critically appraise the existing literature relating to listeners’ perceptual judgments of alaryngeal speech and the influence of auditory versus audio-visual input when it is provided. An additional objective was to complete a study to determine if a suggested listening scenario influenced naïve listener judgments of LC for TE, ES, and EL speech.

Study 1: Critical Review

*This paper was created as a required assignment for the CSD9639 Evidence Based Practice for Clinicians course at Western. While it has been evaluated by course instructors for elements of accuracy and style, it has not undergone formal peer-review.

Methods

Search Strategy

Computerized databases including Scholars Portal, PubMed, and PsychINFO, as well as ASHA Publications were searched using the following search strategy: [(alaryngeal) AND (speakers) AND [(listener judgments) OR (perceptual judgments)]] or [(listener comfort)] or [(alaryngeal) AND (speakers) AND (audio-visual)]. Reference lists from articles were also used to obtain additional relevant articles.

Selection Criteria

For this review, studies were required to include data on at least one mode of alaryngeal communication, including TE, ES, and EL speakers. In addition, listener judgments were required from naïve listeners through rating scales using either auditory or audio-visual input. One article was chosen based on the criteria of LC judgments, which had employed a stuttering population.

Data Collection

The literature search resulted in the identification of six articles. These articles included one randomized block design, two between groups studies, and three mixed design studies.

Results

Eadie, Day, Sawin, Lamvik, and Doyle (2012) used a between groups study of 25 alaryngeal speakers (20 male; 5 female) to examine the relationship between listener-rated speech intelligibility and acceptability, and self-reported QOL. The speakers were at least one year postlaryngectomy and were grouped by mode of communication: TE (n=16), ES (n=2), and EL (n=7). Intelligibility was measured using the Sentence Intelligibility Test (SIT; Yorkston, Beukelman, & Tice, 1996), which was transcribed by 33 naïve listeners. To assess acceptability, 15 listeners rated a recording of a reading passage. QOL was measured using the University Washington Quality of Life (Hassan & Weymuller, 1993) and the Voice Handicap Index – 10 (Rosen, Lee, Osborne, Zullo, & Murry, 2004). Results of a Pearson correlation coefficient showed higher intelligibility and acceptability ratings for TE speakers compared to EL speakers. An appropriate post hoc analysis showed a moderate correlation between listeners' ratings of intelligibility and acceptability, and weak to moderate correlations for acceptability and QOL.

A strength of this study was the high validity. The study employed a relatively large speaker sample size

(n=25) representative of the alaryngeal population. The speakers were selected based on a set of criteria that included native English speakers with no other speech, language, or voice symptoms that may distract the listener. Separate listener groups rated intelligibility and acceptability to prevent learning effects. The stimuli and rating scales were appropriate for the purpose of the study, resulting in good construct validity. The samples were recorded with consideration to sampling rate and background noise. The study also had high reliability. Twenty percent of the reading passage samples for speech acceptability were repeated and a Pearson correlation coefficient yielded acceptable intrarater reliability. Adequate interrater reliability was determined using an appropriate intraclass correlation coefficient.

A limitation of the study was that ES speakers were not included in the post hoc analysis due to the small sample size (n=2) and, therefore, generalizations to this group were limited.

Overall, this study offers compelling evidence in support of a moderate correlation between intelligibility and speech acceptability when presented through auditory input. Furthermore, it demonstrated that listener judgments do not predict self-rated QOL.

Eadie and Doyle (2004) used a randomized block design with 28 TE speakers (22 male; 6 female) to determine listeners' auditory-perceptual judgments (overall speech severity, naturalness, acceptability, and pleasantness) as well as the relationship of these judgments to QOL. The speakers completed the University of Michigan Head and Neck Quality of Life instrument (Terrell, Nanavati, Esclamado, Bishop, Bradford, & Wolf, 1997) and a recording of the Rainbow Passage (Fairbanks, 1960). Fifteen naïve listeners with no history of speech, voice, language or hearing difficulties listened to the audio recordings of the Rainbow Passage and rated the samples on the four auditory-perceptual judgments using direct magnitude estimation (DME) procedures. An appropriate Pearson correlation coefficient showed that although the auditory-perceptual judgments were highly interrelated, they were distinguishable judgments. Furthermore, a moderate correlation was found for the relationship between audio-perceptual ratings and self-rated QOL for the speakers.

