Proceedings

Invitational Round Table
“Evidence-based Voice and Speech Rehabilitation in Head and Neck Cancer”

May 15-16, 2008
Amsterdam

Organization

Department of Head and Neck Oncology and Surgery
The Netherlands Cancer Institute, Amsterdam

Institute of Phonetic Sciences/
Amsterdam Center for Language and Communication
University of Amsterdam

Department of Otorhinolaryngology
Academic Medical Center, University of Amsterdam

Organizing Committee

Prof. Dr. Frans JM Hilgers
Prof. Dr. Ir. Louis CW Pols
Dr. Maya van Rossum
Dr. MWM van den Brekel

Venue: Netherlands Cancer Institute, Amsterdam
Program Invitational Round Table “Evidence-based Voice and Speech Rehabilitation in Head and Neck Cancer”

Thursday May 15, 2008

Chairpersons Frans Hilgers, Louis Pols

13.15: Welcome with coffee
13.45-14.45: Perceptual and acoustic analysis of TE-substitute voice and speech. Corina van As-Brooks
14.45-15.45: Objective evaluation of TE-substitute voice and speech. Ulrich Eysholdt
15.45-16.00: Tea break
16.00-17.00: Articulatory aerodynamics, contact pressures and sense of effort during tracheoesophageal speech. Jeff Searl
17.00-17.45: Poster presentations and poster visit*
   2. Early rehabilitation of Head and Neck Cancer patients at Karolinska University Hospital. Therese Engström et al.

19.00-22.00: Dinner

Friday May 16, 2008

Chairpersons Maya van Rossum, Annemieke Ackerstaff

09.00-10.00: An evidence-based rehabilitation program for tracheoesophageal speakers. Petra Jongmans
10.00-11.00: Quality of life after Total Laryngectomy. Tanya Eadie
11.00-11.15: Coffee break
11.15-12.15: Oral and oropharyngeal cancer: speech outcomes and the need to monitor change over time. Phillip Doyle

The final hour; chairpersons Frans Hilgers, Louis Pols

12.15-13.15: Where do we stand and what are the priorities for research in the near future?
13.15: Adjourn; sandwiches on the go.

All presentations 45 minutes, 15 minutes discussion
* Poster presentations: 15 minutes followed by discussion
# Table of Contents

Introduction........................................................................................................7

Perceptual and acoustic analysis of TE-substitute voice and speech. Corina van As-Brooks ...................................................................................................8

Abstract..........................................................................................................8
1. Introduction ................................................................................................8
2. Equipment and Methods............................................................................8
   2.1. Recording equipment and environment...............................................8
   2.2. Speech samples..................................................................................9
   2.3. Software programs..............................................................................9
3. Results of acoustic analyses in alaryngeal speakers...............................10
   3.1. Spectrographic analysis ....................................................................10
   3.2. Acoustic parameters..........................................................................11
4. Acoustic measures and perceptual relevance .........................................12
   4.1 Perceptual evaluation.........................................................................12
   4.2 Relationships between perceptual evaluations and acoustic measures .................................................................................................................13
5. Conclusion ...............................................................................................13
6. References ..............................................................................................14

Objective evaluation of TE-substitute voice and speech. U Eysholdt, M Schuster, J Lohscheller, F Rosanowski...........................................................15

Abstract........................................................................................................15
1. Real time endoscopy of the vibrating PE segment ..................................15
2. Quantification of speech intelligibility by means of a computerized speech recognition system.......................................................................................16
3. Rating of the laryngectomees’ quality-of-life in relation to his substitute voice ............................................................................................................16
4. References ..............................................................................................17

Articulatory aerodynamics, contact pressures and sense of effort during tracheoesophageal speech. Jeff Searl ............................................................19

Abstract........................................................................................................19
1. Introduction ..............................................................................................20
2. Patients/Materials and Methods...............................................................23
   2.1. Subjects ............................................................................................23
   2.2. Speech Stimuli ..................................................................................24
   2.3. Instrumentation and Measures..........................................................25
      2.3.1. Articulatory Contact Pressure (ACP) ..........................................25
      2.3.2. Oral Air Pressure (Po) ................................................................25
      2.3.3. Subneoglottic Air Pressure (Psub)..............................................26
      2.3.4. Sense of Effort (SOE) .................................................................28
3.1.4. Dyva phoneme intelligibility test with Speech Language Pathologists
3.2. Questionnaires and semantic scaling experiment
   3.2.1. Study-specific structured questionnaires for speakers and their relatives
   3.2.2. Semantic scaling experiment for naïve listeners
4. Discussion and Conclusions
5. Acknowledgements
6. References

Quality of life after Total Laryngectomy. Tanya Eadie

Abstract
1. Introduction
   1.1. Using the ICF as a Framework to Measure QOL
   1.2. Common Tools that Measure QOL after Total Laryngectomy
2. QOL Outcomes after Total Laryngectomy
   2.1. Effects of Laryngeal Cancer and its Treatment on QOL
   2.2. Effects of Body Functions and Structures, Activities, and Participation on QOL
   2.2.1. Functions of Eating and Swallowing
   2.2.2. Voice and Speech Function
   2.2.3. Activities and Participation and QOL
   2.3. Effect of Contextual Factors on QOL
      2.3.1. Environmental Factors: Social Support
      2.3.2. Personal Factors: Coping, adjustment, and depression
      2.3.3. Personal Factors: Gender
      2.3.4. Personal Factors: Age
3. Discussion
   3.1. Implications for Future Research and Clinical Practice
4. Conclusions
5. Acknowledgements
6. References

Oral and oropharyngeal cancer: speech outcomes and the need to monitor change over time. Phillip Doyle

Abstract
1. Introduction
   1.1. Cancer and the Oral and Pharyngeal Cavities
   1.2. The Vocal Tract
   1.3. Monitoring and Indexing Change Secondary to Cancer Surgery
2. Patients/Materials and Methods
   2.1. Procedures and Stimuli
   2.2. Data Analysis
      2.2.1. Acoustic analysis
      2.2.2. Reliability – Formant Frequencies
Evidence-based Voice and Speech Rehabilitation is of increasing relevance in Head & Neck Oncology. The number of patients requiring treatment for cancer in the upper respiratory and vocal tract keeps rising. Moreover, treatment – whether it concerns an “organ preservation protocol” or traditional surgery and radiotherapy – negatively impacts the function of organs vital for communication. A “function preservation treatment” does, unfortunately, not yet exist. This “Invitational Round Table” seeks to assemble the latest and most relevant knowledge on evidence-based voice and speech rehabilitation. Aside from the main topic (voice and speech rehabilitation after total laryngectomy), also effects on communication of the treatment of oral/oropharyngeal carcinoma will be addressed.
Acoustic analyses of postlaryngectomy voice and their perceptual relevance

Corina J van As-Brooks, PhD
Department of Head and Neck Oncology and Surgery, the Netherland Cancer Institute, Amsterdam, The Netherlands

Abstract
Despite the fact that perceptual evaluations are considered the ‘gold standard’ for evaluation of voice quality, there are some drawbacks. They are time-consuming and remain subjective. Hence the use of acoustic measures is interesting. This article describes some common pitfalls of acoustic analyses, the outcomes of acoustic analyses in tracheoesophageal speakers, and their perceptual relevance.

1. Introduction
With the availability of a variety of commercial and free software packages, acoustic analysis has become a tool that is available at any person’s fingertips. However, the acoustic analysis of alaryngeal speech is not straightforward at all and many pitfalls lie ahead. Performing the seemingly easy steps of opening the sound file, running some analyses, and reporting the results may lead to gross errors in the reported results.
The irregularities in periodicity or the total lack of periodicity in many of these voices, can easily cause mistakes in the acoustic analysis. Selection of the type of recording equipment and environment, the speech sample, the software package used, the acoustic parameters analyzed, and checking and correcting the results prior to reporting are all important aspects. In this paper all these subjects will be addressed.

2. Equipment and Methods

2.1. Recording equipment and environment
When recording alaryngeal voices for acoustic analysis, it is important to obtain recordings of good quality. Therefore, it is always recommended to investigate the current state of the art in regards to recording equipment and the microphone used. Currently, portable digital recorders such as the Edirol RW9 in combination with a good dynamic microphone (for example the Sennheiser MD421 dynamic cardioid microphone) are a good and relatively affordable option.
Recordings maybe made in so-called quite rooms that are treated for optimal acoustic damping or in soundproof booths. The latter obviously provide the highest reduction in noise from the environment.

2.2. Speech samples
The type of speech sample that is used may be a sustained vowel (usually the vowel /a/), a (nonsense) syllable or syllable sequence, (nonsense) words or sentences, read-aloud text, retold stories, monologue, dialogue, or conversational speech. The choice of material should fit the goal of the analysis, and often a variety of materials may be used. However, the limitations of acoustic analysis should clearly be kept in mind. For example, F0 and F0 based parameters such as jitter and shimmer can best be measured in the stable part of a sustained /a/ that is produced at a comfortable pitch and loudness. Performing these analyses for example in a vowel produced with the goal to reach maximum phonation time, or in running speech is not advised.

Speech samples other than sustained vowels can be useful for studying consonant or vowel characteristics, temporal, or prosodic aspects of alaryngeal speech.

2.3. Software programs
Some software packages leave more room for the investigator to change the defaults used for the analysis and adapt them for the alaryngeal speaker, other packages do not allow the investigator to make any changes or only minor changes. For the experienced investigator, the packages that allow the user to change setting are the best option. Some packages also allow the user to adjust for example pitch markers that were placed in error. Commercially available packages are not necessarily better than free packages. The software program Praat is freely available on www.praat.org and fulfills all requirements discussed above.

Figure 1. Example of pitch marker errors. From: Van As et al.(1998).
3. Results of acoustic analyses in alaryngeal speakers

3.1. Spectrographic analysis

Before carrying out any in-depth acoustic analyses, it is always worthwhile to carry out a spectrographic analysis of the speech sample. A long-term average spectrum, narrow-band spectrogram, and wide-band spectrogram are all useful means of obtaining more insight into the characteristics of the speech signal. The spectrum gives an impression of the harmonicity in the signal, the formant frequencies, and the noise distribution in the signal, the narrow band spectrogram gives an impression of the harmonics and the strength of the harmonics in the signal and can be helpful in determining F0 or checking F0 values reported by the outcomes of the more specific acoustic analyses, the wide-band spectrogram gives an impression of the articulation of consonants and vowels, formant frequencies, and temporal aspects.

The narrow band spectrum can also be used to select the part of the speech sample with the strongest harmonics for further acoustic analysis. In any case, if the spectrographic analysis shows that there are no clear harmonics in the signal, further acoustic analysis should not include any F0-based parameters.

Van As et al. (2006) describe a signal typing system for tracheoesophageal speech that can be used for this purpose.

![Example of a voice sample classified as type A. The spectrogram shows a stable signal with stable harmonics. The 100 ms selection of the spectrogram shows a clearly periodic pattern. The pitch contour shows a stable fundamental frequency (F0) that is the narrow-band spectrogram (100 ms window time). The first four harmonics are evident in the spectrum of the first second. The low-frequency noise is also visible in the spectrum of the first second.]

![Example of a voice sample classified as type B. The spectrogram shows a stable signal with stable harmonics. The 100 ms selection of the spectrogram shows a stable-gate pattern, with noise. The pitch contour shows a stable fundamental frequency (F0) that is the narrow-band spectrogram (100 ms window time). The first four harmonics are evident in the spectrum of the first second. The low-frequency noise is also visible in the spectrum of the first second.]

FIGURE 1: Example of a voice sample classified as type A. The spectrogram shows a stable signal with stable harmonics. The 100 ms selection of the spectrogram shows a clearly periodic pattern. The pitch contour shows a stable fundamental frequency (F0) that is the narrow-band spectrogram (100 ms window). The first four harmonics are evident in the spectrum of the first second. The low-frequency noise is also visible in the spectrum of the first second.

FIGURE 2: Example of a voice sample classified as type B. The spectrogram shows a stable signal with stable harmonics. The 100 ms selection of the spectrogram shows a stable-gate pattern, with noise. The pitch contour shows a stable fundamental frequency (F0) that is the narrow-band spectrogram (100 ms window). The first four harmonics are evident in the spectrum of the first second. The low-frequency noise is also visible in the spectrum of the first second.
The results of acoustic signal typing were linked to the results of an overall judgment of voice quality, and a chi-squared test demonstrated a significant relationship. Type 1 and 2 signal types were more often rated as good voices, whereas type $ signal types were more often rated as poor.

3.2. Acoustic parameters
Van As et al. (2006) also described a set of acoustic parameters that can be used in tracheoesophageal speakers. Five of them were fundamental frequency based: median fundamental frequency (F0-MED), standard deviation of fundamental frequency (F0-SD), jitter (JIT), percentage of voiced (%Voiced), and harmonics-to-noise ratio (HNR). Two of them were not: glottal-to-noise- excitation ratio (GNE), and band energy difference (BED). All parameters were calculated using Praat, and settings are described in the article.

As discussed above, one cannot expect to be able to calculate all acoustic parameters for all voice samples of laryngectomized patients. Especially not when the patients were selected randomly and not based on 'good or excellent' voice quality. Van As-Brooks et al. (2006) showed that only 77% of the voice samples could be partially or completely analyzed with respect to the pitch based parameters. In the other samples, visual inspection of the results of the signal typing showed that indeed no clear periodicity was present in the signal. Similarly, Van As et al. reported in 1998, that in about 30% of the voice samples
only very short parts could be analyzed. In a study that investigated the influence of stoma occlusion on voice quality (Van As et al., 1998), 21 patients produced three sustained /a/’s each. Of these samples, one second was selected from the middle of the vowel for further analyses. Complete analysis of all samples was only possible in 62% of the patients.

Keeping these limitations in mind, results of acoustic analyses of alaryngeal speech, should always be interpreted with caution. An article reporting these types of results without mentioning any limitations regarding the use of frequency dependent acoustic parameters, is suspicious for overlooking these difficulties and simply reporting the output provided by the software.

A variety of studies is available reporting results of acoustic analysis in alaryngeal speech. Robbins et al. (1984) used measures of frequency, perturbation, and duration in tracheoesophageal, esophageal, and normal speech and found that tracheoesophageal speech is more similar to normal speech than esophageal speech for frequency and duration variables, and that TE speech is louder than esophageal speech. Debruyne et al. (1994) used frequency and perturbation measures, and also frequency-independent measures (spectral slop, harmonic prominence) and found that in TE speech more often a fundamental frequency could be detected and that there was more often a tendency to clearly defined harmonics. Jitter and shimmer values in TE speech were more close to normal speech than those in esophageal speech. Trudeau and Qi (1990) specifically studied male versus female tracheoesophageal speech and concluded that the speech of the women, including pitch, closely resembled that of the men. Van As et al (1998) reported on TE speech versus normal speech and found that almost all parameters reported on were significantly worse in the TE speakers, and also a much larger variability was found. These results were confirmed in a later study that incorporated the use of signal typing (Van As et al., 2006).

4. Acoustic measures and perceptual relevance

4.1 Perceptual evaluation

Although the acoustic analyses are the main focus of this chapter, they should always be relevant in relation to perceptual evaluations. Obviously, a number generated by an acoustic analysis is only relevant when it expresses a perceptual characteristic of the voice quality.

Similar to the acoustic analyses, perceptual evaluations in laryngectomized speakers are not as straightforward as those in normal speakers. The aspects of voice quality that are usually judged for normal pathological laryngeal voices, such a breathiness or hoarseness do often not apply for alaryngeal voice. Conversely, aspects of alaryngeal voices like bubbliness or hypertonicity are not relevant for normal pathological laryngeal voices. Therefore, general perceptual evaluation methods for pathological laryngeal such as GRBAS are often not useful for perceptual evaluation of alaryngeal voice; only the judgment of G (Grade) might be relevant.

In the PhD thesis of Van As (2001) and in Van As et al. (2003) an overview is given that discusses the need for intra- and inter-rater reliability measures; the
factors that influence reliability; the influence of speech materials used; and the various method available. The studies have resulted in a recommendation of the use of naïve raters for gaining insight in daily communication, and trained experts (speech-language pathologists) for gaining insight in clinical performance and for comparison with other clinical measures. The proposed scales for naïve raters are: ugly-beautiful, deviant-normal, low-high, and deep-shrill and the proposed scales for trained experts are: deviant-normal, ugly-beautiful, breathy-not breathy, hypotonic-not hypotonic, low-high, deep-shrill, slow-quick, and dragging-brisk. Additionally, a judgment of overall voice quality by trained experts, gives a good impression of the voice.

4.2 Relationships between perceptual evaluations and acoustic measures
A detailed description of these relationships can be found in Van As (2001) and Van As-Brooks et al. (2006). The acoustic signal types show a significant relationship with overall voice quality, indicating that acoustic signal typing is relevant in relation to overall voice quality.

The correlations between the results of the acoustic measures and the perceptual scales were moderate to strong, showing their relevance with respect to tracheoesophageal voice quality. The complementing scales ugly-beautiful and deviant-normal showed significant correlations with the F0-SD, HNR, %Voiced, and BED. The complementing scales breathy-not breathy and hypotonic-not hypotonic showed significant correlations with HNR, %Voiced, GNE, and BED. The complementing scales low-high and deep-shrill showed a significant correlation with F0-MED. The scales slow-quick and dragging-brisk did not show any relationship with the acoustic measures, which is not surprising as no temporal measures were included in the acoustic analyses.

5. Conclusion
If used correctly, acoustic analyses are a good tool for voice quality evaluation of alaryngeal speech. Especially the construction of a spectrum and/or a narrow-band spectrogram gives a very good first impression of the acoustic content of the voice. Based on these, further acoustic analyses may be carried out. Due to the difficulty in analyzing voices with irregular or absent periodicity, papers reporting on acoustic analysis should always be read critically when it comes to the methodology that was used. Results of papers that presented reliable results show that tracheoesophageal speech is highly variable and remains deviant from normal speech.
6. References


Abstract

Three studies are presented which aim at rater-independent methods to evaluate voice quality, speech intelligibility and the benefit of rehabilitation of the patients: (1) real time endoscopy of the vibrating PE segment, (2) quantification of speech intelligibility by means of a computerized speech recognition system and (3) rating of the laryngectomees’ quality-of-life in relation to his substitute voice.

1. Real time endoscopy of the vibrating PE segment

Objectives: Evaluation of the vibration pattern of the substitute voice generator of patients who have undergone laryngectomy. For automatic quantification of the oscillations of the pharyngoesophageal (PE) segments, image processing of digital high-speed video sequences is applied.

Patients and methods: Endoscopic recordings were taken of 10 men who underwent laryngectomy (mean +/- SD age, 61.5 +/- 5.2 years) during sustained phonation of a vowel using a 90 degree endoscope coupled to a high-speed camera. An image-processing algorithm was developed to automatically define the pseudoglottis in each recording and track its movements.

Results: In a first step, the high-speed recordings of the substitute voice generator have been replayed in slow motion and visually categorized. The forms and oscillation characteristics of the pseudoglottides varied considerably: 3 pseudoglottides were circular, 6 were split shaped, and 1 was triangle shaped. Quasi-periodic openings and closings were observed and automatically detected by the described algorithm in each recording independently from quality of the recording and from morphologic and oscillation characteristics of the PE segment. The frequencies of the extracted oscillations of the pseudoglottides correspond to the structure of the acoustic signals.

Conclusion: Automatic image processing of PE segments derived from high-speed endoscopic recordings enables the detection and quantification of the substitute voice generator’s oscillations in high temporal resolution. These data directly prove that the detected pseudoglottis is the source of the substitute voice. Close relations between substitute voice and functional properties of the PE
segment exist. In the future, these data will be interpreted by applying biomechanical models of the PE segment. Presumably, results may help to optimize surgical and adaptive procedures for specific substitute voice restoration.

2. Quantification of speech intelligibility by means of a computerized speech recognition system

Objectives: Substitute speech after laryngectomy is characterized by restricted aero-acoustic properties in comparison with laryngeal speech and has therefore lower intelligibility. Until now, an objective means to determine and quantify the intelligibility does not exist although the intelligibility can serve as a global outcome parameter of voice restoration after laryngectomy.

Patients and methods: An automatic speech recognition system was applied on recordings of a standard text read by 18 German male laryngectomees with tracheoesophageal substitute speech. The system was trained with normal laryngeal speakers and not adapted to severely disturbed voices. Substitute speech was compared to laryngeal speech of a control group. Subjective evaluation of intelligibility was performed by a panel of 5 experts and confronted to automatic speech evaluation.

Results: Substitute speech showed lower syllables / sec and lower word accuracy than laryngeal speech. Automatic speech recognition for substitute speech yielded word accuracy between 10.0% and 50% (28.7% ± 12.1%) with sufficient discrimination. It complied with experts’ subjective evaluation of intelligibility. The multi-rater kappa of the experts alone differed not from the multi-rater kappa of experts and recognizer.

Conclusion: Automatic speech recognition serves as a good means to objectify and quantify global speech outcome of laryngectomees. For clinical use, the speech recognition system will be adapted to disturbed voices and can also be applied in other languages.

3. Rating of the laryngectomees’ quality-of-life in relation to his substitute voice

Objectives: Health-related quality of life (QoL) and subjective health have become popular constructs for the evaluation of both efficacy and efficiency of diagnostic and therapeutic procedures in medicine. QoL is considered a multidimensional construct encompassing physical, mental and social facets of life. It is an accepted outcome parameter not only in international classification systems such as ICIDH and ICF, but as well in clinical guidelines and disease management programs. Measuring quality of life allows for comparison of different diseases though it certainly lacks disease specific aspects. Thus, it has to be assumed that in patients with distinct functional deficits QoL cannot cover
all aspects that are important for the individual patient. This study focusses on laryngectomees and their self-evaluation of post-laryngectomy speech. It is well known that these patients experience a decreased QoL compared to patients after partial laryngectomy or healthy persons. In this study, the impact of voice restoration on the laryngectomees’ QoL was evaluated.

Patients and methods: In 20 male laryngectomees aged 62 ± 8 years, relations between QoL and voice handicap were evaluated using two instruments as proposed in the international literature, i.e. the Short-Form Health Survey (SF-36) and the Voice Handicap Index (VHI). All patients had successfully been using tracheoesophageal substitute voice for at least one year. Complete data sets were available from all patients. Data were analysed using Microsoft ExcelR and Sigma PlotR, Jandel Corp. software packages.

Results: Results of both the SF-36 and the VHI reveal wide interindividual ranges. There are statistically significant correlations (p < 0.05) between the SF-36’s scales General Health and Vitality and the VHI, whereas no correlations were found between social and psychological scales and the voice handicap.

Conclusion: Wide ranges of the data obtained reveal that obviously both health related quality of life and voice handicap are not affected in a group specific way. So, both tests are of clinical value to depict individual aspects of wellbeing after laryngectomy. The combination of VHI and SF-36 illuminates correlations between general and special subjective aspects. Strikingly, the laryngectomees’ social and psychological status is not related to their voice handicap. Data allow for conclusion that laryngectomees’ substitute voice problems do not affect social aspects of quality of life in a disease specific way, at least as long as voice restoration was successful at all.

Key words: laryngectomy, high speed endoscopy, substitute speech, automatic speech recognition, speech intelligibility, Quality of life, voice handicap index

4. References


Articulatory Aerodynamics, Contact Pressures and Sense of Effort During Tracheoesophageal Speech

Jeff Searl

Hearing and Speech Department, University of Kansas Medical Center, Kansas City, Kansas, United States of America

jsearl@ku.edu

Abstract

Objectives: The purposes of this study were to: 1) describe sense of effort (SOE) associated with speech, oral air pressure (Po), articulatory contact pressure (ACP), and subneoglottal pressure (Psub) in tracheoesophageal (TE) speech, and 2) evaluate the relationships between these measures. The intent was to inform about articulatory alterations in TE speech with a future goal of guiding therapeutic interventions that serve to balance the physical demands of producing intelligible TE speech with the perceived work of generating the speech.

Patients/Materials and Methods: Thirty four TE speakers provided ratings of SOE. ACP, Po and Psub were measured during production of alveolar consonants.

Results: Po, ACP, Psub, and SOE were higher than what is reported for non-laryngectomized speakers. The primary locations identified by subjects as having increased SOE were the lips and tongue, followed by the brain (i.e., cognitive effort), voice, and lungs. Individuals who reported elevated SOE had higher ACP and Po than those who did not report an increase in SOE. SOE, ACP and Po were all moderately to strongly correlated with each other.

Conclusions: More than half of the TE speakers reported elevated SOE. ACP, and also Po appear to be possible physiologic correlates of internal SOE. These parameters could serve a useful therapeutic function as feedback to individuals with a goal of optimizing physical effort while maintaining the perceptual integrity of the consonants being produced.

Keywords: oral pressure, contact pressure, articulation, effort
1. Introduction

Two fundamental alterations to speech production that result from total laryngectomy are the removal of the voice source (i.e., the vocal folds) and separation of the upper from the lower airway with the trachea routed to a stoma at the base of the neck. These anatomical alterations create two primary issues to address in the alaryngeal speech rehabilitation process. First, a means of introducing a vibration into the vocal tract must be established. Second, a mechanism for creating airflows and pressures within the oral cavity and pharynx is needed to allow production of other important sounds for speech such as bursts and frication which are integral to the production of stops, fricatives and affricates.

In general, the tracheoesophageal (TE) speech production process more closely parallels the non-laryngectomy speech process than does esophageal or electrolaryngeal speech. In TE speech, a connection between the lower and the upper airway is re-established through a surgically created fistula in the common wall between the trachea and esophagus (see van As and Fuller [2006] for a brief historical review). Air flowing through a one-way valve placed in the fistula can set tissue in the upper esophagus and lower pharynx (pharyngo-esophageal [PE] segment) into vibration (Singer & Blom, 1980). As such, TE speech utilizes a voice source comprised of biologic tissue inherent to the speaker, unlike artificial larynx speech which utilizes an external device and a mechanical vibration that is external to the body. Additionally, TE speech utilizes pulmonary airflow, unlike esophageal and electrolaryngeal speech, to drive neoglottal vibration. This pulmonary air source provides a large volume of air that provides certain advantages over esophageal speech such as increased phrase durations, fewer pauses, and potentially louder and more stable voice (Max, Steurs & DeBruyne, 1996; Pauloski, 1998; Robbins, Fisher, Blom, Singer, 1984). Additionally, the larger volume of air provides an opportunity to recreate a more consistent and sustained air stream through the pharynx and oral cavity that can be utilized for creating pressure build-up for bursts and turbulent air flow for frication. Without this connection to the lower airway, artificial larynx and esophageal speakers must generate bursts and frication by other means, most commonly by using the lips and tongue to compress the air within the closed oro-pharyngeal space and then quickly releasing it for a burst or releasing and simultaneously moving an articulator to generate a short period of frication.

Not only does the TE speech process more closely parallel non-laryngectomy speech, the speech product has also been found to be a closer match than esophageal of electrolaryngeal speech. For example, although dependent on an individual speaker, TE speech is generally considered to be closer to laryngeal speech in terms of voice quality (Pindzola & Cain, 1988; Williams & Watson, 1987), the mean fundamental frequency (Robbins et al, 1984), pitch variability (Trudeau, 1994), intensity (Max, Steurs & DeBruyne, 1996), and speaking rate.
(Pauloski, 1998). TE speech can provide a functional communication method for those who have had a laryngectomy and it can do so in a relatively short time frame. Despite success with TE speech, however, it has consistently been found to be less intelligible than speech from non-laryngectomized speakers (e.g., Ainsworth & Singh, 1992; Doyle, Danhauer & Reed 1988). Reasons for the reduced intelligibility could vary widely across speakers. The vast majority of research and clinical efforts regarding TE speech have focused on understanding and characterizing the new voice source and seeking means of improving it (e.g., Robbins et al, 1984; van As et al, 1999; van As et al, 2005). Such efforts have high value and have contributed much to maximizing the TE rehabilitation process. Investigations of non-voice source related aspects of TE speech, such as articulation, have received significantly less attention, although it is increasingly clear that other speech behaviors not strictly related to the voice source also can be altered. For example, altered velopharyngeal activity (Searl & Evitts, 2004), prolonged duration of consonants and vowels (Robbins et al, 1984; Searl & Carpenter, 2001), and increased oral air pressure (Po) on stops and fricatives (Saito, Kinishi & Amatsu, 2000) have been reported.

There is sparse empirical data regarding the articulatory adjustments that occur during TE speech. If one goal of the laryngectomy rehabilitation process is to maximize speech intelligibility, it would be useful to understand the TE articulatory behaviors so that appropriate and targeted therapeutic interventions could be designed if needed. The current study, as well as two prior ones (Searl 2002, 2007) were prompted principally with this overall goal in mind. The research focus on articulatory forces and aerodynamics had its genesis from four information sources. The first were general references in clinical training texts regarding the possibility of needing to increase articulatory precision when using TE speech (Gress & Singer, 2005; van As & Fuller, 2007). Second were a set of five articles in recent years noting increased Po on stops and fricatives in TE speech (Motta, Galli & DiRienzo, 2001; Saito et al, 2000; Searl, 2002, 2004, 2007). None of these articles could elucidate specifically why Po was elevated so significantly (on average approximately 2-4 times higher than in laryngeal speech). Searl (2004) speculated that the elevated Po might be an outcome of an intuited or trained response focused on production of stronger bursts and frication to provide a salient (and perhaps unambiguous) acoustic cue to the listener trying to decode a non-normal speech signal. Increased Po is associated with a stronger burst release in non-laryngectomized speakers. Alternatively, the elevated Po might simply be an obligatory outcome of the altered neoglottal aerodynamics. The PE segment has a higher resistance to airflow than do the true vocal folds (Weinberg et al, 1982). This creates a situation where the voice source driving pressure for the laryngectomy is notably higher than that for the non-laryngectomized speaker. The higher subneoglottal pressure could conceivably translate into a strong Po impulse once the PE segment opens (either for voicing on an adjacent voiced phoneme or if the PE segment opens in a devoicing gesture for a voiceless stop consonant). The current study incorporates measurement of subneoglottal air pressure simultaneously with Po
and articulatory contact pressure (ACP) to better understand the possible relationship among these parameters. Regardless of whether the elevated Po is a trained/intuited response or a direct outcome of neoglottal aerodynamics, it does beg the question, what oral articulatory behaviors are enlisted or required to contain the elevated Po?

The third piece of information prompting the current study was a recent report from this lab indicating that lip ACP during bilabial stop consonants in TE speech were 2-3 times greater than in laryngeal speakers (Searl, 2007). From a clinical training perspective, it is of interest to determine whether the ACP needs to be this high; that is, could adequate Po be generated to result in perceptually sufficient bursts and frications without expending more physical resources than necessary. The issue of ‘excess effort’ involved in articulation has not been explored systematically relative to TE speech although the Po data along with the historical approach of training exaggerated articulation in alaryngeal speakers make this a potentially relevant area of interest. In the TE literature, elevated ‘effort’ has been presented in relation to respiratory activity and neoglottal voice generation (e.g., Monahan, 2005; van As & Fuller, 2007) in TE speakers who have strained voice or aphonia. Anecdotally, there are reports within the clinic that speech is at times fatiguing and requires more ‘work’ based on the subjective descriptions of alaryngeal speakers in general. Interestingly, when comparing the magnitude of the Po to the ACP during TE speech in the Searl (2007) study, it is clear that ACP is substantially greater than the Po that is being contained; this held true for both TE speakers and nonlaryngectomized individuals. This opens the possibility that individuals might be trained to reduce ACP (and perhaps reduce the physical work involved in TE articulation), but still have the ability to contain sufficiently elevated Po to generate a strong burst or frication.

One final, related line of thinking contributed to the development of the current series of studies evaluating TE articulation. The Hypo-Hyper Theory of Speech (or Theory of Adaptive Variability) posits that an individual modifies speech behaviors along a continuum depending on the speech situation (Lindblom, 1990; Perkell, Zandipour, Matthies & Lane, 2002). Hyper-articulated or ‘clear’ speech may be utilized in situations where particularly clear speech is desired (e.g., lecturing) or when intelligibility or audibility is at risk (e.g., noisy environment, impaired hearing of the listener). Increased physical effort during hyper-articulation is hypothesized (Lindblom, 1990), although physiological evidence of this has not been sought. The lip ACP data from Searl (2007) might be evidence of hyper-articulated speech although this currently is the sole report on the matter. Other studies documenting elevated Po and less coarticulation in TE speech (as inferred from velopharyngeal aerodynamics [Searl, 2004]) also could be consistent with the interpretation that TE users have modified speech behaviors along a continuum more toward hyper-articulated speech, perhaps to maintain intelligibility. In this case, it may be clinically appropriate to determine the extent to which additional articulatory ‘effort’ is required to maintain an acceptable level of intelligibility. Economy of effort is an under-riding theme in the
hypo-hyper theory of speech; the question would be whether TE speakers are being economical in their articulatory adjustments or are they over- or under-compensating.

The overall goal of this study was to better understand consonant production in TE speech in terms of the articulatory effort and air pressures that are generated. The concept of ‘effort’ is addressed by quantifying the actual ACP generated between articulators during speech and also by global ratings of speech effort from the TE speaker. Sense of effort (SOE) ratings relative to speech production have been used with both normal and disordered speakers for various purposes (Solomon, 2000; Solomon, Robin & Luschei, 2000). SOE ratings have not been reported previously for TE speakers although Searl (2007) offers anecdotal comments in his discussion that some study participants reported varying levels of fatigue or increased ‘work’ to talk. Subneoglottal air pressure measurements are included to allow for more careful interpretation of the tongue-palate contact and Po data.

This study had the following purposes. 1) Describe ACP in TE speech for a set of lingu-alveolar consonants. At present, only data for the lips have been reported. 2) Describe Po and subneoglottic pressure (Psub) during TE speech for a set of lingua-alveolar consonants. Data for both measures have been reported; as such, the current data serve to broaden the data sets available in the literature. 3) Provide the first systematic description of the SOE related to TE speech production. This is to include global ratings of speech effort as well as an indication of the locus (i) of the effort throughout the thorax, neck and head. It is not clear at the moment whether elevated SOE is a phenomenon that rises to a level of significance necessitating attention by researchers or clinicians. 4) Evaluate the strength of the relationships among Po, Psub, ACP, and patient self-ratings of SOE. A strong relationship between Po and ACP has been found previously for the lips, but no data are available for the tongue. Inclusion of Psub will allow more informed interpretation of the Po data and the role that neoglottic alterations compared to oral articulatory changes might be playing in the Po that is measured. Inclusion of the SOE data within the set of planned correlations may allow identification of speech production parameters that vary in relation to internal speech effort state and which could serve as a focus in therapy if the physical expenditure for articulation is inordinately high. Overall, better understanding of the relationships between ACP, Po, Psub and SOE could shed light on the need (or lack thereof) to train higher/lower contact pressures in an effort to optimize both costs (sense of effort) and benefits (intelligible speech).

2. Patients/Materials and Methods

2.1. Subjects

Thirty four individuals who had a laryngectomy and used TE speech participated in this study. The group ranged in age from 46-71 years of age (mean: 61.4 yrs)
and was comprised of 22 males (65%) and 12 females (35%). All speakers met the following criterion: 1) TE speech was the primary mode of communication (patient report); 2) >4 months post laryngectomy (range: 4-38 months); 3) functional hearing ability (hearing screening, best ear, aided allowed); 4) normal speech prior to laryngectomy (patient report); and 5) negative history for stroke, head injury or other neurological disease that might impact speech. Two certified SLPs estimated TE speech intelligibility using a single words and sentence protocol (Yorkston & Beukelman, 1984). Intelligibility scores ranged from 80-97%. The SLPs also rated overall speech proficiency on a 5-point scale with 1=poor and 5=excellent. Ratings for each speaker averaged 4 or greater for all subjects suggesting a fairly highly proficient group of TE speakers.

Only those individuals with a standard laryngectomy were included (i.e., those with surgeries that included pharyngectomy, esophagectomy or glossectomy were excluded). Primary TE puncture was done for 25 of the subjects (74%). Radical neck dissection was completed unilaterally in 11 speakers and bilaterally in 7 others. Radiation therapy was completed by 65% of the participants (9 subjects pre- and 12 subjects post-laryngectomy). Cricopharyngeal myotomy at the time of total laryngectomy is standard in our facility and so 100% of the subjects were known to have some degree of surgical myotomy.

Blom-Singer voice prostheses were utilized by 22 speakers (65%); all of these were indwelling, 20Fr prostheses. Those using smaller diameter prostheses were excluded because the smaller flange size on the 16Fr devices made the prostheses modifications for subneoglottic pressure measurement difficult (see 2.3.3). The remaining 12 speakers used a Provox®2 prosthesis; again, those who may otherwise have qualified for the study, but who used a 17Fr Provox NID were excluded because of prosthesis modification issues. Twenty four speakers (~70%) used digital occlusion to produce voice; of these, 20 occluded using an HME and 4 used a digit directly over the stoma. The remaining 10 speakers used some type of handsfree valve during speech.

All had undergone speech therapy following laryngectomy per patient report, but varied widely in terms of the number of sessions (range: 2-16 visits) and focus of therapy. In general, however, the therapy focused on prosthesis management, voice generation, and use of HME and handsfree devices. Seven of the 34 used a handsfree valve.

2.2. Speech Stimuli

Consonant vowel (CV) syllable strings were constructed using /t/, /d/, /s/, /z/ and /n/ and the vowels /i/, /u/, and /a/ for a total of 15 CV constructions. Each syllable was produced as a 5-syllable string (e.g., tatatatata/) on one breath at comfortable rate and loudness. Three repetitions of each syllable series were produced with stimuli fully randomized for each speaker.
2.3. Instrumentation and Measures

2.3.1. Articulatory Contact Pressure (ACP)

ACP was measured using an Entran EPI-BO flatline transducer (Entran, Fairfield, NJ, USA) mounted on a thin palatal appliance as described in Searl (2003). Briefly, the transducer is 0.5mm thick with a 2.0mm diameter sensing surface at one end of a 2.3mmx6.5mm housing unit and it has response characteristics that have been found to be acceptable for measuring contact pressure during speech (see Searl, 2003; Searl, 2007; Hinton & Luschei, 1992; Thompson et al, 1997). A stone dental casting is made for each speaker. This casting is used to create a 0.5mm thick palatal appliance made from flexible thermo-forming coping material. The Entran transducer is mounted on the palatal mold on the alveolar ridge in midline at a location that yields the highest and most consistent pressure recordings during the 60 minute accommodation period that was imposed to allow speakers to adjust to the presence of the palatal appliance in the mouth (see Searl et al, 2007 for details regarding speaker adaptation to this device). Figure 1 shows the transducer, palatal appliance and also the oral pressure sensing tube (describe below). The transducer signal output was amplified and routed to one channel of a PowerLab 8SP digital recording device (ADInstruments, Colorado Springs, CO, USA). The PowerLab hardware and software system was used for signal conditioning (50Hz low pass filtered; 20kHz digitization, 16 bit precision). See Figure 1 which shows the transducer mounted on the thin palatal appliance.

2.3.2. Oral Air Pressure (Po)

A polyethylene tube (i.d.=2.1mm, o.d.=3mm) was custom molded to run in the corner of the mouth, through the buccal-gingival sulcus to the last maxillary molar where the tube was bent to project to the midline of the hard palate. The tube was attached to the palatal appliance at the bend around the molar and along the course of the projection to the midline of the hard palate. This arrangement was used in order to avoid interfering with anterior tongue movements involved in production of the alveolar consonants. In addition, it was important to insure that the opening of the tube tip was parallel to the oral airstream that would be present during /s/ and /z/ in order to avoid spurious pressure reading of maximum Po. Figure 1 depicts the Po tube attached to the palatal appliance. The tube was connected to a Setra #239 differential pressure transducer (Foxborough, MA, USA) and the output was routed to a second channel of the PowerLab set-up. The Po signal was also low pass filtered at 50kHz to eliminate the voicing component. Both the Po and the contact pressure transducers were calibrated prior to speech recordings from each subject as previously described (Searl, 2003).
2.3.3. Subneoglottic Air Pressure ($P_{sub}$)

To measure air pressure in the upper esophagus, each speaker’s voice prosthesis was modified to accommodate the presence of a polyethylene tube as shown in Figure 2a-b. A small incision was made through the tracheal and esophageal flanges at the 180° mark (i.e., 6 o’clock) so that the sensing tube was run through the flanges on the undersurface of the barrel of the prosthesis. A small amount of adhesive was placed around the tube at each flange to hold it in place and to block air from passing through the incisions. In this way, the sensing tube (1.6mm i.d., 2.2mm o.d.) did not occupy any portion of the voice prosthesis through which air was flowing from the trachea to the esophagus and also allowed the one-way valve to remain fully functional. The tip of this tube projected into the esophagus approximately 2-3mm; the tip was cut at a 90° bevel. An HME base plate was attached to the neck and an In-Health manometer attachment was placed in the collar of the base-plate. The polyethylene tubing was routed out of the stoma through the side port of the manometry collar attachment as depicted in Figure 2b. Again, a small amount of adhesive was applied at the site where the tube passed thru the manometry attachment. The polyethylene tube was then attached to a second differential pressure transducer (Setra #239) and complete stoma coverage could be obtained either with a digit (over an HME or directly over the housing unit itself) or with a hands-free valve in place. Each speaker used whichever method of stoma occlusion that they typically used during TE speech production. This left 4 subjects who typically used direct digital occlusion over the stoma (i.e., non-HME/valve users) who were forced to use digit occlusion over the housing unit opening. They were allowed as much time as needed to experiment with digit coverage using this set-up prior to data recording.
After initial testing, it was determined that a three-way stop-cock valve needed to be positioned in the tubing line after it exited the housing unit collar in order to allow for suctioning through the tube in the event that saliva obstructed the tube tip sitting in the esophagus (as indicated by a baseline shift in pressure being recorded). The investigator was careful to monitor this baseline throughout recording and apply small suction as needed to clear the line. It was also noted that a saliva swallow frequently was successful at clearing the line when it appeared to be plugged as indicated by a return to baseline.

Insertion of the prosthesis was complicated slightly by the presence of the tube, but generally was accomplished without difficulty. The investigator inserted all of the modified prostheses with the polyethylene tube already attached. For the Provox valves, it was possible to use the inserter tool as described by the manufacturer, although passing the modified prosthesis through the inserter did require additional force given the added width of the prosthesis barrel. For the Blom-Singer valves, the gel cap insertion procedure was used as described by the manufacturer. There was sufficient room in the gel cap to accommodate the presence of the polyethylene tube, although the cap could not extend down over the barrel of the prosthesis as far as it normally would if the tube was not present. For the first 12 subjects, flexible fiberoptic nasopharyngoscopy was completed with the scope passed into the upper esophagus. This was done to assess the placement of the modified prosthesis and to determine whether the polyethylene modification survived the insertion process as planned. In no instance was any difficulty noted and the tube appeared to be firmly in place. For the remaining subjects, endoscopy was utilized if there was some question about the insertion process or outcome, but typically this was not necessary. Prior to insertion a fine-tip permanent marker was used to place a mark on the polyethylene tube at the point that it passed out of the tracheal flange of the prosthesis. By looking for the mark on the tube, the investigator could tell whether the tube had dislodged forward or backward relative to the original placement and adjustments could be made as needed. With the modified prosthesis in situ, the neck was prepped for placement of a base unit. The polyethylene tubing was slid through the slit in the collar of the base plate and the base plate was attached to the neck. As with the
other transducers, the Psub transducer was calibrated prior to each subjects recording session.

2.3.4. Sense of Effort (SOE)

Sense of effort was reported in two ways. First, each subject was asked to rate the overall sense of effort that they perceived during typical speech production using a 5-point scale. Each scale point was labeled as follows: 0 = “no extra effort,” 1 = “minimally effortful”, 2 = “mildly effortful”, 3 = “moderately effortful”, and 4 = “severely effortful”. They were asked to consider this as an overall rating of effort and to not think about when their speech is at its best or its worst. They were also asked to consider this a rating of “physical effort” to produce speech rather than “mental” or “psychological” effort. The rating was used for the correlational analyses in this study. For descriptive purposes, a second SOE report was gathered that forced subjects to identify and rank the locations of the SOE in the body. They were shown a lateral schematic of a torso, neck and head that depicted internal structures such as the lungs, top of the esophagus, throat, palate, tongue and lips. These were labeled on the schematic. A brain was also drawn in the schematic within the skull. Subjects were asked to place marks on the schematic where they perceived effort when speaking. They were told that they could place as many or few marks as they wished. If they made more than one mark on the schematic, they were instructed to place a number on each mark to indicate where they perceive the greatest effort (labeled #1), next greatest effort (#2), and so forth.