A strength of the study was the large sample size (n=28) of TE speakers and detailed methods section that would allow for it to be replicated. The stimuli was appropriate and was recorded with consideration to background noise and sampling rate. A DME

procedure was employed with a modulus to measure the auditory-perceptual judgments. The modulus sample represented the midpoint of each auditory-perceptual feature relative to all of the speakers and was repeated every six stimuli to minimize drifting in listener judgments. Order effects were also prevented by using randomized speaker samples and randomized listener judgment order. Reliability was appropriately addressed by repeating 25% of the samples for each of the four listener judgments. Intrarater and interrater reliability were appropriately established using a Pearson correlation coefficient and a Cronbach's alpha respectively. Both intrarater and interrater reliability were adequate for all listener judgments.

In order to make the speakers more representative of the alaryngeal population, more female speakers would need to be included. Additionally, no information was provided on the gender of the listeners.

This study provided compelling evidence that although overall speech severity, naturalness, acceptability, and pleasantness are highly interrelated, listeners are able to distinguish between these dimensions and obtain different measurements.

Most, Tobin, and Mimran (2000) employed a between groups design with five male speakers for each of the following groups: laryngeal, TE, good esophageal (GE) and moderate esophageal (ME) speakers to compare the acoustic and perceptual features of their speech. Each speaker recorded a variety of 16 sentences. Twenty-five naïve listeners listened to a subset of the total sentences and rated them for acceptability as well as transcribed 68 words for intelligibility. All of the speaker samples were compiled and randomized so that each sentence and word was rated by five of the 25 listeners. The mean acceptability ratings were recorded for each of the four speaker groups. An appropriate one-way analysis-of-variance (ANOVA) followed by multiple comparisons showed a significant difference among the four groups. Acceptability ratings were highest for the laryngeal group, followed by TE, GE, and ME speakers. Similarly, the mean for each group's intelligibility ratings were recorded. A one-way ANOVA followed by multiple comparisons showed a significant difference between the laryngeal and ME groups, with laryngeal speakers being more intelligible. In addition, significant correlations were found between acceptability and word intelligibility as well between as acceptability and phoneme intelligibility.

A primary strength of the study was the separation of ES speakers into good and moderate proficiency levels, allowing for more distinguishable groups. Another strength was that the speaker samples were randomized into 25 lists to prevent order effects. Additionally, the presentation of intelligibility and acceptability ratings were counterbalanced.

A limitation of the study was that only male speakers were included, therefore, generalizations to the overall alaryngeal population should be made with caution. For the listeners, no information was provided on gender and only 5 of the 25 listeners rated each sentence and word for intelligibility and acceptability. Only the means, ranges, and standard deviations were reported with no further analysis of interrater reliability for the judgments of intelligibility and acceptability. There were no reported data to account for intrarater reliability.

Overall, this study showed equivocal evidence that acceptability and intelligibility ratings differ for alaryngeal speech methods when presented through auditory input.

Hubbard and Kushner (1980) employed a mixed (between and within subjects) design to examine the effect of presentation mode (visual, auditory, and combined auditory-visual) on intelligibility ratings of 8 good-to-superior ES speakers and 8 laryngeal speakers. Each speaker read five different 2- to 12-word sentences while being recorded on video tape. Three groups of 10 naïve listeners were randomly assigned to either the visual, auditory, or combined auditory-visual mode of presentation and transcribed verbatim sentences read by the ES speakers. The three groups of listeners returned after four weeks and repeated the procedure for the laryngeal speakers, and again, were randomly assigned a presentation mode. An analysis of the raw data using means showed that more words were transcribed correctly in the auditory mode than the visual mode, and more words were correctly transcribed in the combined mode than the auditory mode. An appropriate multifactor ANOVA with repeated measures showed a significant difference in intelligibility between the ES and normal speakers. The effect size showed the group effect was greatest for the auditory mode and least for the visual mode. A t-test for uncorrelated means showed significantly different intelligibility ratings in the combined mode compared to the auditory mode for ES speakers, suggesting that visual information enhanced communication for the ES speakers.

A strength of the study was the selection criteria for the good-to-superior speakers, demonstrating internal validity. A detailed methods section was provided and the stimuli were appropriate for the objectives of the study demonstrating construct validity. A final strength was having a 4 week period before the listeners came back to rate the intelligibility of the laryngeal speakers to minimize familiarity effects.

A limitation of the study was the small sample size for the ES speakers ($n=8$). The male-to-female ratio was also limiting, so generalizations of the findings to the ES speaker population should be made with caution. Another limitation was having only 10 listeners for each of the presentation modes. The results would be more reliable if at least two groups rated each of the presentation modes. Additionally, it was not reported if the sentence presentation for each of the modes was randomized. Therefore, there may have been an order effect. Another weakness of the study was that although means, ranges, and standard deviations were reported for the three presentation modes, there were no further reported interrater reliability measures. Lastly, the intelligibility scores should have been analyzed with consideration to sentence length because the stimuli ranged from 2 – 12 words per sentence.

Overall, the study showed equivocal evidence supporting intelligibility differences for the three modes of presentation.

Evitts, Portugal, Van Dine, and Holler (2010) used a mixed (between and within subjects) design with male ES ($n=1$), TE ($n=1$), and EL ($n=1$) speakers as well as one male laryngeal speaker to determine the effects of mode of presentation (audio-only and audio-visual) on intelligibility ratings. The speakers recorded phonemically balanced sentences from the Hearing in Noise Test (Nilsson, Soli, & Sullivan, 1994) using a device that connected to an audio and audio-visual recorder. Twenty-nine naïve listeners (1 male, 28 female) listened to one of four master lists that consisted of 20 audio-only recordings and 20 audio-visual recordings for each of the speakers as well as 10% repeated samples for intrarater reliability. The listeners wrote what they heard verbatim for the audio-only and audio-visual modes presented. An appropriate 2×4 repeated measures ANOVA showed a significant main effect for mode of presentation, supporting that intelligibility was greater for most speech methods in the audio-visual mode, with the exception of the ES speaker. Results of the repeated measures ANOVA also showed a significant interaction effect between the speech method and mode of presentation.

A strength of this study was that two experienced speech-language pathologists selected a speaker that was highly intelligible and representative of the alaryngeal speech method. Given the visual nature of the study, speakers were also excluded based on facial hair, scars, or other facial features that were not related to laryngectomy but may be distracting to the listener. The results of a one-way ANOVA using transcribing errors showed that neither learning nor fatigue effects were present.

A significant limitation of the study was that only one speaker was chosen to represent each of the speech methods. Therefore, any conclusions from the study must be interpreted with caution. In addition, no female speakers were included. Another weakness was that the EL speaker selected was significantly less intelligible than the other alaryngeal speakers were when assessed by the experienced speech-language pathologists. This discrepancy limited the internal validity because the overall intelligibility rating for the EL speaker was already classified as lower than the other speakers and therefore comparisons across alaryngeal speech groups may not be representative regardless of the presentation mode. Furthermore, the listeners were not equally represented for gender and the criteria required that the listener had *minimal* to no exposure to alaryngeal speech. It was not specified what counted as *minimal* exposure, and this may have influenced intelligibility ratings. Listeners were not familiarized with alaryngeal voices prior to the rating task, so the unusual voice quality may have been distracting for the first few samples.

The words and sentences were phonemically balanced in the task; however, they were not visually, semantically, or syntactically balanced. The visibility of certain sounds and predictable wording may have influenced the intelligibility ratings. Lastly, only listeners with p values <0.5 and r values 0.5 were included in the study, which may have skewed the results.

Overall, the results from this study must be interpreted with caution, primarily because of the small sample size. Results provide equivocal evidence that the mode of presentation effects alaryngeal speech intelligibility, with audio-visual input resulting in higher intelligibility scores.