2.4. Procedures

Participation required two visits. On the first, subjects provided written consent to be in the study and then completed a history form and had their hearing screened. An audio recording was made as they completed a word and sentence intelligibility test for later scoring by two certified SLPs. A dental casting of the upper arch was then made following standard dental casting procedures; the cast was used to create the palatal appliance described above. Finally, they rated their overall SOE related to speech and marked the SOE figure to identify the loci of the effort. Subjects returned for a second visit 2-4 weeks after the first. Their first task on the return visit was to repeat the SOE ratings (overall rating and loci of effort markings) so that the stability of such ratings could be evaluated. They then wore the palatal appliance for one hour continuously. The contact pressure transducer and oral pressure tube were attached to the appliance throughout the accommodation period. During that hour, they engaged in casual conversation with a research assistant to help them accommodate to the appliance. Following the accommodation period, subjects sat in front of a computer screen that was used to display the stimuli in random order. On each screen was shown a reminder for them to say the stimulus on one breath using their typical pitch and loudness. The rate of presentation of the stimuli on the computer screen was under the control of the investigator to allow sufficient pause time between
productions or to ask for a repeat if a misreading or other anomaly occurred (e.g., too fast, soft/loud, etc.). The oral and subneoglottic air pressure baselines were monitored continuously by the investigator to look for drifts that might suggest obstruction of the tubing and small amounts of suction were applied as needed to clear the saliva.

For the analysis, peak pressures (contact, Po, and Psub) were taken from the central syllable of each syllable train. For each subject, the mean pressure (contact, Po, Psub, respectively) was calculated from the three repetitions produced by the speaker and this average pressure was utilized in the analyses below. Ten percent of the samples were measured a second time by the investigator and also by a trained laboratory assistant to evaluate measurement reliability. Correlations between the investigators first and second measurements were .96, .94, and .99 for ACP, Po and Psub, respectively. Paired t-tests for each of the three pressure measures were not statistically significant (p>.05 for each test) indicating good intra-measurement agreement. Likewise, for inter-measurement agreement, correlations between the investigator’s measurements and the trained assistants were high (.93, .93, and .97 for ACP, Po and Psub) and the paired t-tests were not statistically significant at the .05 alpha level.

3. Results

3.1. Data Collapse

Preliminary analysis was completed to determine whether vowel context influenced any of the three pressure measurements. A Consonant (5) x Vowel (3) ANOVA was calculated for each pressure measurement. None of the three had a statistically significant main effect of Vowel (ACP: F=2.11, p=0.74; Po: F=1.62, p=0.66; Psub: F=2.08, p=0.71). Subsequent analysis was completed after collapsing data across vowel contexts when evaluating differences in the pressure measures as a function of the consonant and when completing the correlation analyses.

3.2. Pressure Measurements

3.2.1. Articulatory Contact Pressure

Descriptive statistics for ACP values are in Table 1. A one-way analysis of variance for repeated measures was calculated to evaluate differences in contact pressures across the set of consonants. The computed F value of 1.94 was not statistically significant using an alpha level of .05 (p=0.62) indicating no difference in tongue-palate contact pressures across the set of five consonants.
Table 1. Means and standard deviations for contact pressure, oral air pressure, and sub-neoglottic pressure for each consonant. All pressures are reported in kPa.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>/t/</th>
<th>/d/</th>
<th>/s/</th>
<th>/z/</th>
<th>/n/</th>
<th>Grand mean (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Oral Pressure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.46</td>
<td>1.34</td>
<td>0.90</td>
<td>0.87</td>
<td>0.46</td>
<td>1.01 (0.31)</td>
</tr>
<tr>
<td>(sd)</td>
<td>(0.45)</td>
<td>(0.33)</td>
<td>(0.31)</td>
<td>(0.27)</td>
<td>(0.21)</td>
<td></td>
</tr>
<tr>
<td><strong>Contact Pressure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>9.4</td>
<td>8.1</td>
<td>7.6</td>
<td>7.3</td>
<td>8.7</td>
<td>8.2 (3.7)</td>
</tr>
<tr>
<td>(sd)</td>
<td>(3.8)</td>
<td>(3.9)</td>
<td>(2.9)</td>
<td>(3.3)</td>
<td>(4.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Sub-neoglottic Pressure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.47</td>
<td>3.91</td>
<td>4.17</td>
<td>3.18</td>
<td>3.26</td>
<td>3.60 (1.33)</td>
</tr>
<tr>
<td>(sd)</td>
<td>(1.18)</td>
<td>(1.48)</td>
<td>(1.49)</td>
<td>(0.93)</td>
<td>(1.57)</td>
<td></td>
</tr>
</tbody>
</table>

3.2.2. Oral Air Pressure

Table 1 includes means and standard deviations for Po across the five consonants. A one-way ANOVA for repeated measures resulted in an F of 61.09 and a p value of 0.008. This indicated that there were differences in Po across the consonant set. Post hoc comparisons (Tukey’s honestly significant difference test) were completed and indicated that Po for /n/ was significantly lower than that of the other 4 consonants. Values for /s/ and /z/ did not differ from one another but were both significantly lower than the Po for /t/ and /d/. Po for /t/ and /d/ did not differ significantly. Overall, the results suggest lowest pressure on the nasal phoneme, highest pressure on the stop consonants and intermediate pressure on the fricatives.

3.2.3. Subneoglottic Air Pressure

Subneoglottic pressure means and standard deviations are in Table 1. A one-way ANOVA for repeated measures was computed. The F value of 1.08 was not statistically significant (p=0.833) indicating that pressures did not differ significantly across stimuli.

3.3. Sense of Effort

3.1.1. Consistency of Ratings

A paired t-test was computed using each subject's ratings of overall SOE at the first and second visits to evaluate the consistency of such ratings. The resulting t-value of 1.33 was not statistically significant (p=0.72). This suggests that the overall SOE rating was consistent within the time frame of the study.
3.3.2. Magnitude of Overall SOE

Fifteen subjects (44% of the group) rated SOE as 0, indicating no perception of increased effort associated with speech production. The remaining 19 subjects had a mean rating of 2.2, standard deviation of 0.98 and a range from 1-4 on the 5-point scale. The mean rating suggests that, as a group, these 19 individuals perceived themselves as having slightly more than a mild increase in effort associated with speech production. Overall SOE was rated as minimal by 5 subjects, mild by 9, moderate by 4, and severe by 1.

3.3.3. Locus

Fifteen of 34 (44%) did not mark any location as having increased SOE on the lateral schematic. These were the same 15 subjects who rated overall effort as 0. The remaining 19 subjects marked at least one location on the schematic. The number of subjects identifying increased SOE at each site, and the distribution of rankings is reported in Table 2. From the tabled data it can be inferred that 4 subjects identified only one location of effort, 9 identified two sites, 3 identified three sites, and 3 identified 4 sites. The lips and the tongue were the most commonly identified sites of increased SOE followed by the brain, voice and lungs. The brain ratings were interpreted as indication of “mental effort.” The pharynx was never marked as having increased SOE.

<table>
<thead>
<tr>
<th>Ranking Position</th>
<th>Lips</th>
<th>Tongue</th>
<th>Pharynx</th>
<th>Voice</th>
<th>Lungs</th>
<th>Brain</th>
<th>Number of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>7</td>
<td>--</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>4</td>
<td>--</td>
<td>2</td>
<td>--</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>--</td>
<td>1</td>
<td>--</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

3.3.4. Differences in Pressure Measures as a Function of Effort Rating

Three t-tests for independent measures were calculated to evaluate differences in Po, ACP, and Psub for those subjects without elevated SOE (N=15) compared to those with some degree of elevated effort (N=19). Table 3 reports the mean pressure values for the subjects divided into two groups considered “no extra effort” (rating of 0 on the 5-point scale) and elevated speech effort (ratings >1). Oral pressure and tongue-palate contact pressure were significantly higher for the subjects who reported increased speech effort compared to those who
reported no excess speech effort (Po: \( t=2.95, \ p=0.007 \); ACP: \( t=2.41, \ p=0.015 \)). Psub did not differ significantly between the two groups (\( t=0.76, \ p=0.691 \)).

Table 3. Means and standard deviations for the pressure measurements (kPa) for subjects who rated themselves as having no exceptional effort vs. those with some degree of elevated speech effort (>1 on the 5-point scale).

<table>
<thead>
<tr>
<th>Measure</th>
<th>No Speech Effort (N=15)</th>
<th>Some Speech Effort (N=19)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Oral Pressure (Po)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.85</td>
<td>1.58</td>
</tr>
<tr>
<td>sd</td>
<td>(0.36)</td>
<td>(0.56)</td>
</tr>
<tr>
<td><strong>Articulatory Contact Pressure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean (ACP)</td>
<td>6.5</td>
<td>7.7</td>
</tr>
<tr>
<td>sd</td>
<td>(2.7)</td>
<td>(3.0)</td>
</tr>
<tr>
<td><strong>Subneoglottic Pressure (Psub)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.4</td>
<td>3.7</td>
</tr>
<tr>
<td>sd</td>
<td>(1.1)</td>
<td>(2.6)</td>
</tr>
</tbody>
</table>

3.4. Correlational Analysis

Multiple Pearson product moment correlations among SOE, ACP, Po and Psub were calculated. In order to limit the number of statistics computed, consideration was given to how data could be collapsed across consonants. Correlations with data collapsed across all five consonants are presented in Table 4 to look for more general trends in the TE speech data. Table 3 also presents correlations for subgroupings of the consonants depending on Po status. As noted above, there were significant differences between consonants in terms of Po; it was possible that these differences might obscure or weaken correlations of Po to the other variables in the total data collapse statistics.

Overall, SOE, Po and ACP appeared to be moderately to strongly correlated with one another. SOE was strongly correlated to ACP regardless of the consonant subgrouping (r values ranged from .76 to .86) and was moderately to strongly correlated with Po (r values ranged from .42 to .72). The correlation between ACP and Po was strong for the individual consonant subgroupings (.60 to .79). The subneoglottal pressure measure had notably weaker relationships to the other three variables even though some significant correlations were still present. There was not a significant correlation between Psub and Po for any of the subgroupings or for the combined data set. Excluding the correlations when using combined /t/ and /d/ data, Psub was significantly correlated with SOE and ACP although the r values ranging from .21 to .39 suggested only mildly strong relationships.
Table 4. Correlation Matrices for the four variables of interest. Statistically significant correlations at the .05 level are marked with an asterisk. ACP=Articulatory contact pressure, Po=Oral Pressure, Psub=Subneoglottal pressure, SOE=Sense of effort.

<table>
<thead>
<tr>
<th></th>
<th>PO</th>
<th>ACP</th>
<th>Psub</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOE</td>
<td>.42*</td>
<td>.76*</td>
<td>.24*</td>
</tr>
<tr>
<td>Po</td>
<td>---</td>
<td>.44*</td>
<td>.18</td>
</tr>
<tr>
<td>ACP</td>
<td>---</td>
<td>---</td>
<td>.21*</td>
</tr>
<tr>
<td>/t/ + /d/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOE</td>
<td>.58*</td>
<td>.86*</td>
<td>.20</td>
</tr>
<tr>
<td>Po</td>
<td>---</td>
<td>.79*</td>
<td>.19</td>
</tr>
<tr>
<td>ACP</td>
<td>---</td>
<td>---</td>
<td>.15</td>
</tr>
<tr>
<td>/s/ + /z/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOE</td>
<td>.72*</td>
<td>.80*</td>
<td>.39*</td>
</tr>
<tr>
<td>Po</td>
<td>---</td>
<td>.68*</td>
<td>.13</td>
</tr>
<tr>
<td>ACP</td>
<td>---</td>
<td>---</td>
<td>.29*</td>
</tr>
<tr>
<td>/n/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOE</td>
<td>.57*</td>
<td>.78*</td>
<td>.22*</td>
</tr>
<tr>
<td>Po</td>
<td>---</td>
<td>.60*</td>
<td>.19</td>
</tr>
<tr>
<td>ACP</td>
<td>---</td>
<td>---</td>
<td>.26*</td>
</tr>
</tbody>
</table>

4. Discussion and Conclusions

Maximizing speech intelligibility has been a long-standing goal of both researchers and clinicians working with individuals who utilize TE speech. In addition to maximizing the new voice in terms of its audibility, consistency, quality, and so forth, most clinicians seem willing to recognize the need to address articulatory behaviors, at least for some TE speakers. Unfortunately, there is limited empirical evidence to guide decisions about what aspects of articulation might be the most important to evaluate and address therapeutically. The current study adds to small, but growing knowledge base regarding articulatory behaviors during TE speech with an ultimate goal of guiding clinical decisions about the need (or lack thereof) for articulatory intervention, and specific approaches to that intervention. Consonant production has been an ongoing focus in this line of research, in part because of the important role that consonants play in intelligibility, but also because the anatomical changes from laryngectomy create a substantial alteration in speech aerodynamics (i.e., diversion of the trachea to the base of the neck; re-coupling the lower and upper airways via a fistula; altered neoglottal aerodynamics that might impact air flow upstream of the PE segment). The only significant amount of work on consonant production in TE speech that has received any degree of repeated investigation has been the ability to produce the voiced-voiceless distinction (Jongmans, Hilgers, Pols, van As, 2006; Saito et al, 2000; Searl and Carpenter, 2002). The
current study attempts to broaden the understanding of physiologic characteristics of consonant production in TE speech and to relate these instrumental measures to speakers’ ratings of effort.

4.1. **Pressure Measurements**

4.1.1. Articulatory Contact Pressure

Tongue-to-palate contact pressures during alveolar consonants produced by this group of TE speakers were substantially higher than what has been reported for a similar set of consonants in non-laryngectomized adults (Searl, Evitts and Davis, 2007). Both studies used the same instrumentation, although the speech stimuli varied somewhat. The slight differences in stimuli, however, are not likely to explain the approximate doubling of contact pressures demonstrated by the TE speakers. Lip ACP data from TE speakers also found that TE speakers produced bilabial phonemes with ~2 times higher pressure (Searl 2007). It is interesting to note that in both the current study and Searl (2007) that the nasal phonemes also were produced with this elevated contact pressure. Given the low oral pressure requirement for nasal phoneme production, one might hypothesize that only limited articulatory contact pressures would be invoked to contain the limited air pressure. However, it seems the case that the speech system does not attempt to modulate contact pressures during speech depending on a particular consonant (no differences as a function of consonant have been found for TE or laryngeal speakers in the studies from our lab). Rather, it seems that the speech system merely targets a contact pressure level that will be sufficient to contain air pressures across the range of consonants to be produced. For TE speakers, it appears that the speech system simply re-adjusts the ACP to a notably higher level. From an economy-of-effort perspective, it seems reasonable for the speech system to not have to specifically adjust contact pressures dependent on the consonant. However, the TE data do suggest that the overall level of physical effort expended during consonant articulation is greater than in non-laryngectomized speakers.

4.1.2. Oral Air Pressure

The Po data for this set of alveolar consonants are consistent with what has been reported previously (Motta et al, 2001; Saito et al, 2002; Searl, 2002, 2004, 2007). The pattern of differences in Po across the stimulus set were predictable with lowest Po on the nasal /n/, highest pressure on the stop consonants /p/ and /b/, and intermediate values for the fricatives /s/ and /z/. Overall, the Po values are notably higher than what is reported for non-laryngectomized adults (again, 2-3 times higher). The lower pressure on the nasal phonemes in TE speech is presumably for the same reason that low pressures occur on nasal in non-laryngectomy speakers. That is, the velopharyngeal port is open, and all else being equal in terms of respiratory drive and oral articulatory behaviors, a Po drop will occur because air can escape through the nose.
4.1.3. Subneoglottal Pressure

Although pressure within the esophagus of TE and esophageal speakers has been reported previously, the data are sparse. Subglottic pressure in non-laryngectomized speakers is expected to be within a range of approximately 5-10cmH2O (~0.49-0.98kPA) (Wilson & Leeper, 1992). The overall mean $P_{sub}$ for the TE speakers in this study was 3.5 times greater than the range expected from normal speakers. This is not surprising given earlier studies that have described the elevated resistance of the neoglottis (e.g., Weinberg et al., 1982).

A few comments regarding the procedure for obtaining the $P_{sub}$ measure in this study are warranted given the novelty of the approach. The intent was to avoid placing a tube within the lumen of the prosthesis which would diminish the cross-sectional area through which pulmonary air would flow and which would have held open the one-way valve at the esophageal end of the prosthesis. The current approach was generally successful although the prosthesis insertion process was somewhat uncomfortable for a small number of subjects, presumably because of the increased overall diameter of the prosthesis passing through the fistula. Once in place, subjects did not complain of pain or other sensations of discomfort, although when directly asked they generally reported that they could feel something different. No leakage of saliva around the prosthesis was noted for any of the subjects and their TE voices were not perceptibly different when the modified prosthesis was in place (voice recordings with the usual and the modified prosthesis are available for judging, but that has not been done to date). The problem of saliva blocking the esophageal tip of the sensing tube was generally easy to manage using the suction arrangement. Since that time, we have found it simpler to pass a length of acrylic fishing line through the sensing tube all the way to the esophageal tip to disperse saliva bubbles. The fishing line is flexible, but has enough stiffness to not coil up in the tube. By placing marks along the length of the sensing tube all the way out to the distal end that is attached to the transducer, a comparable length of fishing line can be passed through the tube to reach the esophageal tip.

4.2. Sense of Effort and Relationship among Measurements

As far as the author is aware, this is the first attempt to systematically describe the SOE associated with speech in general from a group of individuals using TE speech. Increased effort has been discussed for TE speakers with a focus mainly on issues of voice production (e.g., excess respiratory effort, neck tension, strained voice or aphonia, etc.). A sizeable portion of the group (~45%) did not report any sense of increased effort. However, 55% of the group reported at least some amount of increased effort, most in the mild range of the rating scale. It may be tempting to diminish reports of effort in the minimal or mild range. However, speech should not be effortful. In fact, non-laryngectomized individuals are usually not aware of any sense of effort associated with speech production unless they have been fatigued or they are required to significantly increase
loudness (see Solomon, 2000). For that reason, speakers with any degree of elevated SOE were grouped together for portions of the analysis.

Inspection of the data regarding the locus of the SOE suggests that the lips and the tongue were the most commonly identified sites of effort with over 60% of subjects marking either or both articulators. ACP also differed significantly when comparing those who reported no increased effort to those who reported some elevated effort. It seems reasonable to conclude that subjects SOE ratings and identification of the lip and tongue as sites of effort are a reflection of the physical effort involved in articulation as documented in the ACP measure. The correlational analysis also can be interpreted as support for that position. That is, ACP may be a measurable, physiologic correlate of the subjective rating of SOE. If true, it may be possible to use the ACP signal as feedback to an individual in an effort to manipulate SOE. Care would have to be taken to track the aerodynamic (Po) and perceptual outcome (consonant intelligibility or precision) when such manipulations are attempted. The intent would be to limit physical effort but still produce a strong consonant with a burst or frication that is sufficient for the listener to accurately decode what is being said. The fact that Po and ACP were also strongly correlated means that caution must be taken to not have the unintended effect of weakening the stop consonant or fricative. Of course, such future work is contingent on first determining whether the findings of increased SOE by TE speakers is replicated in other studies and determining whether the increased effort rises to a level that warrants consideration for manipulation.

The number of subjects identifying the voice as a source of effort was somewhat lower than anticipated. Knowing that the rehabilitation process generally focuses on the new voice and that neoglottal resistance and Psub are expected to be elevated (which it was in this study), one might have expected subjects to focus more on the neoglottis when identifying locus of effort. However, this was a highly proficient group of speakers who were not having trouble with the voicing aspects of TE production. It seems reasonable to expect that in other TE speakers, the locus of effort very well could be on the voice, with or without concurrent sense of effort in the articulators. Also, it may be that subjects had simply accommodated to the elevated driving pressures below the neoglottis so that it was no longer a salient perception to the speaker.

The fact that almost half of the group identified the brain as a site of effort is also telling. This was interpreted to reflect a sense of elevated cognitive load. Further exploration of this concept of cognitive effort might be important, particularly if means of reducing this effort can be identified. It may be that factors such as time-post laryngectomy or TE puncture are critical, or it may be that manipulations of other behaviors that reduce physical effort (e.g., reducing ACP) might also translate into lower cognitive loads.

The relationship of Psub to the other variables is less easy to interpret. Psub was not significantly correlated to Po. This suggests that the elevated Po that has
been consistently documented in TE speakers is not likely a direct outcome of elevated pressures below the neoglottis. It may be that the oral articulatory behaviors that result in increased ACP on these consonants is primarily responsible for increasing Po. For example, when a TE speaker is producing an alveolar stop, they may have been trained (or intuited the need) to generate a strong burst. To do so, they may have the tongue contact the palate more forcefully, and perhaps with greater surface contact (though this has not been documented to date). The net result might be further reduction of the oropharyngeal cavity volume, and subsequently greater compression of the air within the vocal tract (i.e., increased Po), relative to a production that is done less forcefully and with less complete contact between tongue and palate. Future work that incorporates methods of surface area contact between articulators, or even better, volume of the vocal tract, could help elucidate this issue.

Psub was correlated with SOE, but the strength of the correlations was weaker than that between ACP and SOE. The correlation between Psub and ACP also was fairly weak (although statistically significant for some of the consonant set). The ACP-Psub correlation suggests that individuals who demonstrate increased ACP also have higher Psub. TE speakers might be considered to vary along a continuum of physical effort (reflected in contact and air pressure measures) and that the level of effort may not be specific to one aspect of the process, but rather pervades most or all of the production parameters. Future studies that allow for simultaneous recordings of measures that might reflect effort across the speech system (respiratory, neoglottal, oral articulatory, etc.) are needed to explore this situation further. Results of this study suggest that contact pressures as well as some air pressure measures could serve as a physiologic correlate to perceptions of speech effort.

4.3. Future Directions

Studies are currently underway in which TE speakers are provided with ACP and/or Po biofeedback as they attempt to decrease ACP. The goal is to see if an individual can reduce the physical work of articulation (ACP), but maintain Po. Acoustic recordings of such efforts will be used to help determine whether the perceptual product is degraded as these manipulations are made. Studies of effort (both subjective and physiologic) in electrolaryngeal and esophageal speakers are of interest. It will be particularly interesting to compare measures of effort within speakers who are able to utilize more than one method of alaryngeal communication to see if one is more likely than another to induce greater articulatory demands or perceptions of greater articulatory effort.