O'Brian, Packman, Onslow, Cream, and O'Brian (2003) employed a mixed (between and within subjects) design study to establish a) if LC was a viable construct in stuttering research and b) if

measurements differed from speech naturalness. Ten adults who stutter (7 male; 3 female) and 10 adult controls matched for age and gender made a video recording while speaking. Each of the adults who stutter made a video recording post-treatment and pre-treatment. Two separate samples from a video recording were chosen for the controls to account for equal numbers of samples for both groups. Thirty naïve listeners were split into two groups to rate the speakers. Group A listened to the recordings and rated LC on a 9-point, equal interval scale. Group B listened to the same recordings and rated speech naturalness using a 9-point scale. Appropriate Pearson correlation coefficients were calculated for pre-treatment listener comfort and speech naturalness (-.96), post-treatment ratings (-.49), and control ratings (-.86). The correlations were negative because the values on the scales were reversed. The pre-treatment correlation for speech naturalness and LC was high and the post-treatment correlation was only moderate, which may indicate that the LC and speech naturalness scales were measuring two different features. If the scales were measuring the same thing, it would be expected to remain highly correlated for all conditions.

A primary strength of this study was the high construct validity with a detailed methods section that would allow for replication. Additionally, the criteria for speakers included treatment type and post-treatment duration. The only treatment type included in this study was the Camperdown Program, which strengthened the internal validity. The stimuli were valid because a video recording is representative of face-to-face communication. Appropriate intraclass correlations were calculated for the rating scales and both were in the acceptable range, with speech naturalness being superior in reliability. Interrater reliability was measured by comparing the value assigned by each listener to the value assigned by the other listeners. The LC scale showed lower interrater reliability (56% of scores within ± 1) than the speech naturalness scale (68% of scores within ± 1). An ANOVA showed listener ratings on the LC scale had a higher variance compared to the speech naturalness scale. Intrarater reliability for both LC and speech naturalness were adequate.

Limitations of the study included the small sample size of adults who stutter ($n=10$), which would limit the generalization of the results. No information was provided on the presence of secondary stuttering behaviours. The presence of secondary behaviours such as blinking or twitches may have made the listener less comfortable, therefore influencing the judgment.

Overall, this study offers suggestive evidence that the LC scale has moderate reliability in the stuttering population and measures something different from speech naturalness.

Discussion

The current literature suggests that listener judgments provide a valid measure of alaryngeal voice and speech (Eadie, Day, Sawin, Lamvik & Doyle, 2012; Eadie & Doyle, 2004; Most, Tobin & Mimran, 2000). Additionally, listening scenario may influence listener judgments of alaryngeal speech because the input of audio-only and audio-visual influences ratings of intelligibility (Hubbard & Kushner, 1980; Evitts, Portugal, Van Dine & Holler, 2010). Lastly, O'Brian et al. (2003) found that "listener comfort" is a viable measure in a stuttering population, and therefore it may be a useful perceptual dimension in alaryngeal communication.

Study 2: Experimental Study

The following study was conducted to determine the potential influence of a suggestive listening scenario on a perceptual judgment termed "listener comfort" (LC) for TE, ES, and EL speakers. The two listening scenarios compared were based on the suggestion of: a) a social communication situation and b) a telephone communication situation. It was important to investigate LC on the telephone because it can be a major avenue for personal and business communication for laryngectomees.

Methods

Speakers

Audio recordings from TE, ES, and EL speakers (6 male and 6 female speakers per group) were obtained from an alaryngeal speech sample database. In addition, the recordings of the 6 male and 6 female ES speakers were modified by digitally removing non-phrasal pauses from the recordings, resulting in 12 modified-esophageal (MES) speech samples and a total of 48 speech samples overall. All speech samples were obtained from native English speakers who had undergone total laryngectomy and were judged by an experienced clinician to have excellent speech intelligibility.

Listeners

Twelve (6 male and 6 female) normal hearing, young adults participated as naïve listeners in the study. All

participants were students at Western University (mean age = 23 years; 7 months; range = 21 – 27 years). None of the listeners had formal training in voice disorders or listening to alaryngeal speech. Those who agreed to participate provided full informed consent prior to serving as listeners. The Research Ethics Board at Western University gave approval for this study (Ethics Number: 105993).

Audio Recordings

Each speaker recorded The Rainbow Passage (Fairbanks, 1960) and the second sentence, “*The rainbow is a division of white light into many beautiful colours*” was used as stimuli. The speech samples were judged by an experienced clinician to be of good quality and highly intelligible. Four randomized lists comprised of the 48 speech samples and 8 repeated samples for reliability were made.