5. Acknowledgements

Thanks are given to all of the individuals with a laryngectomy who willingly gave of their time. A portion of this work was supported by a Technology Innovation Enhancement Grant, Bowling Green State University.
6. References


Abstract

Objectives: to develop an evidence-based therapy program aimed at improving tracheoesophageal speech intelligibility. The therapy program is based on particular problems found for TE speakers in a previous study as performed by the authors.

Patients/Materials and Methods: 9 male laryngectomized individuals were included in the study. Seven had had a total laryngectomy, one had had a PM flap and one a gastric pull-up. Mean age was 64.9 years and mean time after TLE was 5.6 years. A therapy program was developed based on the most common difficulties of TE speakers found in literature and on a previous perception experiment of the authors. Where possible, therapy techniques were based on evidence-based research. TE speakers participated in a nine-session therapy program spread over five weeks. Several pre and post tests were used to measure the effect of the therapy program: the speakers were asked to read out speech material, both before and after the therapy program, including almost all levels of speech: phoneme level, word level, sentence level and spontaneous speech. Different types of raters were used for each of the different tests. Naïve, inexperienced listeners rated spontaneous speech in a semantic scaling experiment; naïve, but experienced raters (phoneticians) rated phoneme and sentence level; experienced listeners, well known with TE speech, rated phoneme level. Questionnaires were filled out by the speakers themselves and a close relative. A Voice Handicap Index was also used, but will not be discussed here.

Results: Scores improved for fricatives in initial position, and for the voiced-voiceless distinction. For medial position, all manners of articulation, except the approximant /l/ improved, as well as the voiced-voiceless distinction. Sentence intelligibility also improved. The questionnaires show that speakers are more positive about their own speech post therapy. The scores for the semantic
scaling experiment did not improve. However, the semantic scale results on articulation and intelligibility did improve.

Conclusions: Although already effective in its present set-up, some extension of this 5-week rehabilitation program is necessary to ensure an optimal progression from phoneme level to spontaneous speech level.

Keywords: TE speech production, evidence-based therapy

1. Introduction

A total laryngectomy changes the anatomy and physiology considerably, which has far-reaching consequences for the voice quality and speech intelligibility of the laryngectomees. In this paper, only tracheoesophageal (TE) speakers will be discussed. Especially the voice quality of TE speakers has been studied quite extensively, also in Dutch [Van As, 2001]. The choice for that topic was motivated by the fact that especially the voice source has altered dramatically. The question is if a loss in voice quality also entails a loss in intelligibility. However, TE speech intelligibility has received relatively little attention abroad and in the Netherlands. There are studies that have looked at TE speech intelligibility but unfortunately, most of these studies were limited in scope and most of them only focused on phoneme intelligibility [Doyle, Danhauer & Reed, 1988; Hammarberg, Lundström & Nord, 1990; Miralles & Cervera, 1995; Roeleven & Polak, 1999; Oubrie, 1999; Boon-Kamma, 2001; Searl, Carpenter & Banta, 2001; Lundström & Hammarberg, 2004]. Nevertheless, these studies rendered interesting and valuable information on TE speech intelligibility and all showed that TE speech is less intelligible than normal laryngeal (NL) speech. The results of these studies were later confirmed by our own study in which almost all levels of TE speech were studied: from phoneme level to spontaneous speech [Jongmans et al., 2006]. It appeared that the TE speakers in our perception experiment showed an overall consonant intelligibility rate of 72% and a vowel intelligibility rate of 74%. This rate is much lower than the 93% consonant correct score for normal laryngeal speakers as found by Pols (1983) for CVCVC nonsense words and the 84% vowel correct score as found by Koopmans-van Beinum (1980) for CVC words. Other problems found for TE speakers are conveying sentence accent accurately [Van Rossum et al., 2002] and problems with intonation and timing [Roozen, 2005].

A diminished intelligibility should be taken seriously as intelligible speech is a prerequisite for communication. Optimizing intelligibility will improve communicative ability (e.g., conversing over the telephone), independence, and ultimately quality of life [Ackerstaff et al., 1994]. However, despite the evidence that TE speakers have problems with a number of phonemes and features, as well as with intonation and timing, speech language pathologists (SLPs) do not routinely provide the patient with a rehabilitation program that covers both voice and speech quality. At present, the main goal is usually to establish an audible voice of acceptable quality. Also, prosthesis care (deservedly) receives much
attention during therapy. This contrasts with the training of electrolarynx (EL) and esophageal (E) speech, during which articulation does receive more attention. Studies have compared the different speech modes and found TE intelligibility more favorable than EL or E speech [e.g. Williams & Watson, 1987; Pindzola & Cain, 1988; Debruyne et al., 1994 & Max, Steurs & De Bruijn, 1996]. It might be that TE speech is perceived to be so much more intelligible than EL and E speech, that it causes a shift in the SLP’s focus of training from articulation training to prosthesis use/care and stoma occlusion. Additionally, busy hospital schedules and limited resources might prevent SLP’s from investing extra time in intelligibility training. Many researchers and SLP’s throughout the world, however, recognize that there is more to gain for the patient through specific intelligibility training. To quote Searl (2004): “I simply want to remind myself and others not to settle for merely usable speech, but rather to work towards the best speech possible for each laryngectomee.” In the same article he expresses surprise that about 10% loss of intelligibility in stroke patients is considered serious enough to warrant training whereas professionals are “content to let an intelligibility deficit of this magnitude (or greater!) go unaddressed in the laryngectomee patient.”

Apart from the abovementioned reasons there is another explanation why intelligibility in TE speech might be neglected. For Dutch, no large scale studies analyzing TE speech intelligibility had been performed so far. Thus, there was no experimental evidence that intelligibility might be compromised. Without this evidence, there would not be much incentive to develop an (evidence-based) therapy program. However, results from our earlier studies have shown that the rationale for developing an evidence-based therapy program is obvious.

The goal of our study was to develop a program that has the potential to improve intelligibility in TE speech. The typical TE speech problems that were found in our perception experiments and elsewhere in the literature form the basis of this program. A systematic literature search was performed to see if any evidence-based therapies or training techniques could be found for laryngectomees in general and TE speakers in particular. This search rendered thirteen results, seven studies of which were evidence-based [Fitzpatrick, Gould & Nichols, 1980; Mase-Goldman, Allen & Nichols, 1988; Zeine & Brandt; 1988, Christensen & Dwyer; 1990; De Maddalena & Pfrang, 1993; Bellocc, Vaissière & Brasnu, 1997; Wong Chung et al., 1998]. However, only two of those seven studies were based on TE speakers [Bellocc Vaissière & Brasnu, 1997; De Maddalena & Pfrang, 1993]. Whenever no adequate technique could be found, the expertise and experience of the two SLPs involved in the development of the program was used.

The training consisted of nine one-hour lessons, divided over five weeks. Offering the same program to all speakers makes statistical analysis easier and sessions are easier to plan. Also, standardly, Dutch health insurance pays for nine sessions maximally. The program was divided into two main parts: lessons 1-6 dealt with problematic phonemes and lessons 7-9 dealt with sentences and
spontaneous speech. This structure was to ensure a progression from phoneme level to spontaneous speech.
In order to meet the demands of evidence-based rehabilitation, an effect study has been performed to measure if the training program has been successful. In effect studies like the present one, five levels of evidence can be distinguished (Moore, McQuay & Gray, 1995). This study has level III evidence: Evidence from well designed trials without randomization, with a single group performing pre and post tests. In order to assess the effect of therapy, we performed several pre and post tests. By comparing pre and post test scores, it can be established whether therapy has been successful. With this study we hoped to answer two main questions: can TE speech intelligibility be improved by specific speech therapy and is the set-up of the developed program suitable for its purpose. In this paper, the effect study and its outcomes will be discussed. For more details see Jongmans (2008).

2. Patients/Materials and Methods

2.1. Subject selection

In order to obtain a subject group as homogeneous as possible, selection criteria were formulated. The most important criteria were that speakers had undergone a standard total laryngectomy, that time after the operation was longer than six months and that speakers did not have excessive problems producing voice as this usually indicates other physiological or anatomical problems. Ten subjects could be maximally accommodated and eventually, nine TE speakers participated in the study. Speakers had a mean age of 64;9 (range 54;11-82;9) and mean time after TLE was 5;6 (range 0;7-13;7). All speakers but one had received radiation. Due to the limited pool of potential participants, two of the speakers included had additional resections: one with a Gastric pull-up and one with a PM flap.

2.2. Recordings

Recordings for the pre and post tests were made in a sound-treated room with a Sennheiser MD421 Dynamic Microphone and an Edirol (Roland) R-1 portable 24 bit digital wave audio recorder. The maximum recording level was chosen for all speakers and then fixed.

2.3. Pre and post tests

The pre and post tests consisted of three different listening experiments performed by different groups of listeners as well as study-specific questionnaires filled out by the subjects and their relatives. The idea behind this variation in tests and listeners is that TE speech intelligibility can be tested at different levels (from phoneme to discourse level) and that each listener group will evaluate speech intelligibility in a different way. Speech pathologists, for example, having their
professional frame of reference, will use different strategies when judging TE speech than naïve listeners will. By using several tasks it is hoped to create as complete a picture of intelligibility as possible implying different levels: from overall subjective impressions to transcription at phoneme level.

The pre and post tests can roughly be divided into two groups: on the one hand phoneme and sentence intelligibility tests (sect. 2.3.1) and on the other hand qualitative judgments of intelligibility and the influence of the speaker's intelligibility on daily life (semantic scales and questionnaires), see sect. 2.3.2.

2.3.1. Phoneme and sentence intelligibility tests

The participants were asked to read out loud syllables and Semantically Unpredictable Sentences (SUS), e.g. “Het scherpe boek vormt de tuin” (the sharp book shapes the garden) both before and after the therapy program [Benoit, Grice & Hazan, 1996]. The syllables consisted of the Dutch consonants /p b t d k h x f v s z l n m/ in combination with the vowels /a i u/ in CV and CVC (nonsense) words. These consonants were trained specifically in the therapy program. All Dutch vowels, including diphthongs were tested in CVC words. Where possible, minimal pairs were used differing only in the length of the vowel (for example bom versus boom). A set of five SUS sentences were also recorded for each speaker.

Ten phonetically trained, naïve listeners were asked to identify phonemes and semantically unpredictable sentences. Listeners had a mean age of 45.4 (range 27-65). The listening experiment was performed online, so that people could perform the experiment in the comfort of their own home. After listening to a stimulus, they typed in what they perceived in normal spelling, which is unambiguous in Dutch. The program automatically rendered response files that were used for further analysis.

The participants also read out loud 25 existing words belonging to the so called ‘Dyva’ test, a standardized test normally used for dysarthric speakers [Dharmaperwira-Prins, 1998]. The advantage of this wordlist is that it includes all Dutch vowels and consonants plus a range of consonant clusters. The fact that the test is standardized makes it easier to compare the outcomes of the subjects. Another reason why this test was used is that this is the only known standardized intelligibility test in the clinical setting. This lack of tests may be filled in the future by the development of a new instrument for testing intelligibility, developed in Belgium by De Bodt, Guns & Van Nuffelen (2006). Unfortunately this test was not available yet at the time of testing.

Six Speech Language Pathologists (all female) participated because they are considered analytical listeners, who are not distracted by the voice quality of the speaker and hence may give a different judgment as compared to naïve experienced listeners (phoneticians). SLP’s had to have at least 6 months
experience with TE speakers. Mean age of the SLP’s was 40 (range 26-59). They also performed the experiment online. During the test, the listeners saw on the screen which word was spoken at that moment. First, the quality of the ‘target’ consonants and vowels were judged, which were marked on the screen: if they were produced correctly, the listener could go to the next stimulus; if they were not produced correctly, listeners deleted that particular target sound in the text box. Then, the listener typed in exactly what she perceived in normal spelling (the whole word). There was also room for additional comments. The original test only counts the number of errors (how many target sounds were erased) and not the type of errors. However, the results of the phoneticians in the first experiment will provide information on the type of errors made, just as the whole response given by the SLP’s.

2.3.2. Semantic scaling and study-specific, structured questionnaires

Speakers were asked to read out loud an adapted text from the newspaper about maltreated bears and were then told to retell the story. Ten naïve listeners (mean age 52 years; range 40-68 years) were asked to evaluate speech intelligibility and voice-quality-related aspects of (semi) spontaneous and read-aloud speech. These types of speech come closest to normal connected speech. In this experiment, the naïve listeners rated how well TE speakers can be understood and how their voice sounds. For this purpose, semantic bipolar 7-points scales were used. Furthermore, there was one scale which concerned an overall judgment, this scale had only three options: good, moderate and poor. It was included here as it had been proven to be a useful scale in the voice quality related perception experiment performed by Van As et al. (1999).

A judgment was wanted on specific aspects of TE speech, especially on those aspects that were trained during the therapy program. Voice quality is known to be affected as well and as this may influence the overall intelligibility, voice quality related scales were also included. The scales represent the following aspects: intelligibility, articulation, intonation, rhythm, loudness and voice quality. There is overlap in meaning for some of the scales. The difficulty with semantic scales is that the terms used may not mean the same to all listeners [Fagel, Van Herpt & Boves, 1983]. Results were expected to show that some terms give better results than others and that based on these results a subset of scales can be distinguished that represents the underlying perceptual dimension.

Each speaker and a close relative also filled out study-specific, structured questionnaires. The questions used were in part taken from a laryngectomy quality-of-life questionnaire devised and in long-term use in the Netherlands Cancer Institute and then complemented by specific questions formulated for the current study [Ackerstaff et al., 1994]. The questions were all phrased as multiple-choice and required the expression of the opinion on how speech intelligibility of the subject affected daily life. The questionnaires were used to get an idea of how satisfied the speakers and their relatives are with the speech of
the TE speaker and also if there is a difference in satisfaction rate between the
TE speakers and their relatives.

2.4. Statistics

For the phoneme and sentence intelligibility tasks, first Cronbach's alpha was
used to calculate inter-rater reliability. The McNemar test for two related samples
was used to compare the results from the pre and post test to see if any
significant differences can be established. For the questionnaires, a Wilcoxon
signed ranks test for related samples was used to investigate the difference
between the pre and post test scores. A Mann-Whitney test for independent
samples was used to investigate the difference between the speakers and their
relatives. For the semantic scales, first the reliability of each scale was calculated
using Cronbach's alpha. Only scales with a coefficient higher than .70 were
included for further analysis. A Principal Component (PC) analysis was
performed to investigate the relations between the semantic scales. After the PC
analysis a VARIMAX rotation was performed, by which the variance of the
component loadings is maximized per factor. Only components with
eigenvalues>1 were included.

2.5 Praat

The program Praat was used to prepare stimuli for the listening experiment
[www.praat.org].

3. Results

In this section, first the results of the phoneme and sentence identification
experiment as performed by the phonetically trained listeners and the results of
the Dyva experiment as performed by the speech language pathologists (SLPs)
will be discussed. After that, the results of the questionnaires and the semantic
scaling experiment will be presented.

3.1. Phoneme and sentence intelligibility tests

3.1.1. Phoneme and sentence intelligibility test by phonetically trained listeners:
consonants

For the consonants the inter-rater reliability was .747 and .796 for initial and
medial position in the pre test and .747 and .767 for initial and medial position in
the post test, respectively.

The following table shows the overall percentage correct scores for all nine
speakers individually, as well as the mean scores, both for initial and medial
consonants in the pre and post test.
Table 3.1. Percentage correct scores for the individual speakers and the mean scores per test and per position. The bold numbers indicate the speaker with the greatest deterioration in score; the bold and italicized numbers indicate the highest improvement.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre initial</td>
<td>62</td>
<td>75</td>
<td>71</td>
<td>61</td>
<td>67</td>
<td>56</td>
<td>71</td>
<td>58</td>
<td>83</td>
<td>67</td>
</tr>
<tr>
<td>Post initial</td>
<td>72</td>
<td>76</td>
<td>79</td>
<td>68</td>
<td>77</td>
<td>66</td>
<td>69</td>
<td>72</td>
<td>86</td>
<td>74</td>
</tr>
<tr>
<td>Pre medial</td>
<td>68</td>
<td>82</td>
<td>78</td>
<td>69</td>
<td>75</td>
<td>63</td>
<td>75</td>
<td>66</td>
<td>79</td>
<td>73</td>
</tr>
<tr>
<td>Post medial</td>
<td>80</td>
<td>78</td>
<td>70</td>
<td>86</td>
<td>84</td>
<td>67</td>
<td>83</td>
<td>79</td>
<td>82</td>
<td>79</td>
</tr>
</tbody>
</table>

The table above shows a range of 56-83 percent correct for initial position in the pre test. The range is smaller (66-86) in the post test. There are large differences between the percentages of improvement of individual speakers. Speaker 8 shows the highest improvement for initial position (a difference of 14%), whereas speaker 7 shows a small deterioration. For medial position, the pre test shows a range of 63-82 percent correct, against a range of 67-83 percent correct in the post test. Again considerable differences between individual improvements are found: speaker 4 shows the highest improvement of 17%, whereas speaker 3 shows a decrease of 8%. This table also shows that speakers do not always score consistently: a high improvement in initial position does not necessarily mean a high improvement in medial position, just as a decrease in score in initial position does not mean the speaker will show a decrease in medial position.

The following table lists the mean percentage correct score for the pre and post test per manner of articulation for initial position. The phonemes included are /p b t d k h x f v s z l n m/.

Table 3.2. Mean percentage correct score per manner of articulation and the overall score for the pre and post tests for initial position, with the ranges between brackets and the significance in the last column.

<table>
<thead>
<tr>
<th>Initial position</th>
<th>Pre test (range)</th>
<th>Post test (range)</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosives</td>
<td>77 (66-89)</td>
<td>79 (61-92)</td>
<td>NS</td>
</tr>
<tr>
<td>Fricatives</td>
<td>50 (35-83)</td>
<td>62 (31-87)</td>
<td>p&lt;.01</td>
</tr>
<tr>
<td>Nasals</td>
<td>88 (72-97)</td>
<td>87 (75-93)</td>
<td>NS</td>
</tr>
<tr>
<td>Approximant /l/</td>
<td>77 (50-95)</td>
<td>83 (60-95)</td>
<td>NS</td>
</tr>
<tr>
<td>Mean</td>
<td>67 (56-83)</td>
<td>74 (66-86)</td>
<td>p&lt;.01</td>
</tr>
</tbody>
</table>

It can be seen in Table 3.2 that there is an overall significant improvement between the pre and post test. Closer inspection of Table 3.2 shows that this improvement is entirely caused by the fricatives that score much higher in the post test. The other categories only show marginal improvements and the nasals show no improvement at all.
In the next table, the mean results for medial consonant position are given.

Table 3.3. Mean percentage correct score per manner of articulation and the overall score for the pre and post tests for medial position, with the ranges between brackets and the significance in the last column.

<table>
<thead>
<tr>
<th></th>
<th>Medial position</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre test (range)</td>
<td>Post test (range)</td>
<td>Sign.</td>
<td></td>
</tr>
<tr>
<td>Plosives</td>
<td>81 (64-98)</td>
<td>86 (66-97)</td>
<td>p&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Fricatives</td>
<td>62 (49-79)</td>
<td>69 (59-83)</td>
<td>p&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Nasals</td>
<td>73 (40-92)</td>
<td>86 (72-97)</td>
<td>p&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Approximants</td>
<td>90 (57-100)</td>
<td>92 (70-100)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>73 (63-82)</td>
<td>79 (70-86)</td>
<td>p&lt;.01</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3 for medial position shows even more promising results than Table 3.2. Again, there is a significant overall improvement. However, in medial position, three of the four categories show significantly higher scores. Only the approximant /l/ does not improve significantly. The nasals show the greatest improvement in percentages.

When initial position and medial position are compared with each other, it can be seen that more categories improve significantly in medial position. The mean scores, both for the pre test and the post test are also higher in medial position. It seems to be easier for TE speakers to produce consonants between vowels (and apparently it is also easier to apply the techniques learned for consonants in medial position). However, it is not only interesting to look at the overall scores, but also to look at the individual scores. Table 3.4 summarizes the significant increases and decreases in score for the individual speakers per manner of articulation and per position.

Table 3.4. Statistically significant increases (+) and decreases (-) in phoneme intelligibility per patient and per category.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial plosives</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial fricatives</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial nasals</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Initial approximants</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Medial plosives</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial fricatives</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial nasals</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial approximants</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3.4 shows that even when an overall significant improvement is found, it does not necessarily mean that all speakers improved their score.
3.1.2. Phoneme and sentence intelligibility test by phonetically trained listeners: vowels

For the vowels, the inter-rater reliability was .78 for the pre test and .76 for the post test, respectively. One of the ten listeners was left out for this particular part of the experiment as this person failed to complete this vowel test. Table 3.5 shows the individual overall scores for the pre and post test.

Table 3.5. Individual and mean scores for vowels in the pre and post test. The numbers in bold indicate the speaker with the greatest deterioration; bold italicized numbers indicate the highest improvement.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>98</td>
<td>89</td>
<td>92</td>
<td>97</td>
<td>76</td>
<td>90</td>
<td>93</td>
</tr>
<tr>
<td>Post</td>
<td>98</td>
<td>99</td>
<td>93</td>
<td>91</td>
<td>91</td>
<td>92</td>
<td>86</td>
<td>95</td>
<td>94</td>
<td></td>
</tr>
</tbody>
</table>

The first thing to notice from Table 3.5 is that the overall scores are all very high, except the pre test score of speaker 8. This is the person with most room for improvement, which he exploits rather well with an absolute improvement of 10%. Speakers 3 and 4 show a decrease in score, but their post test scores are still close to normal laryngeal scores (Koopmans-van Beinum, 1980). As with the consonants, it is also interesting to look at the scores for the types of vowels: long and short vowels and diphthongs. Table 3.6 shows these results.

Table 3.6. Mean scores per vowel type and overall with the ranges between brackets and the significance in the last column.