Procedures

The first of two listening sessions began with a familiarization task that allowed listeners to listen to four alaryngeal speech samples (a male or female sample of TE, ES, EL, and MES speakers). These four speech samples were not obtained from speakers associated with the experiment and were used only to familiarize the listeners with the unique acoustic characteristics inherent to alaryngeal speech. The listeners were then provided with instruction on how to rate the samples.

In the rating task, listeners were presented with one of four randomized lists of 56 speech samples. Using the speech samples, the listeners were required to make a judgment of “listener comfort” for a suggestive social situation or suggestive telephone conversation. The following definition of LC was used for judgments of a social situation: “How comfortable you would feel listening to the person’s speech in a social situation. Your response should reflect your feelings about the way the person was speaking (i.e., how comfortable you would feel listening to them), not what the person was saying or how their personality affected you” (O’Brian et al., 2003). The same definition was provided for “listener comfort modified” (LC(M)), but specified in a telephone conversation (O’Brian et al., 2003). In the first session, half of the listeners rated samples based on the social situation definition and the other half of the listeners rated samples specific to the definition related to a telephone conversation. For both listening scenarios, listeners were asked to mark their ratings of the judgment on a 100mm visual analogue scale (VAS) after listening to each speech sample. The listeners were able to listen to the speech samples an

unlimited number of times prior to making a judgment. However, once a rating was provided, listeners were asked to not return to past samples and/or alter their prior ratings.

The second listening session occurred 7-10 days following the first session. In this session, the participant was asked to rate another randomized set of the speech samples, but judgments made in the second session were for the opposite listening scenario that they rated in the first session (either a social situation or a telephone conversation). The same audio recordings and procedure outlined above were used for the second listening session; however, listeners were not explicitly told that they were the same audio recordings.

Reliability

Internal validity was obtained by having the listeners rate 8 repeated speech samples at the end of both sessions (1 male and 1 female TE, ES, EL, and MES speaker). Listeners were not informed that these samples were duplicated. The ratings from the repeated samples were compared to the first ratings of the samples to evaluate intrarater reliability. An analysis of the raw data showed that with the exception of one sample, listeners rated the repeated samples +/-15 points of the first rating at least 50% of the time. Repeated samples were rated within +/-10 points at least 33% of the time and +/-5 points at least 25% of the time.

Data Analysis

A Pearson correlation coefficient was used to examine the relationship between judgments of listener comfort in a suggested social situation and a suggested telephone conversation. Independent t-tests were calculated to determine the relationship between speaker mode and suggested listening scenarios. A predetermined level of $p < 0.05$ was used for all analyses.

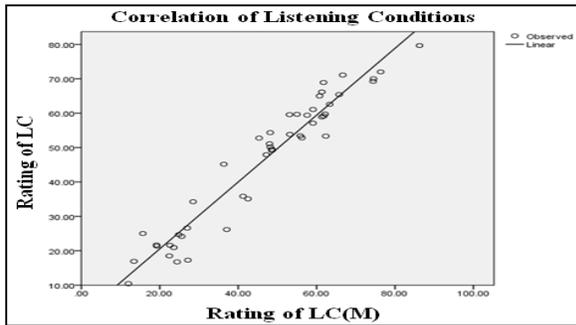
Results

An analysis of the raw data showed that no significant differences were found for speaker gender, so measures of central tendency were calculated using both male and female ratings combined (Table 1). A Pearson correlation coefficient showed a significant relationship between the two suggestive listening scenarios, with $r = 0.966$ (Graph 1 and Graph 2). Results from the independent t-tests showed TE speakers were significantly more comfortable to listen to for both the suggestive social and telephone listening scenarios when compared to

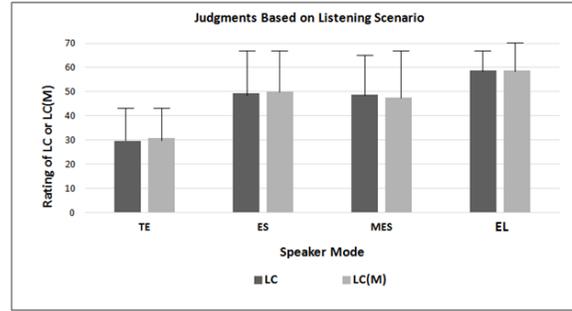
ES, EL, and MES speakers (Table 2). No other significant differences were found between the speaker groups. An analysis of the raw data also demonstrated that within each speaker group, there was a wide range of listener comfort judgments for both suggestive listening scenarios (Graph 3 and Graph 4).