<table>
<thead>
<tr>
<th></th>
<th>Pre test (range)</th>
<th>Post test (range)</th>
<th>Sign. (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short vowels</td>
<td>87 (68-100)</td>
<td>89 (72-98)</td>
<td>NS</td>
</tr>
<tr>
<td>Long vowels</td>
<td>95 (75-100)</td>
<td>96 (87-100)</td>
<td>NS</td>
</tr>
<tr>
<td>Diphthongs</td>
<td>96 (74-100)</td>
<td>96 (80-100)</td>
<td>NS</td>
</tr>
<tr>
<td>Mean</td>
<td>93 (76-98)</td>
<td>94 (86-99)</td>
<td>NS</td>
</tr>
</tbody>
</table>

Table 3.6 shows that none of the categories improve significantly, which was to be expected considering the high overall pre test scores. When the individual scores are considered, only two speakers show a significant difference between their pre and post test: speaker 9 shows a significant improvement for the short vowels from 76% to 96% correct. For the long vowels speaker 4 shows a significantly lower score for the post test (from 100% to 91%). However, even the lower post test score for speaker 4 is still very high.

3.1.3. Phoneme and sentence intelligibility test by phonetically trained listeners: SUS sentences

The inter-rater reliability for the SUS intelligibility pre test is .88 and for the post test .85, respectively. Table 3.7 shows the individual percentage correct scores and the mean scores for the SUS sentences in the pre and post test.
Table 3.7. Individual scores for the SUS sentences and the mean score. The bold and italicized percentages indicate the greatest improvement. * indicates a significant difference.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>46</td>
<td>80</td>
<td>70</td>
<td>26</td>
<td>32</td>
<td>20</td>
<td>70</td>
<td>30</td>
<td>40</td>
<td>46</td>
</tr>
<tr>
<td>Post</td>
<td>64</td>
<td>86</td>
<td>78</td>
<td>44*</td>
<td>54*</td>
<td>46*</td>
<td>86*</td>
<td>34</td>
<td>44</td>
<td>60*</td>
</tr>
</tbody>
</table>

It is apparent from Table 3.7 that there are large individual differences with pretest scores ranging from 20% to 80% and post test scores from 34% to 86%. It can also be seen that all speakers improve their scores, most of them quite considerably, with the greatest improvement for speaker 6 who also had most to gain. However, significance could only be established for speakers 4 and 7 (p<.05) and speakers 5 and 6 (p<.01). It should be emphasized that also the overall improvement is significant (p <.01).

3.1.4. Dyva phoneme intelligibility test with Speech Language Pathologists

Inter-rater reliability for the Dyva test results is .72 for the pre test and .71 for the post test based on six speech language pathologists. Even though in this test the target phonemes consist of consonants, consonant clusters and vowels, in the final scoring no distinction is made between these different categories. The test in its original set-up is not concerned with the kind of mistakes made, but only with the number of mistakes made. In the original test, speakers belong to one of five categories: normal, lightly impaired, moderately impaired, severely impaired, and maximally impaired. However, as the Dyva test in the present study is used for a different category of ‘pathological’ speakers, and dysarthric patients are obviously not comparable to laryngectomized individuals, this study chose to look at the percentage correct score of the individual speakers for the pre and post test, just as in the other phoneme identification tests. Table 3.9 shows those results.

Table 3.8. Individual score for the ‘dyva’ test and the overall score in percentages. The bold and italicized percentages indicate the greatest improvement. * indicates a significant difference.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>89</td>
<td>95</td>
<td>96</td>
<td>91</td>
<td>94</td>
<td>89</td>
<td>92</td>
<td>87</td>
<td>84</td>
<td>91</td>
</tr>
<tr>
<td>Post</td>
<td>93*</td>
<td>97</td>
<td>96</td>
<td>90</td>
<td>94</td>
<td>91</td>
<td>94</td>
<td>90</td>
<td>95*</td>
<td>93*</td>
</tr>
</tbody>
</table>

Table 3.8 shows that all speakers scored high in the pre test. This means that there is not much room for improvement. However, six out of nine speakers improve their score in the post test, with speaker 9 showing the greatest improvement. Even though the improvements are only small and only significant for speaker 1 (p<.05) and 9 (p <.01), the overall improvement is significant.

Even though this test was validated for a different type of speaker, the outcome of this validated test shows that overall the impairment in intelligibility at the
phoneme level only seems to be moderate and that there is a trend towards a positive effect of the therapy program on TE speech intelligibility.

3.2. Questionnaires and semantic scaling experiment

In this section, first the results of the different questionnaires will be discussed, followed by a discussion of the semantic scale judgments.

3.2.1. Study-specific structured questionnaires for speakers and their relatives

The study-specific structured questionnaires consisted of 11 questions for the speakers and 10 questions for the relatives with slight adaptations in the formulation. The questions were the same for both types of raters, except that the 11th question applied only to TE speakers themselves. Each question had four reply options, with the first option being the most negative one and the last option the most positive. The numbers one to four were allocated as scores to these options.

In the following table the overall average scores are shown for the speakers and their relatives for the pre and post test scores. These scores are based on eight speakers and eight relatives, as one speaker failed to hand in his questionnaire (speaker 9). A word of caution is necessary: the questionnaires are limited in length and there are only two raters per speaker: the speaker himself and one relative. The limited amount of data makes it difficult to perform statistics.

Table 3.9. Average scores for the speakers and their relatives on the questionnaire. The significance level indicates whether a difference exists between pre and post test (rightmost column) or between speakers and relatives (last row).

<table>
<thead>
<tr>
<th></th>
<th>Pre</th>
<th>Post</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speakers</td>
<td>2.88</td>
<td>3.18</td>
<td>P&lt;.01</td>
</tr>
<tr>
<td>Relatives</td>
<td>3.21</td>
<td>3.35</td>
<td>NS</td>
</tr>
<tr>
<td>Significance</td>
<td>P&lt;.01</td>
<td>NS</td>
<td></td>
</tr>
</tbody>
</table>

From Table 3.9 it can be seen that, on average, speakers themselves scored significantly higher in the post test, meaning they were more positive about their intelligibility than at the start of the therapy. It should be noted, though, that only two speakers improve significantly. The judgment of the relatives did not improve significantly. Relatives generally appear to have a more favorable opinion of the speaker’s intelligibility than the speakers themselves. This difference is significant in the pre test, but disappears in the post test.

It is also interesting to look at the scores for each of the eleven questions that are shown in Table 3.10. Some of the questions are discussed below. It is however difficult to perform statistics, as there are only 8 speakers or relatives per question.
Table 3.10. Scores per question for the speakers’ and their relatives’ pre and post test. NA means this question was not scored in the questionnaire of that relative. Scores are based on eight subjects.

<table>
<thead>
<tr>
<th>Questions</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>speakers</td>
<td>Pre</td>
<td>3.4</td>
<td>3.3</td>
<td>3.1</td>
<td>1.7</td>
<td>3.1</td>
<td>2.9</td>
<td>2.2</td>
<td>2.9</td>
<td>2.0</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>3.6</td>
<td>3.4</td>
<td>3.1</td>
<td>2.2</td>
<td>3.3</td>
<td>2.9</td>
<td>3.2</td>
<td>3.2</td>
<td>2.6</td>
<td>3.3</td>
</tr>
<tr>
<td>relatives</td>
<td>Pre</td>
<td>3.6</td>
<td>3.4</td>
<td>3.2</td>
<td>2.1</td>
<td>3.6</td>
<td>3.6</td>
<td>2.9</td>
<td>3.8</td>
<td>2.9</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>3.7</td>
<td>3.4</td>
<td>3.3</td>
<td>2.8</td>
<td>3.7</td>
<td>3.4</td>
<td>3.2</td>
<td>3.4</td>
<td>2.9</td>
<td>3.1</td>
</tr>
</tbody>
</table>

It can be seen from Table 3.10 that the speakers show higher scores on all questions in the post test, but these improvements are mostly small. Three questions show a higher improvement: questions 4, 7 and 9. These were also the questions that initially scored lowest and hence had most room for improvement. Question 4 concerns the intelligibility at parties. It can be assumed that by training clear speech, intelligibility in especially these circumstances will improve. Question 7 concerns the speech rate. Again, this is an aspect that has been trained specifically during therapy. Question 9 concerns the pleasantness of the speech. This aspect has not been trained as such, but it is likely that when other aspects of intelligibility improve, the overall pleasantness will improve as well.

The relatives show a slightly different pattern: not all questions show improvement, but both increases and decreases in scores are small. Just as with the speakers, the largest improvements can be found for question 4 and 9, followed by question 7. The speakers and relatives seem to agree on these aspects, although it should be mentioned that no significant correlation could be established between the speakers’ own judgments and their relatives’ judgments.

3.2.2. Semantic scaling experiment for naïve listeners

Ten naïve and inexperienced listeners judged the subjects’ speech by using semantic bipolar 7-points scales. Two types of speech material were used for this experiment: read-aloud text and a recount of the same story. Inter-rater reliability was .887 both for the retelling and reading aloud task in the pre test, and .872 and .884 for the retelling and reading aloud task in the post test, respectively. Table 3.11 shows the overall results for both types of speech material for all 18 scales together. The higher the mean score, the better speech intelligibility was judged to be.

Table 3.11. Overall score plus ranges for retelling and reading aloud in the pre and post test of the semantic scale judgment.

<table>
<thead>
<tr>
<th></th>
<th>Pre (range)</th>
<th>Post (range)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retelling</td>
<td>4.07 (2.97-4.98)</td>
<td>3.83 (2.64-4.29)</td>
<td>P&lt;.01</td>
</tr>
<tr>
<td>Reading aloud</td>
<td>3.91 (2.98-5.07)</td>
<td>3.94 (3.03-4.97)</td>
<td>NS</td>
</tr>
</tbody>
</table>
Table 3.11 shows that the rating for the retelling task is significantly lower in the post test than in the pre test. The rating for the reading aloud task improves slightly, but not significantly.

In this type of experiment, it is more interesting to look at the separate scales in order to see if listeners distinguish between certain categories or if speakers improve or score worse on particular categories. A Principal Component analysis was performed to see which scales represent the same underlying category. Due to limited space, we cannot discuss the results of the Principal Component analysis at great length. However, in general it was found that listeners do not distinguish between intelligibility and articulation scales, and between rhythm and intonation scales. In addition, the general judgment scale (good-moderate-poor) was always a general judgment on voice quality rather than speech intelligibility. Unfortunately, the contents of the components differed somewhat between the reading aloud and the retelling task, and between the pre and post test. This made it difficult to make comparisons between the pre and post tests. A reduction of scales according to Fagel, Van Herpt & Boves (1983) also did not result in similar components for the pre and post tests. To make comparison possible, it was decided to take the components from the pre test and to compare the scores of the scales belonging to these components with the scores in the post test. The name given to the components is meant to represent the common theme of the component. For that reason, it has been decided to choose the term ‘intelligibility’ for both the intelligibility and the articulation scales. The following tables give the scores per component for the pre and post test.

Table 3.12a. Scores for the different components for the pre and post test for the retelling task, based on the pre test components.

<table>
<thead>
<tr>
<th>Retelling</th>
<th>pre test</th>
<th>post test</th>
<th>sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intelligibility/rate/rhythm</td>
<td>4.35</td>
<td>4.12</td>
<td>p&lt;.01</td>
</tr>
<tr>
<td>Voice quality</td>
<td>2.99</td>
<td>2.82</td>
<td>p&lt;.05</td>
</tr>
<tr>
<td>Livelihood/loudness</td>
<td>4.36</td>
<td>3.94</td>
<td>p&lt;.01</td>
</tr>
</tbody>
</table>

Table 3.12b. Scores for the different components for the pre and post test for the reading aloud task, based on the pre test components.

<table>
<thead>
<tr>
<th>Reading aloud</th>
<th>pre test</th>
<th>post test</th>
<th>sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intelligibility</td>
<td>3.98</td>
<td>4.28</td>
<td>p&lt;.01</td>
</tr>
<tr>
<td>Voice quality</td>
<td>2.85</td>
<td>2.85</td>
<td>NS</td>
</tr>
<tr>
<td>Livelihood/rate</td>
<td>4.46</td>
<td>4.29</td>
<td>p&lt;.05</td>
</tr>
<tr>
<td>Loudness</td>
<td>4.32</td>
<td>3.68</td>
<td>p&lt;.01</td>
</tr>
</tbody>
</table>
From Table 3.12a above it can be seen that intelligibility deteriorates for the retelling task. This component also consists of the scales on rhythm and rate. In the reading aloud task, however, intelligibility improves according to the scores. The voice quality scales deteriorate slightly in the retelling task and stay the same in the reading aloud task. It has to be kept in mind, however, that there were no exercises in the program to improve voice quality. Component 3 of the retelling task includes the scales on loudness and liveliness. Even though lower scores are seen for the post test, it does not necessarily mean the speaker has performed worse, just as for the scales on liveliness and rate (component 3 for the reading aloud task). Clear speech was used in the training program to improve connected speech. This technique advocates a slower speaking rate and a comfortable loudness level without straining the voice, rather than very loud speech. The scores in this test suggest that speakers actually employed this technique correctly, but in a somewhat exaggerated manner. This may have caused their speech to sound more boring or unnatural.

4. Discussion and Conclusions

As was discussed in the introduction, a training program was developed for TE speakers, with its contents based on the most common problems found in TE speech. Where possible, the techniques in the program were based on evidence-based studies, or else on experience-based studies (training manuals). Nine TE speakers received nine one-hour sessions divided over five weeks. In order to study the effect of the training program, several pre and post tests were used, results of which have been described in this paper.

The pre and post tests can roughly be divided into two categories: the phoneme and sentence intelligibility tests, as performed by phonetically trained listeners and speech language pathologists on the one hand, and on the other hand study-specific questionnaires and a semantic scaling experiment. By using different types of experiments with different types of raters, we hoped to create a picture of TE speech intelligibility as complete as possible.

Results showed that consonant intelligibility improved, especially in medial position. The production of the voiced-voiceless distinction also improved significantly and as the most common problem for TE speakers is the production of this distinction, this is a good result. The intelligibility of the (semantically unpredictable) sentences also improved. After only five weeks of training, these results are very positive. However, when the reading aloud and retelling task is taken into account, no improvement can be found. A deterioration in score was even found for the retelling task. At this point, one can only speculate as to why these results were found. Relatively little time was devoted to connected (spontaneous) speech in the therapy program. In only three sessions, subjects were instructed on clear speech and then asked to practice on sentences, short reading passages and short monologues and dialogues. It may be the case that more time is needed for the progression from phoneme/word level to
spontaneous speech. The fact that reading aloud was practiced during therapy probably explains the fact that a (non significant) improvement was found for the reading aloud task, but not for the retelling of the story (which was not practiced). In addition, the use of clear speech may not be an automatic process yet. Another reason for the low scores could be that retelling a story is a difficult task. Even though the subject was told the information did not need to be correct, subjects did try and remember what they had been reading, causing a slow speech rate and many hesitations. And since patients after the therapy were trying harder, the retelling task became more difficult and is even less comparable with running spontaneous speech. In retrospect, it would have been better to let the subjects talk spontaneously about matters that occur in their daily life.

Another positive result, however, was found for the questionnaires as filled out by the speakers themselves. At closer inspection, this result was mainly based on the improvement of two speakers. However, it is difficult with a short questionnaire consisting of 4-point answers (1-4) to obtain significant results. Nevertheless, subjects apparently were more critical about their own intelligibility than their relatives were. After completing the therapy program, however, they felt more comfortable with their speech and this difference of opinion with their relatives disappeared. The fact that the relatives did not improve their scores is likely to be caused by the fact that mean scores were already quite high. Improvement would have meant near perfect scores, which even after intensive training would be very unlikely.

In the introduction, two main research questions were formulated. The first question was if TE speech intelligibility can be improved by specific speech training. This question can be answered affirmatively as training consonant, vowel and sentence intelligibility seems to be beneficial for this group of speakers. However, a better articulation of phonemes alone is not sufficient for intelligible speech and it was found that no improvement could be found for the retelling and reading aloud task. This leads to the second question about the set-up of the program, which is more complicated to answer. With ‘set-up’, both the structure of the program and whether the pre and post tests were appropriate is meant. Where the structure of the program is concerned, adjustments will have to be made, mostly based on the opinion of the SLP’s and the experience of the author. To improve spontaneous speech as well, more time spent on training spontaneous speech intelligibility is required. The best option would probably be to increase the number of sessions. For this, the health insurance authorities have to be convinced and the results obtained with the present program provide good arguments for this increase in the number of sessions. Where the phonemes are concerned, it seems that the right techniques were chosen and that the time spent on training them was sufficient.

It can be concluded that as a result of our study we now have a promising rehabilitation tool to improve TE speech intelligibility: the fact that phoneme
intelligibility improved after only five weeks of training is a positive results that shows that TE speakers can be trained. In addition, all participants were positive about the program and the effect it had on them. Even when no objective improvement could be found, these laryngectomized individuals became more aware of their speech and felt more confident. It is our expectation, based on the results, that overall intelligibility can be improved as well. Based on the evaluation of the therapy program, this program can be optimized further with the aim of achieving the expected effects. And, obviously, the revised program should be evaluated again to see if the aims have been achieved. The main aim of a new evaluation study would be to further validate this training program and to make it available for general use in hospitals for all TE speakers.

5. Acknowledgements

We are grateful to Stichting Breuning Ten Cate for their unrestricted research grant, to Maya van Rossum who developed the therapy program and to Rob van Son who programmed all online experiments.

6. References


Quality of Life After Total Laryngectomy

Tanya L. Eadie

Department of Speech and Hearing Sciences, University of Washington,
Seattle, WA, USA
teadie@u.washington.edu

Abstract

Objectives: To use the framework proposed by the International Classification of Functioning, Disability and Health (ICF) to review outcomes related to quality of life (QOL) after total laryngectomy.

Patients/Materials and Methods: Results from papers published on QOL after total laryngectomy were organized according to the ICF framework. Outcomes were summarized relative to the contribution of disease and treatment-related factors, those related to changes in body functions and structures, activities, and participation, and those related to contextual factors such as family and social support, and personal variables such as coping strategies, age, and gender.

Results: QOL after total laryngectomy is multidimensional. Outcomes are affected by management of the airway (stoma), success of alaryngeal speech and communication, and swallowing. Factors related to treatment (absence or presence of radiation or chemotherapy), those related to one’s environment (family and social support), and/or person (depression, coping strategies, age, and gender) are interactive and also predict outcomes. Longitudinal studies that include psychosocial variables are lacking.

Conclusions: The type of instrument used to assess quality of life, time after diagnosis, and study design must be considered when interpreting QOL postlaryngectomy. There is a continued need to perform multi-center, prospective studies to increase levels of evidence after treatment for advanced laryngeal cancer to ensure the most beneficial treatment for each individual is pursued.

Keywords: quality of life, total laryngectomy, International Classification of Functioning, Disability, and Health (ICF)
1. Introduction

Cancer of the larynx is among the most common cancers of the upper aerodigestive tract. Every year more than 11,300 new diagnoses of laryngeal cancer occur in the United States, with more than 3,660 associated deaths [American Cancer Society, 2007]. Worldwide, incidence and mortality rates are variable, following patterns and changes in exposure to alcohol and tobacco, even still with a 10-15 fold variation in male laryngeal mortality rates between males across European countries [Bosetti et al., 2006]. In the Developed World, 50% to 60% of patients present with early laryngeal cancer, defined by the American Joint Committee on Cancer as a T1 or T2 tumor without nodal involvement or distant metastases [Chen et al., 2006]. Advanced laryngeal cancer, on the other hand, corresponds to stages III and IV of the TNM classification, with current recommendations for management published by the National Comprehensive Cancer Network (NCCN) [National Comprehensive Cancer Network, 2008]. The recommendations are based on results of several phase II and phase III treatment studies, as well as expert opinion, and are constantly revised to assess the impact of therapy on survival and quality of life (QOL) outcomes.

While improved cure rates, prolonged disease-free survival, and organ preservation are the primary focus of the treatment of advanced laryngeal cancer, the implicit purpose of organ preservation is improved function and QOL. However, organ preservation does not necessarily equate to conservation of organ function, nor does it always increase QOL [Genden et al., 2008]. Thus, outcomes related to QOL after treatment for advanced laryngeal cancer should be reviewed with careful consideration of these issues. The difficulties faced by individuals diagnosed and treated for advanced laryngeal cancer are a topic of growing concern. QOL is a multidimensional construct that minimally includes physical, psychological, and social domains of functioning, and applies to the level of one’s general well-being and life satisfaction [Rogers, Fisher, Woolgar, 1999]. QOL has a particular relevance for those diagnosed and treated for advanced laryngeal cancer because of the myriad difficulties these individuals experience with everyday functioning, including difficulties with airway management, speech, and swallowing. However, multiple factors beyond breathing, speech and swallowing affect QOL in individuals with advanced laryngeal cancer, including time since diagnosis, treatment type, methods of coping, mental health, social support, gender and cultural issues, to name but a few [Hanna et al., 2004; Terrell, Fisher, Wolf, 1998; van den Brink et al., 2006; Vilaseca, Chen, Backscheider, 2005]. Consideration of multiple components of a person’s well-being and functioning are, therefore, necessary to facilitate the most effective treatment, rehabilitation, and end-of-life care in this clinical population [Hammerlid et al., 1999].
It is widely agreed that those diagnosed and treated for laryngeal cancer are best served by an interdisciplinary team of health care providers. In this regard, the conceptual framework underlying the World Health Organization’s (WHO) International Classification of Functioning, Disability and Health (ICF) provides a highly appropriate context for the structure of providing services to those with laryngeal cancer [ASHA, 2001; Eadie, 2007; 2003; WHO, 2001]. The primary advantage of the ICF centers around the fact that its basis is a biopsychosocial model of health and functioning; it requires that a variety of concerns are addressed to understand the impact and implications of any given disease or health condition. The usefulness of the ICF is highlighted by the example in which an individual is disease free (e.g., at five or 10 years after cancer diagnosis), yet that person continues to experience difficulty in everyday activities as a result of the cancer and/or its treatment. Measuring a person’s functioning, disability, and health-related QOL thereby requires consideration of multiple factors beyond that of “time of survival”, for example, psychological, social, and emotional status, as well as perceptions of satisfaction found in work, home life, spirituality, family, education, sexual function, or income [Doyle, 1994; Llewellyn, McGurk, Weinman, 2005].