Scenario	Statistic	TE	ES	MES	EL
LC	Mean	29.5	49.37	48.64	58.66
	Median	25.17	56.75	55.21	53.83
	Mode	21, 26	59	-	53
	SD	14.92	19.46	19.02	9.84
LC(M)	Mean	30.76	50	47.56	58.61
	Median	27.08	50.63	57.04	56.25
	Mode	27	48, 61	59	-
	SD	13.33	18.83	20.59	11.59

Table 1: Central tendency data for all speech modes in both listening scenarios.



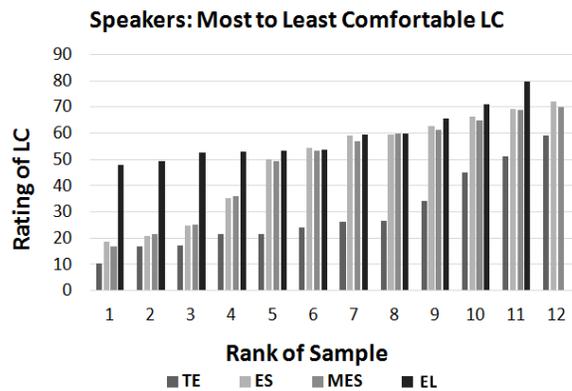
Graph 1: A Pearson correlation coefficient showed a significant relationship between LC and LC(M) listening conditions ($r = 0.966$).



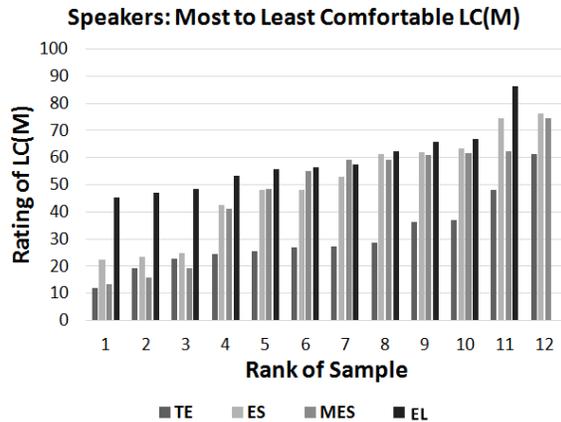
Graph 2: Means and SDs of speakers for each communication mode across both listening scenarios.

Speaker Comparison	LC p-value	LC(M) p-value
TE vs. ES	0.01	0.009
TE vs. EL	0.001	0.0007
TE vs. MES	0.012	0.027
ES vs. EL	0	0 (0.206)
ES vs. MES	0	0 (0.76)
EL vs. MES	0	0

Table 2: Independent t-tests showed TE speakers had significantly better ratings for both listening scenarios when compared to ES, EL, and MES speakers. No other significant differences were found.



Graph 3: All speakers ranked from most to least comfortable for LC. The range within each communication mode is displayed.



Graph 4: All speakers ranked from most to least comfortable for LC(M). Speakers were found to vary based on listener ratings.

Discussion

This paper sought to identify situationally-bound judgments of listener comfort for postlaryngectomy voice and speech.

This study accounted for variables that could potentially influence listeners' judgments of LC. Firstly, four randomized lists of speaker samples were created in order to minimize presentation bias. Additionally, the presentation of LC and LC(M) rating scales were also counterbalanced to minimize order effects of the judgments made.

As demonstrated by Graph 1 and Graph 2, ratings of "listener comfort" did not differ significantly for any of the speaker groups dependent upon listening scenario. This suggests that the suggestive listening scenario of either a social or telephone conversation does not impact perceptual judgments of "listener comfort". Upon analysis, significant differences were observed in listener judgments of TE speakers when compared to ES, AL and MES speakers for both listening conditions. This suggested that listeners are more comfortable listening to TE speakers than other speaker modes, regardless of the listening scenario.

Clinical Implications

It is important to note that in the current study, listeners were only provided with auditory input regardless of the listening scenario. Further research is needed that provides the visual information for the social situation scenario, as the listener would have face-to-face information in a real life scenario.