For many years, total laryngectomy (i.e., total removal of the larynx) followed by radiotherapy was considered the standard treatment for advanced laryngeal cancer. In 1991, a prospective randomized clinical trial was performed by the Veterans Affairs (VA) Laryngeal Cancer Study group in the United States [Department of Veterans Affairs Laryngeal Cancer Study Group, 1991]. In this study, more than 300 patients with stage III or IV laryngeal cancer were randomized to 2 arms of treatment: induction chemotherapy followed by radiotherapy (RT), or total laryngectomy followed by radiotherapy. In most patients (62%) in the chemotherapy and RT arm, the larynx was preserved. Of most significance, no differences in survival rates were observed between the nonsurgical and surgical arms of the study. A similarly designed study in Europe also demonstrated no difference in overall survival between these treatment options, along with a 42% laryngeal preservation rate in the nonsurgical arm (Lefebvre et al., 1996). These landmark studies demonstrated the potential value for nonsurgical therapy in achieving laryngeal preservation with survival rates comparable to total laryngectomy (i.e., the “gold standard” at that time).

Recently, Chen and Halpern (2007) reported the results of a study involving more than 7000 patients from a nation hospital-based cancer registry. Individuals in this study included those who had undergone total laryngectomy, RT alone, or combined chemotherapy-RT. Contrary to the results of the prior clinical trials [Lefebvre et al., 1996; VA study group], Chen and Halpern (2007) found that overall, total laryngectomy was significantly associated with increased likelihood of survival compared with RT alone or chemo-RT. Specifically, they found that total laryngectomy was associated with significantly greater survival for those with stage IV disease; among patients with stage III disease, total laryngectomy and chemo-RT had similar impacts on survival, with both of these options...
showing increased survival compared with RT alone. Chen and Halpern (2007) also found that overall survival was decreased among men, black patients (compared to white patients), and patients with Medicare or Medicaid or those who were uninsured (compared with those in private insurance). The authors stated that their findings may have reflected differences in patient mix between those enrolled in previous clinical trials with those in the broader population of patients with advanced laryngeal cancer involved in their study. Results from all of these studies highlight the need to take into consideration patient factors such as performance status, access to support services and close follow-up, as well as institutional factors when selecting a treatment regimen. In addition, results highlight the need and continued appropriate role of total laryngectomy in the treatment of advanced laryngeal cancer, as either a primary or salvage procedure, particularly among those with stage IV disease [Chen, Halpern, 2007]. Finally, even with comparable survival rates among different treatment options, the relative impact of the treatment on the patients' QOL becomes critical in selecting the “optimal” therapeutic approach.

Given that total laryngectomy continues to be a primary or secondary treatment (salvage) for advanced laryngeal cancer, and given the significant biological, psychological, and psychosocial impact after total laryngectomy, it is important to continue to review and evaluate post-laryngectomy outcomes related to QOL. Thus, the objective of this article is to provide a review of the literature related to QOL after total laryngectomy, to identify factors that influence QOL, and most importantly, to identify gaps in our knowledge related to QOL in this population. Although there are different methods for structuring this type of review (i.e., theme analysis, systematic review), the present review will be structured using the framework of the ICF to ensure inclusion of biopsychosocial factors. Through consideration of the ICF and its biopsychosocial model of care, proactive approaches to intervention for those treated for advanced laryngeal cancer will almost certainly emerge and be enhanced over time.

1.1. Using the ICF as a Framework to Measure QOL

In 2001, the WHO proposed a biopsychosocial model that integrates the concepts of health and functioning in the ICF (WHO, 2001). The ICF defines functioning and disability as the interaction between a “Health Condition” and “Contextual factors” relative to three levels: a) “Body Functions and Structures,” b) “Activities,” and c) “Participation.” “Body Functions and Structures” are defined as physiological functions of body systems or anatomical elements such as organs, limbs, and their components; “Activities” are the execution of specific actions, and “Participation” encompasses involvement in life situations (WHO, 2001). All three levels of functioning are therefore affected by contextual factors which include the physical, social, and attitudinal environment in which a person lives (i.e., “Environmental factors”), as well as aspects related to the person, including an individual’s gender, ethnicity, age, fitness, lifestyle, coping, social background, and education (i.e., “personal factors”).
The primary advantage of the ICF is based in its biopsychosocial model of health and functioning. Using this model to understand the impact of a disease or health condition requires the consideration of a number of factors, such as biological, psychological, or social well-being. The ICF model is multidirectional in its explanation of functioning and disability associated with health domains. In brief, the ICF not only explains how disease (e.g., advanced laryngeal cancer) creates an impairment in Body Structure (e.g., total removal of the larynx) and Body Function (e.g., alaryngeal speech), but it also demonstrates how Activity Limitations and Participation Restrictions may arise from social causes (e.g., embarrassment associated with alaryngeal voice and perceived disfigurement creates social isolation) [Eadie, 2007]. Thus, Contextual Factors such as the environment and personal characteristics interact with levels of functioning and influence outcomes. For example, the attitudes of friends or family, or physical barriers (e.g., background noise) in an environment may impact a person’s ability to perform an activity or participate in a life situation. Personal Factors such as age, gender, ethnic and cultural membership, and personal motivation also may affect functioning in all of the other components of the ICF. For example, willingness to engage in social behaviors under the Activities/Participation domain after surgery may be affected by an individual’s motivation to succeed, a Personal Factor. Viewed collectively, the biopsychosocial model can provide a broad framework from which one may evaluate the consequences of laryngeal cancer and its treatment on each individual. Readers are referred to Eadie (2003; 2007) for a more detailed description of the application of the ICF and its components after total laryngectomy.

In addition to providing a framework for functioning and disability, the ICF also provides a theoretical model for understanding QOL. As stated by the WHOQOL Group (1995, p. 1405), there exist primary needs for enquiry into the perception by individuals “of their [own] position in life in the context of the culture and value systems in which they live.” Thus, QOL, and specifically health-related QOL, may be seen as a coalescence of all factors involved in health and functioning within the ICF model. It is a construct which is best measured by individuals with health conditions, who make judgments of their current status and compare them with their own internalized standards of what is possible or ideal [Cella, Tulsky, 1990]. It is this subjectivity that has challenged the validity of QOL outcomes: there is no guarantee that each dimension of QOL measured in the instrument will be important and have the same salience to the respondent. In fact, the same measure of QOL can take on different meanings for individuals with different subjective views of reality [Stennar, Cooper, Skevington, 2003].

Despite some of these challenges in measuring QOL, a number of tools have been developed and validated. They are intended to measure generic health-related QOL, as well as measuring this construct in specialized populations (e.g., head and neck cancer), and for specific purposes (e.g., for voice-related QOL). Thus, to evaluate QOL outcomes associated with total laryngectomy, one must first examine tools that are available for assessing physical, social, and mental
well-being in this population. This summary will provide the basis for the review of QOL outcomes after total laryngectomy, which will follow the review of the assessment tools.

1.2. Common Tools that Measure QOL after Total Laryngectomy

Three types of QOL tools are commonly employed by researchers and clinicians in health-related fields. These tools include: (a) discipline-free measures, which transcend disciplines and generically define a problem, and have typically been used to evaluate health care; (b) disease-specific measures, which are used to measure health and functioning specific to a population, and; (c) discipline-specific measures, which explicitly define difficulties using specialized theoretical frameworks and terminology, and typically focus on particular domains within the scope of a particular discipline [McDowell, 2006]. Table 1 provides a summary of the most commonly used QOL instruments used to evaluate post-laryngectomy QOL. They are listed from top to bottom in order of generic QOL measures, disease-specific (cancer, head and neck cancer) instruments, to discipline-specific measures (e.g., those related to communication and swallowing specific to the disciplines of speech-language pathology (S-LP) and otolaryngology).

Table 1. Most commonly used QOL measures after total laryngectomy, including discipline free, disease-specific, discipline-specific instruments.
<table>
<thead>
<tr>
<th>Year</th>
<th>Author(s)</th>
<th>Type of Instrument</th>
<th>Instrument (Abbreviated)</th>
<th>Instrument Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>Ware et al.</td>
<td>General Health</td>
<td>SF-12</td>
<td>Medical Outcomes Study Short Form 12-item Health Survey</td>
</tr>
<tr>
<td>1993</td>
<td>Ware</td>
<td>General Health</td>
<td>SF-36</td>
<td>Medical Outcomes Study Short Form 36-item Health Survey</td>
</tr>
<tr>
<td>1999</td>
<td>Björdal et al.</td>
<td>Head and Neck Cancer</td>
<td>EORTC-HN35</td>
<td>European Organisation for Research and Treatment of Cancer Head and Neck Quality of Life Questionnaire 35</td>
</tr>
<tr>
<td>1993</td>
<td>Hassan, Weymuller</td>
<td>Head and Neck Cancer</td>
<td>UWQOL</td>
<td>University of Washington Quality of Life Scale</td>
</tr>
<tr>
<td>1997</td>
<td>Terell et al.</td>
<td>Head and Neck Cancer</td>
<td>HNQOL</td>
<td>University of Michigan Quality of Life Questionnaire</td>
</tr>
<tr>
<td>1993</td>
<td>Cella et al.</td>
<td>Cancer (applicable to other)</td>
<td>FACT-G</td>
<td>Functional Assessment of Cancer Therapy – General</td>
</tr>
<tr>
<td>1990</td>
<td>List et al.</td>
<td>Head and Neck Cancer</td>
<td>PSS_HN</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>Jacobson et al.</td>
<td>SLP-Otolaryngology</td>
<td>VHI</td>
<td>Voice Handicap Index</td>
</tr>
<tr>
<td>1999</td>
<td>Hogikvan, Semutheran</td>
<td>SLP-Otolaryngology</td>
<td>V-RQOL</td>
<td>Voice related Quality of Life Scale</td>
</tr>
<tr>
<td>2001</td>
<td>Chen et al.</td>
<td>SLP-Otolaryngology</td>
<td>MDADI</td>
<td>MD Anderson Dysphagia Index</td>
</tr>
</tbody>
</table>
Studies investigating outcomes related to total laryngectomy often include discipline-free measures that permit comparisons between QOL after total laryngectomy to QOL associated with other conditions. For example, many studies have used either the 12-item or 36-item versions of the Medical Outcomes Study Health Survey (SF-12 and SF-36, respectively) [Ware, 1993; Ware et al., 1996] (see Table 1) as a generic health indicator. The SF-36 generates scores for 8 domains: physical functioning, role limitations attributable to physical problems, bodily pain, general health perception, vitality, social functioning, role limitations because of emotional problems, and mental health.

While discipline-free measures are often used to document general health status, disease-specific instruments are often used to supplement information provided by the generic health measures. Disease-specific instruments can further be subdivided into those that are generic to a disease (e.g., cancer), or site specific (e.g., head and neck cancer). Because of the widespread effects on individuals who have undergone treatment for head and neck cancer, most site-specific QOL inventories include domains related to physical, emotional, psychological, and social functioning, and specifically include domains related to speech and swallowing, which may not be reflected on generic health measures. Common instruments used in the QOL literature for advanced laryngeal cancer are listed in Table 1. Because of their strong psychometric properties and frequency of use, it is important to recognize how these measures uniquely contribute to a discussion related to this clinical population.

Recently, Eadie (2007) provided a summary of instruments related to speech-language pathology used to describe functioning total laryngectomy using the ICF framework. Most discipline-specific measures (e.g., speech-language pathology instruments) are focused at evaluating post-laryngectomy communication and swallowing at the Body Functions and Structures level, as well as some at the Activity level. For example, Body Structures may be evaluated using laryngoscopy, while Body Functions may be assessed either directly (e.g., using videostroboscopy to view the function of the pseudoglottis in tracheoesophageal speech production; using fiberendoscopic examination of swallowing (FEES) to assess the function of the pseudoglottis in swallowing), or indirectly (e.g., acoustic measures of fundamental frequency (fo), perturbation measures; auditory-perceptual ratings; nutritional mode, ratings of function during modified barium swallow (MBS) examinations, etc.). In addition, speech impairments may be assessed with variables related to speaking rate (e.g., pause time) or intelligibility of speech sounds (e.g., using articulation measures that identify problematic phonemes). In contrast, communication activities may be assessed with speech intelligibility measures, such as the Sentence Intelligibility Test (SIT) [Yorkston, Buekelman, Tice, 1996], or auditory-perceptual measures (e.g., speech acceptability) that capture how much an alaryngeal voice distracts a communication partner from the content of the message, or functional measures that assess swallowing function with certain consistencies of food [Eadie, 2007]. In some cases, these measures also include interactions with Environmental
Factors (i.e., attitudes of other communication partners or mealtime partners) [Bennett, Weinberg, 1973]. Finally, it is notable that no exclusive measures of participation exist within the discipline of speech-language pathology [Eadie et al., 2006]. Although some discipline-specific measures include questions that measure aspects of Participation (communication or swallowing-related), no measures appear to be solely dedicated to measuring this construct as defined by the ICF [Eadie et al., 2006]. Instead, scales such as the Voice Handicap Index (VHI) [Jacobson et al., 1997], Voice-Related Quality of Life (V-RQOL) scale [Hogikyan, Sethuraman, 1999] measure voice-related QOL, whereas the M. D. Anderson Dysphagia Index [Chen et al., 2001] measures swallowing-related QOL. These measures include specific questions related to Body Functions and Structures, Activities, and Participation, as well as some Personal and Environmental factors that are specific to voice and swallowing. Because of the importance of both voice and swallowing after treatment for advanced cancer, these instruments are being used increasingly as outcome measures that may be more sensitive indicators of function in these areas (see Table 1). Yet, because QOL measures encompass all aspects of functioning across the ICF, it is important to understand the relationships between measures of Body Functions and Structures, Activities, and Participation and measures of QOL.

2. QOL Outcomes after Total Laryngectomy

The following section provides a brief summary of outcomes related to total laryngectomy, including its known impact on QOL. This summary is based upon use of many of the measures listed in Table 1. First, a summary is provided that examines the relationship between disease-and treatment-specific factors to QOL. Second, the roles of Body Functions and Structures, Activities/Participation, and Contextual factors are examined in their relationship with QOL. The review provides a context for recommended approaches to assessment and treatment of this clinical population, as well as implications for future research.

2.1. Effects of Laryngeal Cancer and its Treatment on QOL

Traditional healthcare frameworks have focused on outcomes related to the disease or treatment of a disease. Thus, the bulk of the literature for advanced laryngeal cancer relates to outcomes that focus on effects of the disease or specific treatments of the disease. Many factors related to the disease contribute to QOL after diagnosis and treatment for advanced laryngeal cancer. For example, acute and chronic physical symptoms (dysphagia, xerostomia, speech problems, pain and fatigue) following treatment are of substantial concern to most individuals [Ledeboer et al., 2005]. Symptoms vary by disease stage, treatment modality, and tumor site. For example, de Graeff et al. (2000) found more difficulties in individuals with advanced stage disease, as well as those who had experienced an increased number of treatments. However, more recent studies have failed to find an association between cancer stage and QOL, but
rather treatment type and modalities seem to override stage [Terrell et al., 2004]. For example, the extent of neck dissection has been shown to lead to poorer QOL [Terrell et al., 2004]. One general finding related to site of the disease is that individuals with oropharyngeal tumors tend to report more pain and dysphagia, and those with laryngeal tumors tend to report worse speech scores [Bjordal et al., 1999].

In the past 10 years, the effect of total laryngectomy, with or without radiotherapy (RT) has been studied extensively in comparison to chemoradiation (i.e., organ preservation) protocols [Finizia et al., 1998; Fung et al., 2005; Hanna et al., 2004; Hillman et al., 1998; Terrell et al., 1998]. More recently, studies have investigated the consequences of total laryngectomy versus other surgical approaches, such as supracricoid laryngectomy [Dworken et al., 2003; Weinstein et al., 2001] or partial laryngectomy procedures [Swenaik et al., 2005]. In general, results regarding QOL have been variable due to different research designs (e.g., cross-sectional vs. longitudinal; retrospective vs. prospective), sample sizes, characteristics of the subjects (e.g., laryngectomy as salvage vs. laryngectomy as primary treatment; laryngectomy with or without RT etc.), and time since diagnosis (e.g., immediately post-operatively vs. 3 months vs. 6 months vs. 2 years vs. 10 years etc.). In general, most studies that have used generic health-related QOL instruments have found poorer QOL in individuals who have undergone total laryngectomies when compared to those who have undergone laryngeal preservation protocols or other surgical approaches when QOL is measured directly after treatment, and up to one year after treatment. However, these differences have been found to dissipate over time, with most studies finding no differences in overall QOL between total laryngectomy versus CT and RT or partial laryngectomy many years post-treatment [Deleyiannis et al., 1999; Finizia et al., 1998; Sewnaik et al., 2005; Weymuller et al., 2000].

A few studies have investigated QOL after total laryngectomy using longitudinal designs [Armstrong et al., 2001; Campbell et al., 2000; List et al., 1996; Morton, 2003; Terrell, Fisher, Wolf, 1998; Weymuller et al., 2000]. Armstrong et al. (2001) investigated overall QOL in 34 laryngectomees from the pre-operative stage up to 6 months after surgery using the SF-36 and two non-standardized measures. Individuals in their study not only reported low pre-operative QOL scores, but they continued to exhibit significant and persistent difficulties in communication and swallowing, as well as continued poor emotional status 6 months after their surgery. The researchers concluded that while individuals showed improvement in most domains, that there were significant problems remaining at the 6 month period, suggesting that further rehabilitation was necessary and appropriate. Other studies that have focused on the time immediately following the laryngectomy also suggest that health-related QOL is poorer in those who have undergone total laryngectomy than those in laryngeal preservation protocols [van den Brink et al., 2006]. While some of these difficulties persist over time, they are usually not reflected in overall QOL scores, with many studies demonstrating no significant differences between those who have undergone total laryngectomy.
and those who have been treated with radiotherapy or other organ preservation protocols several years after treatment [Stewart, Chen, Stach, 1998]. For example, Finizia et al. (1998) investigated QOL in 14 patients who used tracheoesophageal speech compared to 14 patients who received radical radiotherapy (RT) for advanced laryngeal cancer. The majority of patients (12 in each group) reported that it had been more than 12 months since their last treatment. The only difference found between the groups for EORTC-QLQ-C30 or EORTC-QLQ-H&N35 scores related to a question about hoarseness, with those in the RT group reporting worse scores. Likewise, Sewnaik et al. (2005) reported no significant differences in QOL between a group of 12 patients who had undergone partial laryngectomies and 11 patients who had undergone total laryngectomies at least one year after their operations.

Although overall health-related QOL scores may not demonstrate differences, some researchers have found differences between total laryngectomy and other treatments for advanced laryngeal cancer when QOL is analyzed using domain scores. Terrell et al. (1998) performed a follow-up investigation to the original survival study by the Veterans Affairs group. In this study, patients were randomized either to the RT plus induction chemotherapy (CT) treatment arm, or to the RT plus total laryngectomy group. Twenty-five individuals from the surgery plus RT group and 21 from the CT plus RT group completed the SF-36 and the HNQOL instruments, the Beck Depression Inventory, as well as smoking and alcohol consumption surveys 10 years post-treatment. Terrell et al. (1998) found that patients with intact larynges had significantly less bodily pain (SF-36), better mental health (SF-36), as well as better HNQOL emotion scores. More patients in the total laryngectomy group also were found to be depressed. Although Terrell et al. (1998) found some differences between the two treatment groups using the generic QOL (SF-36) instrument, others have found no differences in overall QOL scores using similar instruments, or core questionnaires (e.g., cancer specific) [Finizia et al., 1998; Hanna et al., 2004]. In fact, Vilaseca et al. (2006) found that 49 patients who had undergone total laryngectomy more than 2 years post-treatment, showed very few differences in their overall health-related QOL scores (SF-12) than normal controls.

While overall QOL may not be affected, differences in QOL consistently have been found between those who have undergone total laryngectomy versus those who have undergone laryngeal preservation protocols using disease-specific scales that also permit analysis of subdomains. Hanna et al. (2004) reviewed 42 patients with advanced stage III or IV laryngeal cancer who were treated either with concurrent chemoradiotherapy or total laryngectomy with post-operative radiation therapy. QOL was measured using the EORTC-QLC-C30, as well as the EORTC-QLQ-H&N35. They found on the core questionnaire, that there were no statistically significant differences in overall QOL between the two groups. However, when subscales were analyzed using the head and neck cancer specific module, they found that those who had undergone total laryngectomy reported significantly greater difficulties with smell and taste, use of painkillers,
and coughing, whereas those treated with chemoradiotherapy reported significantly more problems with dry mouth. The authors concluded that both chemoradiation and total laryngectomy affect QOL of patients treated for advanced laryngeal cancer in different ways, and that these differences could only be revealed by functional and subscale analysis, and not in the overall QOL scores. Similarly, Sewnaik et al. (2005) found the only differences in QOL of patients after partial laryngectomy versus total laryngectomy was found to be smell and taste related (as measured by the EORTC-QLQ-H&N35), with the total laryngectomy group performing more poorly.

Results from these studies reveal an interesting trend: while functional disabilities persist, overall QOL does not seem to be affected after total laryngectomy after a 12 month period of adjustment. Interestingly, results obtained using disease-specific scales have not shown consistent differences related to speech or swallowing when total laryngectomy is compared to laryngeal preservation protocols many years post-treatment [Hanna et al., 2004; Terrell et al., 2004]. These results are surprising because difficulties related to airway management, speech, and swallowing are among those impairments most frequently reported after treatment for advanced laryngeal cancer [Gritz et al., 1999; List et al., 1996]. There are several reasons why these results have been observed. First, disease-specific QOL instruments may not be sensitive indicators of how much a person’s voice, communication, or swallowing affects QOL. Thus, it may be important to include discipline-specific QOL instruments to measure these outcomes.