Furthermore, the auditory signal is changed when on the telephone, so that would need to be accounted for as well.

The findings from the current study demonstrated that TE speakers are considered to be more comfortable to listen to regardless of the suggestive listening scenario. These ratings of LC may influence laryngectomees' choice of postlaryngectomy communication method because listeners are more comfortable listening to TE speech. In addition, this study may also be of counselling and educational value for laryngectomees in order to establish optimal rehabilitation results.

Acknowledgments

Special thanks to Dr. Philip Doyle for his support and supervision of the study as well as the collection of voice samples. Thank you to Steve R. Cox, PhD (C) for collecting some of the voice samples, Sebastiano Failla for editing the samples, and to members of the Voice Production and Perception Laboratory at Western University.

References

- Doyle, P. C. (1999). Postlaryngectomy speech rehabilitation: Contemporary considerations in clinical care. *Journal of Speech-Language Pathology and Audiology*, 23, 109-116.
- Doyle, P. C., & Eadie, T. L. (2005). Chapter 6: The perceptual nature of alaryngeal voice and speech. In P.C. Doyle, & R.L. Keith (Eds.), *Contemporary considerations in the treatment and rehabilitation of head and neck cancer: Voice, speech, and swallowing*. Austin, TX: Pro-Ed.
- Eadie, T. L., Day, A. M. B., Sawin, D. E., Lamvik, K., & Doyle, P. C. (2012). Auditory-perceptual speech outcomes and quality of life after total laryngectomy. *Otolaryngology – Head and Neck Surgery*, 148(1), 82-88.
- Eadie, T. L., & Doyle, P. C. (2004). Auditory-perceptual scaling and quality of life in tracheoesophageal (TE) speakers. *The Laryngoscope*, 114, 753-759.
- Eadie, T. L., & Doyle, P. C. (2005). Quality of life in male tracheoesophageal (TE) speakers. *Journal of Rehabilitation Research and Development*, 42(1), 115-124.

- Evitts, P. M., Portugal, L., Van Dine, A., & Holler, A. (2010). Effects of audio-visual information on the intelligibility of alaryngeal speech. *Journal of Communication Disorders, 43*, 92-104.
- Fairbanks, G. (1960). *Voice and Articulation Drillbook*, 2nd ed. New York: Harper & Row.
- Fung, K., & Terrell, J. E. (2004). Outcomes research in head and neck cancer. *Journal of Otorhino-Laryngology and Its Related Specialties, 66*(4), 207-213.
- Hassan, S. J., & Weymuller, E. A. (1993). Assessment of quality of life in head and neck cancer patients. *Head & Neck, 15*(6), 485-496.
- Hubbard, D. J., & Kushner, D. (1980). A comparison of speech intelligibility between esophageal and normal speakers via three modes of presentation. *Journal of Speech and Hearing Research, 23*(4), 909-916.
- Most, T., Tobin, Y., & Mimran, R. C. (2000). Acoustic and perceptual characteristics of esophageal and tracheoesophageal speech production. *Journal of Communication Disorders, 33*, 165-181.
- Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America, 95*(2), 1085-1099.
- O'Brian, S., Packman, A., Onslow, M., Cream, A., & O'Brian, N. (2003). Is listener comfort a viable construct in stuttering research? *Journal of Speech, Language, and Hearing Research, 46*, 503-509.
- Robbins, J., Fisher, H. B., Blom, E. C., & Singer, M. I. (1984). A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production. *Journal of Speech and Hearing Disorders, 49*, 202-210.
- Rosen, C. A., Lee, A. S., Osborne, J., Zullo, T., & Murry, T. (2004). Development and validation of the Voice Handicap Index-10. *The Laryngoscope, 114*(9), 1549-1556.
- Terrell, J. E., Nanavati, K. A., Esclamado, R. M., Bishop, J. K., Bradford, C. R., & Wolf, G. T. (1997). Head and neck cancer - Specific quality of life: Instrument validation. *Archives of Otolaryngology-Head & Neck Surgery, 123*(10), 1125-1132.
- Yorkston, K., Beukelman, D., & Tice, R. (1996). Sentence Intelligibility Test [Speech analysis]. Lincoln, NE: Tice Technology Services.