In the past 10 years, several studies have included voice-related QOL instruments such as the VRQOL or VHI in their evaluation of QOL outcomes after total laryngectomy [Dworkin et al., 2003; Fung et al., 2005; Kazi et al., 2006; 2007; Stewart et al., 1998; Schuster et al., 2004; Sewnaik et al., 2005; Weinstein et al., 2001]. Results from these studies suggest that after total laryngectomy, there is a continued moderate difficulty with voice; outcomes after total laryngectomy are significantly poorer than those reported for laryngeal preservation protocols [Fung et al. 2005; Stewart et al., 1998], or healthy controls [Fung et al., 2005; Schuster et al., 2004; Weinstein et al. 2001]. Inconsistent results have been found between individuals who have undergone total laryngectomies with those who have undergone partial laryngectomies; some studies reported similar scores [Dworkin et al., 2003; Sewnaik et al., 2005], while one study found that voice-related QOL was poorer after total laryngectomy than supracricoid partial laryngectomy [Weinstein et al., 2001]. Schuster et al. (2004) compared voice-related QOL in 20 male laryngectomees who used tracheoesophageal speech with individuals who had voice disorders related to organic and functional etiologies, as well as healthy controls. The average total VHI scores for the laryngectomy group was 45.5 (SD = 24.10) (i.e., a moderate handicap), which was significantly higher than patients with functional voice disorders, but differed only slightly from patients with organic laryngeal dysphonia. Results from these studies show that while voice-related QOL may be affected after total laryngectomy, there is significant variability across individuals,
with some alaryngeal speakers reporting scores that are just as good, or even better than, individuals with preserved larynges [Schuster et al., 2004; Stewart et al., 1998]. Factors that may impact results include method of postlaryngectomy communication, age, gender of the speaker, presence of other physical problems, or overall health. While instrument sensitivity may impact interpretation of outcomes after total laryngectomy, a related complicating factor in interpreting changes related to QOL lies in the patient’s ability to adapt to their illness, in addition to coping with the after-effects of treatment for cancer. This “response shift” can be a valuable strategy for coping with the reality of chronic changes by recalibrating one’s expectations for health and functioning and the relative valuation of health states. The result can also be that changes in patient-rated QOL over time do not correspond to objective measures of loss of function, or disability.

2.2. Effects of Body Functions and Structures, Activities, and Participation on QOL

By applying the ICF framework, one may not only address consequences of the disease, but one may evaluate the function of the individual at three levels (Body Functions and Structures, Activities, and Participation), as well as their relationship with QOL. For example, total laryngectomy involves complete surgical removal of the larynx (i.e., a severe structural impairment), resulting in complete loss of laryngeal voice production (i.e., a severe impairment of function). The trachea is brought forward and sutured at the base of the neck, forming a tracheostoma. Consequently, the individual no longer inspires or expires through his mouth or nasal cavities, which also may lead to an impaired sense of smell. In addition to changes in phonation and breathing, the individual also may experience changes in taste, shoulder function, laughing, crying, and swallowing [Rohe, 1986]. All of these physical changes have the potential to impact one’s QOL. For example, difficulties with breathing have led to development of heat-moisture exchanger systems; with use, these systems may positively impact QOL [Hilgers, Ackerstaff, 2000]. Other functional impairments that may affect QOL include eating, swallowing, voice and speech. Outcomes are reviewed in the next sections.

2.2.1. Functions of Eating and Swallowing

Only a few studies have explored the impact of total laryngectomy and other types of treatments on eating and swallowing, though incidence of dysphagia following total laryngectomy has been reported to range from 10% to 60% [Balfe et al., 1982; Crary & Glowasky, 1996]. The inconsistency in incidence figures seems to stem from the various definitions of “successful” swallowing, as well as the time at which the assessments take place. Some researchers have equated swallowing “success” as oral intake without supplements. For example, Fung et al. (2005) evaluated the impact of organ preservation therapy with or without total laryngectomy as salvage on voice and swallowing function. They found that while
swallowing function was good in all patients, those with an intact larynx were more likely to obtain nutrition by oral intake without supplements, thereby showing an advantage over total laryngectomy. Dworkin et al. (2003) evaluated a series of subjects treated for advanced laryngeal cancer with supracricoid laryngectomy (n=10) versus total laryngectomy (n=10). Patients were evaluated using FEES, MBS, and other acoustic and aerodynamic measures, as well as listener impressions of speech and voice. They found that patients from both groups performed comparably with regard to speech intelligibility and voice quality. However, whereas individuals in the total laryngectomy group did not demonstrate the need for protracted exercises to remove their feeding tubes postoperatively, those in the supracricoid laryngectomy group required many sessions of swallowing therapy to obtain this objective and eliminate tube feeding supplementation. This result is important because distress in patients and caregivers has been found to be related to presence of a feeding tube [Verdonck-de Leeuw et al., 2007].

Finally, some researchers have defined swallowing “success” as ability to manage a normal diet. Ackerstaff et al. (1994) evaluated the functional disorders and lifestyle changes following total laryngectomy and found that as many as 25% of patients reported alterations to their diet, including the avoidance of certain consistencies as well modifications to their style of eating. Hillman et al. (1998) noted that 76% of their laryngectomy patients at 24 months post-surgery managed a normal diet, while Ward et al. (2002) reported that 58% of laryngectomy patients managed a normal diet, at least one year after their surgery. Ward et al. (2002) further reported that patients experiencing long-term dysphagia identified significantly increased levels of disability, handicap and distress. Interestingly, they found some patients who presented with the ability to manage a normal diet at follow-up who still experienced a mild level of handicap and distress related to their swallowing function. Interviews with these patients revealed that although these individuals could manage a normal diet, compensations such as taking additional time to complete a meal were negatively perceived. Other psychological and emotional issues (e.g., lack of smell and taste of foods) also were reported as negatively impacting the social pleasures of swallowing and eating.

2.2.2. Voice and Speech Function

Voice and speech may be restored after total laryngectomy through training and use of alternative methods of verbal communication, including esophageal, tracheoesophageal (TE), and electrolaryngeal speech. The success of these methods is often determined by evaluating the acoustic and perceptual parameters of speech for the dimensions of fo, sound pressure level, quality (e.g., perturbation measures), speaking rate, and intelligibility of speech sounds. Several studies have noted that both female esophageal and TE speakers may have difficulty achieving gender-appropriate fo levels [Eadie, 2003]. Esophageal speakers also may have lower than normal speaking rates because the
esophagus has to be recharged with air throughout speech production [Doyle, 1994]. In addition, reduced speech intelligibility often occurs with use of the electrolarynx particularly for the manner of voicing [Weiss & Basili, 1985], although it is rare to find any alaryngeal speaker who is 100% intelligible [Hillman et al., 1998].

When alaryngeal speech samples are presented to listeners, they identify alaryngeal speech as significantly less intelligible, acceptable and natural than normal laryngeal speech [Finizia et al., 1999; Pindzola & Cain, 1988]. The reduced intelligibility and acceptability of the speech signal by listeners might be hypothesized to penalize individuals who use alaryngeal speakers in their interactions with communication partners, thereby decreasing QOL. Yet, the relationship between both objective and subjective dimensions of alaryngeal voice and speech to QOL varies across studies. Meyer et al. (2004) investigated the relationship between word and speech intelligibility scores from a standardized instrument and QOL as measured by the UWQOL, FACT, FACT-H&N and PSS-HN. They found that in their group of individuals who had undergone total laryngectomy (n=16), there were no significant relationships between word or speech intelligibility measures with any QOL measures. Other studies have examined the relationship between voice quality, acceptability, and QOL after total laryngectomy. Finizia et al. (1999) found that upon self- and listener-assessments of voice quality and acceptability, TE speakers were judged to be less acceptable and worse in voice quality than both normal speakers and speakers who had undergone RT. Yet, these differences did not seem to translate to differences in QOL, at least as measured by generic or diseasespecific instruments [Finizia et al., 1998]. Eadie and Doyle (2004) investigated listeners’ perceptions of 28 TE speakers and the relationship between these measures and self-rated QOL, as measured by the HNQOL. Listeners’ perceptions of voice quality for speech acceptability, naturalness, voice quality, and pleasantness were only mildly to moderately related to HNQOL subscale scores (Pearson’s r ranged from .378 to .494). Thus, while results from auditory-perceptual studies would suggest that alaryngeal speakers remain identifiably different than normal speakers, these differences do not necessarily translate into poorer QOL.

Some investigators have hypothesized that post-laryngectomy communication mode may affect QOL. For example, because TE speakers are usually judged as being most intelligible and pleasant when compared to esophageal and artificial laryngeal speakers, it might be expected that they might exhibit the best QOL outcomes of any type of alaryngeal speaker. Some researchers have found that type of alaryngeal speech method may affect QOL outcomes. Finizia and Bergman (2001) reported no significant differences in QOL between those who used TE speech and those who had been treated with radical radiotherapy. However, they reported that a small group of speakers who used an electrolarynx (n=5) showed higher levels of dysfunction and emotional distress when compared with the TE group and those treated with RT. The authors found it
difficult to make definitive conclusions about how TE speakers function compared to electrolaryngeal speakers due to the small sample size. However, they hypothesized that speech rehabilitation with TE prostheses after laryngectomy may be as effective as conservative treatment with RT for advanced laryngeal cancer relative to communication and psychosocial adjustment. Eadie and Doyle (2005) similarly hypothesized that strong QOL scores (HNQOL) reported by 30 men who used TE speech in their study could have been due to a sample of exclusive TE speakers (vs. all types of communicators in the study by Terrell et al. (2004)). However, not all studies support the notion that TE speakers are the most successful among individuals who use an alaryngeal communication mode [Blood et al., 1992]. For example, Carr et al. (2000) concluded that there did not seem to be a measurable improvement in one’s QOL or ability to communicate using TE speech over any other post-laryngectomy communication method.

One reason why strong relationship may not be found between speech intelligibility or voice quality to QOL may relate to the types of instruments used to measure QOL. Specifically, if studies had included measures of voice-related QOL, it is unknown whether these relationships would have changed. Op de Coul et al. (2005) examined the sensitivity of EORTC questionnaires (both QLQ-C30 and QLQ-H&N35) to voice-related problems in 80 laryngectomized individuals who used TE speech. They found a good overall and voice specific-QOL level using the EORTC questionnaires. However, they found that use of these questionnaires masked some voice-related concerns, which were only revealed with use of additional, symptom-specific questionnaires. For example, despite an overall satisfaction with many aspects of voice (e.g., intelligibility, loudness, pitch, fluency), in their study more than half of the subjects (63%) reported serious problems with communicating in noisy environments. That is, had they used a measure such as VHI, perhaps these problems would have been detected. Several researchers have noted poorer functioning after total laryngectomy for voice-related QOL when compared to laryngeal preservation or partial laryngectomy procedures [Fung et al. 2005; Stewart et al., 1998; Weinstein et al. 2001]. However, it is equally possible that strong relationships between speech, voice, and QOL do not exist because after many years of living with total laryngectomy, individuals are bothered more by the physical consequences of surgery (i.e., presence of tracheostoma) and interference with social activities and social participation than impaired communication per se.

2.2.3. Activities and Participation and QOL

Total laryngectomy can affect one’s communication in a variety of social settings (e.g., at work, in the community), which may isolate an individual. It also may lead to disruptions in mealtimes with family and friends. Further, the presence of a tracheostoma may increase the risk of work disability [Taylor et al., 2004], or alter or reduce the frequency, ability, or enjoyment of performing activities of daily living (e.g., cannot blow on hot soup; bathing and eating may be altered) or participating in leisure activities (e.g., skiing; gardening; singing in a choir).
The relatively weak relationship observed between speech, voice, and swallowing impairments and QOL may reflect differences in how these constructs are measured. For example, most impairments (reduced or altered Body Functions and Structures) are measured by clinicians using instrumental as well as perceptual measures, whereas participation, and QOL are measured from the perspective of the individual with the health condition. For example, Palmer and Graham (2004) investigated relationships between demographic variables as well as functional abilities with overall QOL after total laryngectomy. They found that self-perceived ability to communicate exhibited the strongest relationship with QOL \( (r = .52) \). In addition, they found that individuals who communicated with others several to many times daily reported higher QOL scores than those who reported infrequent communication. These results indicate that social interaction and participation in social activities, rather than speech abilities or swallowing per se may be more important indicators of QOL [Mohide et al., 1992]. Unfortunately, as noted by Eadie (2007), there are fewer instruments available for measuring disruptions in activities and participation than impairments, which limits the study of the relationship between participation restrictions and QOL. A few studies have used adapted scales to measure functional outcomes. Hillman et al. (1998) assessed functional outcomes related to communication, swallowing and eating, and employment status using a Communication Profile (CP). The CP consisted of 24 statements about the individual’s reaction to various communication situations/environments (e.g., “I avoid answering the telephone”), as well as data related to living environment, employment, swallowing and eating, and extent of speech and voice therapy. At the end of a 24-month rehabilitation period, CP scores for TE and esophageal speakers were slightly higher than or comparable to those achieved pre-surgery. Thus, although some “functional” disabilities persist, most domains related to QOL eventually return to pre-treatment levels after 12-36 months, including those related to communication [Bjordal et al., 2001; Deleyiannis et al., 1999]. Overall, results support the notion that even loss of the larynx may not impact overall QOL, but that QOL is a construct that changes over time, is specific to each person, and is affected by multiple factors.

### 2.3. Effect of Contextual Factors on QOL

One of the reasons why outcomes related to QOL after total laryngectomy are often inconsistent relates to the effect of a great number of contextual factors on functioning at other levels. These contextual factors include environmental facilitators/barriers such as one’s family, friends, and the attitudes of the society in which one lives. Personal factors include an individual’s coping strategies, age, gender, and mental health. Llewellyn et al. (2005) conducted a systematic review to evaluate studies investigating relationships between psychosocial/behavioral factors with health-related QOL in individuals with head and neck cancer. They found that in the 16 studies they reviewed, four main factors impacted QOL. These included: (a) personality (a personal factor), (b) social support (environmental factor), (c) satisfaction with consultation and information (interaction between person and environment), and (d) behavioral factors (e.g.,
alcohol consumption, smoking). In addition, depression levels before treatment were predictors of depression post-treatment, and QOL post-treatment. These results of this study highlight the importance of considering all of these factors when interpreting QOL outcomes.

2.3.1. Environmental Factors: Social Support

The importance of environmental factors on outcomes is highlighted by research showing that post-treatment distress is proportional to the extent of communication problems with friends and relatives [Friedman et al., 1994]. In contrast, openness to discussing malignancy in the family, emotional support, and a larger social network may positively affect rehabilitation outcomes [Mesters et al., 1997]. Verdonck-de Leeuw et al. (2007) found that distress in patients who had been treated for head and neck cancer, including laryngeal cancer, was related not only to the presence of a feeding tube, speech, and swallowing problems, but also to a passive coping style, fewer social contacts and non-expression of emotions. While speech intelligibility or voice quality alone may not relate strongly to overall QOL after total laryngectomy, communication is a necessary component of socialization. Thus, communication abilities and satisfaction with communication in everyday activities may more strongly relate to one’s QOL by affecting relationships with others. For example, Palmer and Graham (2004) found that individuals who communicated with others several to many times daily reported higher QOL scores than those who reported infrequent communication. Others have suggested that it is not the number of people involved in the social support after total laryngectomy, but rather, how satisfied individuals are with their support in terms of perceived availability from family, friends, and professionals [Richardson, Bourque, 1985]. One study reported that being single was also predictive of poorer QOL in the first 3 months following laryngectomy [van den Brink et al., 2006]. Results suggest that social interactions and interpersonal activities strongly and positively affect the QOL for those who are laryngectomized. In addition, familiarity of the communication partner and social support, as well as education of family members may increase success, whereas unfamiliar partners may reduce communicative participation and emotional functioning [Law, 2005; Mathieson, Stam, Scott, 1990; Zeine, Larson, 1999]. Likewise, changes in one’s physical appearance may result in alterations in body-image and self-concept, which may affect one’s socialization and QOL. These results have clear implications for interdisciplinary health care teams, and warrant consideration in general symptom assessment (i.e., beyond those that are solely physical in origin).

2.3.2. Personal Factors: Coping, adjustment, and depression

An individual’s ability to cope, as well as his or her mental health may affect one’s success post-laryngectomy. Blood et al. (1992) investigated coping, adjustment, self-esteem, perceived communication abilities, and preferred communication modes of 41 individuals who had undergone total laryngectomy. Patients also
were judged by external listeners for speech acceptability. Results revealed that individuals who were well adjusted also tended to view their voice rehabilitation as more beneficial and rated themselves as more easily understood by listeners than those subjects who demonstrated psychological distress and poor adjustment. Blood et al. also found a relationship among speech acceptability and adjustment, coping, self-esteem, and general well-being. Specifically, voice acceptability ratings were positively correlated with problem-focused and seeking-social-support coping strategies. Further, self-esteem and general well-being both showed positive correlation with voice acceptability ratings, with the highest correlations being found for voice acceptability and self-esteem ($r = .63$). However, low-to-moderate correlations were found for adjustment, self-esteem, coping, and general well-being. This led the investigators to suggest that individuals with good adjustment skills may be more optimistic about their voices, and that “Positive relationships between perceptual ratings and adjustment and coping encourage speech-language pathologists to continue their role in ‘holistic’ treatment of the patient” (p.68).

Several studies also have investigated the relationship between depressive symptoms and QOL. Hammerlid et al. (1999) studied mental distress in individuals with head and neck cancer, and found that one-third had mood disorders during their first year. De Leeuw et al. (2001) found that individuals with pre-existent depressive symptoms and women were more likely to develop depression after 6-36 months. De Graeff et al. (2000) also showed that depression and low performance status are risk factors for physical and psychological morbidity after treatment for head and neck cancer, and that these symptoms are predictive of QOL. Results from several studies suggest that depression is one factor that should be assessed pre- and post-operatively, and that psychosocial intervention may be warranted to improve QOL in this population [Hammerlid et al., 1999; Llewellyn et al., 2005].

2.3.3. Personal Factors: Gender

Several studies have investigated the differential impact of gender on outcomes. First, results do not show a consistent difference between men and women with regard to depression [Katz et al., 2003]. Although some studies report higher levels of depression in women than men associated with disfigurement, these effects seem to be buffered by levels of social support. However, women indicate differential means of support than men, with women who are laryngectomized relying on family, reading, church services, and non-laryngectomized social groups as primary support systems [Graham, Palmer, 2002]. In contrast, men rely on laryngectomy group meetings, their work setting, and reading.

Second, studies that have investigated differences in overall QOL scores between men and women after total laryngectomy have shown inconsistent results. Some studies have shown that there are no differences between men and women with regard to QOL, and that gender has no predictive value for QOL
outcomes, or even voice-related QOL [Eadie, Doyle, 2004; de Graeff et al., 2000; Kazi et al., 2007; Palmer, Graham, 2004], although women showed increased voice handicap in one study [Kazi et al., 2006]. Still, other studies have suggested some differences in how women and men experience symptoms [Vilaseca et al., 2006]. Despite the failure to find such differences using QOL instruments, some researchers have found that listeners differentially penalize female alaryngeal speakers more than male speakers when they make judgments of speech acceptability [Eadie, Doyle, 2004; Eadie et al., 2008]. If these attitudes are reflected in their communication partners, the quality of social interactions may be experienced differently by male and female laryngectomees.

One qualitative study included five women who were treated for head and neck cancer, including two women who had undergone total laryngectomies. Results showed that all women experienced significant physical limitations that interfered with their ability to carry out daily tasks. All women described the ability to cope and adapt to changes in appearance and disfigurement, as well as their ability to maintain a positive attitude, and the importance of sharing their experiences with others [White, 2004]. Results from these studies demonstrate the complex relationship between gender and QOL that may produce differential effects on levels of function [Katz et al., 2003].

2.3.4. Personal Factors: Age

There is a paucity of data to suggest that age alone is associated with a higher risk of surgical complications. However, while age alone (in lieu of co-morbidities) may not impact surgical morbidity, it has a complicated relationship with QOL. For example, it is often assume that elderly individuals fare worse after treatment, but few data support this view [de Graeff et al., 2000]. In fact, some studies have shown an opposite effect, with younger patients more likely to express dissatisfaction with appearance and anxiety [Vilaseca et al., 2006]. Likewise, Woodard, Oplatek and Petruzzelli (2007) found better functional and QOL outcomes in at least one domain for individuals who were older than 65 years. They hypothesized that younger patients may have had better pre-treatment functional status, and therefore, when they are subjected to an extensive surgical procedure, their baseline functional status drastically declines. In contrast, older patients who may be retired may have had a more sedentary pre-operative lifestyle. Consequently, there may be little change in their baseline functional status after surgery. Therefore, their perception of QOL may not be as negatively affected. To fully evaluate this hypothesis, more longitudinal studies that include both pre- and post-operative measures should be performed.

3. Discussion

3.1. Implications for Future Research and Clinical Practice

The importance of using QOL questionnaires in evaluating individuals treated with laryngectomy has grown in the past decade because it shows promise as a
means of deciding between treatments when no survival advantage is afforded by one modality over another. In addition, QOL outcomes have become more mainstream as biopsychosocial models of health have gained acceptance. However, there are major obstacles to performing research on QOL outcomes in general, including those that are practical and those that are clearly methodological in origin. Practical considerations include the amount of time and resources required to collect and analyze large amounts of questionnaire data, the difficulty in selecting an appropriate instrument, and the bias created by missing data on those most severely affected by their disease. That is, there is a survivor bias in that individuals with recurrent disease are often excluded from evaluation; excluding these individuals from analysis results in artificially inflated QOL scores [Schwartz et al., 2001]. Methodological considerations include the inability of instruments to capture an individual’s adaptation to the conditions of their disease state, the difficulties in comparing outcomes across populations, and, lack of interpretation of what QOL score differences mean for clinical practice. In addition, specific domain scores may not be sensitive enough to detect real differences in function with small sample sizes. One part of this problem could be addressed by designing studies that investigate smaller, yet better defined groups of speakers. In addition, multi-center studies that investigate outcomes in a prospective, longitudinal design are necessary so that outcomes are better supported. These designs would allow better determination of relationships between psychosocial variables and how they may predict QOL at different times after surgery.

Results from this review showed that many factors (e.g., depression, coping skills, social support) interact with functioning and disability at all levels and lead to either poorer or enhanced QOL. These factors are often overlooked in traditional assessment measures, including QOL outcomes, and must be understood to facilitate optimal care in this population. However, not enough is known about the contribution of these factors at present, and therefore, we must employ additional methods (e.g., qualitative methods) to identify and define such areas of concern. In addition, not enough is known about how personal factors, such as cultural background, education, socioeconomic status, and gender affect QOL. These factors are important to investigate as the demographic characteristics of those who undergo total laryngectomy change, and to ensure that outcomes may be compared across countries and healthcare settings [Morton, 2003]. Through such endeavors, clinicians and researchers may better understand the influence of these factors on QOL, leading to more effective rehabilitation and consequently, better outcomes.

4. Conclusions

All individuals with advanced laryngeal cancer who undergo total laryngectomy will face numerous challenges, difficulties, and learn to cope with significant changes and restrictions in functioning following treatment. A successful program of short- and long-term rehabilitation must be designed to comprehensively
assess and provide effective intervention for these individuals. These considerations go beyond that of the narrow outcome of “time of survival” and need to include myriad issues related to QOL. By endorsing a comprehensive approach to assessment and intervention such as that proposed by the ICF, we are sure to continue to make meaningful changes in the lives of individuals who have undergone total laryngectomies.

5. Acknowledgements

I would like to acknowledge my mentor, Dr. Philip Doyle, for his helpful comments on a prior version of this manuscript. I would also like to thank Robyn Schwab and Amanda Politziner who helped with its organization.

6. References


Oral and Oropharyngeal Cancer: Speech Outcomes and the Need to Monitor Change Over Time

Philip C. Doyle, Ph.D.

Voice Production and Perception Laboratory, Department of Otolaryngology and Reconstructive Surgery, Schulich School of Medicine and Dentistry, Doctoral Program in Rehabilitation Sciences, and Communication Sciences and Disorders, University of Western Ontario, London, ON, Canada
pdoyle@uwo.ca

Abstract

Objectives: Despite the long history of surgery as the primary treatment for cancer of the oropharyngeal structures, a relative paucity of information on speech outcomes currently exists. While the importance of this voice and speech performance is clearly identified as an essential outcome along with chewing and swallowing, there remains inconsistency in how outcomes for speech are obtained and interpreted. It is clear that surgical treatment for oral and oropharyngeal tumors and subsequent reconstruction has a direct influence on the integrity of the human vocal tract. As a result, surgery and reconstruction holds the significant potential for alteration of signal transmission characteristics of the vocal tract. This presentation provides a brief summary of existing data on outcomes and serves to raise several questions related to a need for research. Further, preliminary data obtained in our laboratory using established methodology to assess the acoustic structure of English vowels in individuals who have undergone surgery for oral cancer is presented for consideration.

Patients/Materials and Methods: Following a cursory review of the literature, results from a preliminary study of 11 participants (3 women, 8 men) who had received surgery for treatment of oral cancer are presented. All participants had undergone surgery for oral tumors and received subsequent reconstruction using a radial free forearm flap. A group of age- and gender-matched control participants were also studied. All participants produced multiple productions of 10 English vowels consistent with methods outlined by Peterson and Barney (1952); formant frequencies were then determined via spectrographic analysis.

Results: Data obtained revealed remarkable similarity across the speaker groups for values obtained for formants 1, 2, and 3 (F1, F2, and F3). When proportional relationships between F1 and F2 were compared, no differences between surgical and control participants were identified. Overall, and in the presence of substantial surgical reconstruction, these 11 surgical participants
appear capable of modifying their vocal tracts so as to adequately code the basic and essential acoustic requirements for the 10 vowels studied.

**Conclusions:** These preliminary findings demonstrate that traditional and fundamental methods of vocal tract assessment, namely determination of formant frequencies through acoustic analysis, can serve as a valuable index of postsurgical vocal tract function and signal transmission. These acoustic data represent one specific measure of vocal tract transmission characteristics that could be exploited in additional analyses of the postsurgical system. The lack of significant differences between these groups does not, however, suggest that the speaker groups are equivalent with respect to overall speech production capabilities. The current data provide an initial step toward establishing standardized methods of post-treatment speech evaluation.

**Key words:** oral cancer, oropharyngeal cancer, vocal tract, speech acoustics, speech rehabilitation

1. Introduction

The treatment of oral and oropharyngeal malignancies poses numerous and often considerable challenges that cross multiple areas of concern. As with all malignant conditions and the associated surgical treatment options, aspects of oncologic safety are paramount. Medical management will focus on eliminating disease, but intricately tied to the concern of oncologic safety comes a desire to retain functional capacity (Teichgraeber, Bowman, & Goepfert, 1986) and optimize “quality of life” (Doyle, 1994). When considering the functional capacity of the upper airway, primary considerations have typically centered on aspects of verbal communication, deglutition, and swallowing; beyond issues related to survival, these three functional domains are widely agreed to comprise facets of critical importance to outcomes in those treated for oral and oropharyngeal cancers (Logemann et al., 1993; Matthews & Lampe, 2005). These areas also are linked inextricably to those broad domains of functioning that are most frequently addressed in relation to QOL, namely, myriad facets of physical, psychological, and social capacity (Doyle, 1994; Myers, 2005; Rieger, Dickson, Lemire, et al., 2006). Yet the ability to speak, as well as to chew and swallow without difficulty, remain at the forefront of any effort to document the success of cancer treatment regardless of modality (Urade et al., 1987).

The determination of a successful or unsuccessful outcome in those diagnosed with oral and oropharyngeal malignancies must always be conceptualized in a multidimensional fashion. However, for the purpose of this paper, the information presented is directly focused on the distinct area of speech capacity in the post-treatment period. The emphasis on verbal communication in this paper should not be interpreted to imply that deglutition and swallowing (or other areas of communication functioning) are of less importance—they are not; collectively, these functions are critical elements of outcome measurement that seeks to provide evidence for treatment effectiveness and success and its impact
on the individual. Nevertheless, the dynamic processes associated with the
generation of voice and speech and ultimately, the ability to systematically
quantify and monitor the speech signal, require careful attention. Thus, this
rather focused topic will be addressed exclusively in the remainder of this paper.
Consequently, several key areas will be identified in an effort to provide a
framework for the documentation of potential post-treatment changes, as well as
a means of evaluating ongoing changes in such deficits. Perhaps the most
important issue of concern in the context of speech outcomes is directly related
to inconsistent and variable methods of evaluating outcomes. As a result,
“evidence” becomes inconsistent both in its presentation and interpretation.
Dylan et al. (2007) reviewed 12 articles that reported some level of speech
outcome, but identified that there “…were few studies that included precise
objective outcome measures.” (p. 3). One issue observed by these authors is
that some oropharyngeal resections are complicated by not only their inherent
surgical variability, but may also create additional functional complexities when
laryngectomy and/or surgical-prosthetic voice restoration (tracheoesophageal
[TE] puncture) is involved. It is important to note, however, that despite these
complexities, many studies have ultimately applied a practical approach to
assessment that is directed toward one’s global communication activities for daily
tasks such as phone use. Though broad, such approaches may offer more
refined, individual profiles of post-treatment speech (communication) outcomes
as a marker of evidence.

In order to facilitate the discussion to follow, a cursory overview of several
issues is necessary. This will initially involve a discussion of: (1) oral and
oropharyngeal structures and the composite structures and functions of the vocal
tract, (2) surgical intervention for oral cancer and the subsequent reconstruction
of surgical defects that result from tumor excision, and finally, (3) the ability to
monitor changes to the vocal tract acoustically in the context of post-treatment
follow-up. Throughout these sections, a selective review of information from the
literature will be provided. Although the information included in this review is not
in any manner exhaustive, information provided is judged to provide a framework
for data presented later in the paper. Next, and as a direct outgrowth of this
work, some suggestions for future research will be offered for consideration.
Finally, although references to specific reconstruction methods may be noted
along the way, it is beyond the scope of this written document to address the
differences in those surgical techniques in detail at this time.

1.1. Cancer and the Oral and Pharyngeal Cavities

The oral and oropharyngeal cavities are comprised of numerous, highly
integrated anatomical structures which lie superior to the inlet of the voice
generation mechanism of the larynx. Consequently, this important region exhibits
substantial complexity both anatomically and physiologically. Basic structures
include the lip, alveolus, teeth, floor of mouth, buccal tissue, palate, velum,
tongue, and pharynx, as well as the region of the tonsillar bed, base of tongue,
maxilla, and mandible. Collectively, these structures ultimately comprise the
vocal tract, a variable length tube (cavity) that exists from vocal folds to lips (Kent, 1993; Kent & Read, 1992). The vocal tract is a dynamic region that varies considerably in response to linguistic requirements and the acoustic representations dictated by those requirements for the purpose of verbal communication. Because of the critical nature of the vocal tract on speech production, alterations in the vocal tract hold considerable importance relative to subsequent alterations in the acoustic speech signal that represents the end product of speech. More specifically, abnormalities in the structural integrity of the vocal tract, concurrent changes in its physiologic function, and compensatory strategies that may be employed to delimit such changes must be considered in a multidimensional fashion (Morrish, 1988). However, in regard to any efforts to evaluate speech outcomes in this heterogenous population, it is apparent that standards have not been clearly defined. One critical shortcoming of much of the work conducted in the area of speech outcomes in those treated for oral and oropharyngeal cancers relates to the fact that no standard approach or protocol appears to be used for speech assessment. While a number of studies in the literature do indeed use sound, rigorous methodology and a rational approach to gathering and providing outcome measures on speech, considerable differences do exist. Additionally, terminology used and the descriptors employed are also problematic. In fact, Lorenz and Alam (2003) have identified a tendency for speech assessments to be based on somewhat vague definitions such as “...all had speech intelligible enough for communication” (p. 234). Interestingly, these authors also note past concerns in the literature that suggest the importance of speech “quality” on measure of speech intelligibility proper. That is, while intelligibility may be the ultimate target or a given assessment (Kent, Weismer, Kent, & Rosenbek, 1989), and in fact is the true end product of speech communication, changes in the overall “quality” of the voice/speech signal may provide a substantial confound to judgments made and/or their interpretation.

While many methods employed for measuring speech outcomes are indeed reasonable and sound from a procedural and methodological point of view, the inherent degree of variability in methods from study-to-study is considerable. Consequently, the ability to reasonably and validly compare across literature contributions is often quite limited. In essence, the variability in tumors treated, differences in the amount of tissue resected, the amount and type of reconstruction utilized, multiple modalities of treatment, concerns of medical complications and associated treatment morbidities, as well as numerous other factors makes such comparison daunting at best. Work by Clark et al (2006) has offered a detailed evaluation of operative morbidity associated with pharyngeal reconstruction in an attempt to identify and describe early and late complications. The data obtained provided predictive data on complications and morbidity in relation to the extent of surgical defect, type of reconstruction, and method of initial treatment. These data would appear to provide valuable information which suggests that speech outcomes might be better understood if one can understand more about the details of surgery and associated complications;
differences are indeed substantial from individual to individual—no two surgical procedures are identical.

Uwiera, Seikaly, Rieger, Chau, and Harris (2004) reported somewhat paradoxical information on those undergoing hemiglossectomy; while a significant difference was observed between pre- and postoperative intelligibility measures at the single-word level, no differences were identified for the sentence materials. These data confirm the relative limitations of using particular speech materials in measuring post-treatment speech outcomes. Minimally Uwiera et al. provide support for using multiple measures of intelligibility (Lariviere, Seilo, & Dimick, 1975). This finding raises additional questions concerning the need for developing standard methods of intelligibility assessment in this population. In the absence of standardized approaches, development of a credible evidence base will be restricted and our ultimate ability to accurate detect outcome will be restricted. With this caveat provided, the information to follow may provide an initial snapshot of more obvious concerns and areas of research need in association with treatment of oral and oropharyngeal cancers.

1.2. The Vocal Tract

When seeking to understand the functional integrity of the vocal tract, many factors come into play. The vocal tract functions as a dynamic region that in large part permits the uniqueness of the speech signal for each individual. It is the combined relationship between the larynx and the vocal tract that makes each speaker’s verbal communication a unique phenomenon. In the context of oral and oropharyngeal cancers, two additional considerations emerge. Changes in the vocal tract always hold the potential to alter the uniqueness of one’s speech communication. First, the general structural integrity of the vocal tract must always be considered when addressing cancer treatment for tumors within the oropharyngeal system. Second, one also must consider the ability of the postsurgical system to grossly and finely function so as to allow the vast complexity of adjustments. While adaptation and compensation is common and not unexpected, restrictions will exist. Changes to the physiologic capacity of the vocal tract have a direct influence on the rapid and highly coordinated movements necessary for intelligible speech. Degradation in the overall function of the vocal tract creates numerous obstacles that will influence the character of speech produced.

Concerns about what defines “acceptable” speech outcomes has been raised by Butler and Lewin (2004). The primary issue pertains to the fact that in some instances, intelligibility may be defined very broadly, ranging from somewhat vague descriptions to those that are much more elaborate and operationally

---

1 Although the present commentary specifically addresses the functional capacity of the vocal tract secondary to surgical management of cancer, these concerns are of equal importance to other treatment modalities such as radiation and chemoradiation. All methods of cancer treatment hold the potential to disrupt normal structure and function of the vocal tract. Consequently, the methods described herein are equally applicable to other treatment populations.
defined indices of change. For example, according to Butler and Lewin, intelligibility can be judged according to whether it was “understandable by the medical personnel without the need for interpretation by the family, multiple requests for repetition, or written clarification of what was said.” (p. 500-501). Although this definition does provide a means of broadly classifying intelligibility and serves as a generalized index of communication ability, it may lend itself to varied interpretation and, hence, create problems in objectively indexing speech performance. Loss of such definitions creates the possibility for the inaccurately interpretation of existing data, and perhaps more importantly, its comparison to new data.

In regard to the problems encountered in accurately quantifying one’s post-treatment speech capabilities, some reports have provided rather contradictory information relative to speech “intelligibility” (Pigno & Funk, 2003). More specifically, in their case report, Pigno and Funk note that their patient’s speech had improved, being “intelligible but noticeably in error.” (p. 121). Though this information does support some generalized index of the level of communicative function specific to speech, a description of the “errors” noted may hold substantial importance in understanding the nature of deficits that occur and the potential for rehabilitation to improve speech intelligibility. Small deficits may create substantial problems even in the presence of compensation. Similarly, some studies have adopted use of a global 5-category rating system for classifying general intelligibility (Chow, Chan, Chow, Fung, & Lam, 2007). While categorical or cluster ratings do hold some generic assessment value, it does not always characterize the collective entities that comprise speech and how they merge together as a coherent unit. Further, the net effect of such changes to the ear of the listener are essentially unknown. Thus, fine details as well as “gestalt” measures must be considered in the context of generating evidence of outcome in those with oral and oropharyngeal cancers. Doing so can merge the speaker and the listener as a dynamic, dyadic unit specific to communication effectiveness.

Consequently, evaluation of the functioning of both the vocal tract proper and the end-product of speech as judged by the listener in the post-treatment period is essential. Doing so may serve as an objective means of determining component deficits that may then correlate with reductions in gross speech intelligibility and its component parts. One must acknowledge, however, that changes in the function of the vocal tract may involve considerable adaptation over time (perhaps even over several years) and there appears to be little if any data on this topic. Single session observations are unlikely to provide evidence that will withstand scrutiny relative to external validity. As a result, it is imperative that efforts to document acoustic, and ideally auditory-perceptual correlates of acoustic change in those treated for oral and oropharyngeal malignancies, be performed over reasonable periods of time; measurement over time is absolutely essential. Doing so will facilitate our ability to understand ongoing changes that are the result of the loss of anatomic tissue and the degree and extent of the
surgical defect(s), approaches to reconstruction, and longer-term changes due to healing, reduction in tissue swelling, concurrent influences of combined treatments, residual influences of treatment (e.g., the formation of scar tissue), and the individual’s ability to adapt and compensate for dynamic speech restrictions that occur (Doyle & Keith, 2005). Yet discrepancies in how intelligibility is defined and how such definitions overlay to one’s functional capacity persist and continue to create challenges in the interpretation of our evidence. For example, in regard to a group of participants who had been treated for aggressive orofacial cancer with submental flap reconstruction, Chow et al (2007) noted that “…all retained normal conversational function, except for two who were able to communicate with intelligible speech. (p. 500). Statements such as this, coupled with limitations in procedural standards for measuring outcomes, add further challenges to fully understanding the subsequent functional outcomes in this diverse population.

Changes within the vocal tract exist in numerous planes and ultimately, the key issues of how the vocal tract works to create specific acoustic structures necessary for “normal” speech are found primarily in the length of the vocal tract and the varied cross-sectional areas over that length. In essence, during speech the vocal tract is continually (and rapidly) being modified in shape via varied degrees of constriction as a result of tongue positioning, jaw movement, and lip shape (e.g., rounding). The net effect of these changes create varied “cavities”, each of which permits further acoustic modification of the fundamental sound source that emanates from oscillation of the vocal folds. Changes may be viewed from the perspective of simple acoustics (e.g., extent of frequency change), to more dynamic and global areas of functioning that are reflected in measures of speech intelligibility, to the composite characteristics that define each speaker.

Haughey, Taylor, and Fuller (2002) evaluated 43 subjects who had undergone resection and fasciocutaneous free flap reconstruction for cancer of the tongue and floor of mouth. Thirty of these individuals received reconstruction of the oral tongue while 13 had reconstructions of the tongue base. Overall, they reported that speech intelligibility scores were significantly better for those who received tongue base reconstructions (98% intelligibility) when compared to those with reconstruction to the oral-tongue (76% intelligibility). This work also suggested that 20% of participants experienced either medical or surgical complications. Assessment of speech intelligibility was derived from averaged scores based on a 50-item word list with judgments made by five naïve listeners. From these data, substantial variability was noted with scores ranging from 20-100%. Results revealed a mean intelligibility score of 98% for those undergoing tongue base procedures with a mean of 68% for those with oral tongue reconstructions. This amount of variability has been reported previously in the literature relative to anterior tongue defects in association with relatively low mean intelligibility scores (Bokhari, & Wang, 2007; Rentschler & Mann, 1980; Leonard, 1983; Leonard & Gillis, 1990; Mah et al., 1996; Pauloski et al., 1993, 1998).
Chien, Su, Hwang, et al. (2006) presented data for a group of 20 participants who received total glossectomy for base of tongue cancers and 7 who had subtotal glossectomy for tumors of the oral tongue. The reconstruction method of choice for these participants was either a radial forearm free flap (RFFF) or anterolateral thigh flap. Based on their data, Chien et al. indicated that approximately 90% demonstrated “intelligible” speech postoperatively. Similarly, Biglioli, Liviero, Frigerio, et al. (2006) reported information on a group of individuals who had undergone partial glossectomy and RFFF reconstruction as a means of preserving nerve function for the tongue. Overall, increased satisfaction was reported by those receiving neurofasciocutaneous (sensate flaps) procedures when compared to those undergoing non-sensate reconstruction methods. In fact, the level of satisfaction was almost double for those in the sensate group (78% vs. 43%). The sensate flap group also was found to exhibit higher “articulation” scores via objective evaluation; however, only 44% of participants in that group demonstrated such results, thus raising numerous questions about the methods of assessment and the associated influence of the extent and degree of tumor excision and reconstruction. Further, questions related to the potential interactions between speech performance and satisfaction can be reasonable raised and beg the question of determining if speech performance and satisfaction are indeed linked.

Seikely et al. (2003) reported findings from 27 participants who were diagnosed and treated for oropharyngeal cancer. Within this group, 18 provided complete data for both speech and swallowing outcomes. Speech data were obtained prior to surgery, at one month before radiotherapy was initiated, and at some point between 6-9 months postsurgery and following the completion of radiation. The results of this work suggest that the amount of tissue resected for base of tongue tumors did not have any significant influence on either speech or swallowing. However, from the standpoint of speech acoustics and aerodynamics, Seikaly et al. recommend that additional study is warranted for those that undergo resection of the soft palate of half or greater. Although post-treatment intelligibility measures were good, the nature of stimuli used was deemed to be of importance. Once again, differential findings for word intelligibility and sentence intelligibility were noted; this raises questions related to the content of the stimuli used for assessment of speech intelligibility as it is an essential factor to consider in determining evidence-based outcomes. Finally, these authors did note the importance of “drop out” of participants on the data obtained. The concern of participant drop out rates have been raised previously in the clinical literature (Colangelo, Logemann, Rademaker et al., 1999) and such considerations may carry substantial value in assessing the range of true outcomes in multiple areas of functioning following treatment. If individuals cannot be monitored when at their worst, can the data generated be validly interpreted?

Determining meaningful measures offers a considerable challenge in the clinical population who undergo treatment for oral and oropharyngeal cancers and no
tried and true approach is universally accepted. Therefore, our research group in the Voice Production and Perception Laboratory (VPPL) at the University of Western Ontario has initiated a preliminary approach in efforts to assist our understanding of the postsurgical speech production system with an emphasis on vocal tract transmission characteristics. Although we are keenly interested in the end-product of speech and its interaction with the listener as part of the auditory-perceptual process, our efforts to date have primarily focused on quantitative evaluation and analysis of the acoustic signal at the vowel level of performance. The work to be presented herein has been directed toward efforts to objectively document the influence of surgical treatment on the production of English vowels. As noted earlier in this paper, our focus on vowel production does not discount the importance of other areas including the larger concern of one’s overall level of speech intelligibility, and/or general aspects of communicative effectiveness as a critical means of social interaction, and of course, the larger dimension of quality of life. Yet our interest in vowel generation provides what our group believes is an important and fundamental index of postsurgical functional speech capacity for those undergoing surgery for cancers of the oral region.

In the work to be described, it was anticipated that intrinsic acoustic properties of a vowel (i.e., formant frequencies) might be degraded due to the extent of the surgical resection, tissue reconstruction, or development of scar tissue and/or decreased mobility of the tongue or other moveable structures within the vocal tract. These factors also are likely to influence one’s overall speech intelligibility and have in fact been documented for many years (Skelly, 1973). Several studies have reported that speech intelligibility in those who had undergone partial glossectomy was inversely proportional to the amount of tissue removed and the subsequent adequacy of the tongue-palate valve (Bokhari & Wang, 2007; Dios et al, 1994; Michi et al, 1989; Michiwaki et al, 1990; Soutar & McGregor, 1986; and others).

Changes in one’s ability to successfully “valve” regions within the vocal tract in the postsurgical period have been a fruitful area of study. In order to optimize reconstruction following total glossectomy in an effort to reduce speech deficits, Sanger, et al (2000) modified the forearm flap method of reconstruction to include use of a portion of the brachioradialis muscle. Their decision to augment the basic flap procedure was driven by observations that simple flaps could reliably resurface tongue defects, but they did not appear to adequately permit palate and tongue relationships necessary for efficient speech. Thus, the use of the brachioradialis was intended to decrease deficits associated with posterior bulk in the vocal tract. Through this alteration in the flap procedure, it was anticipated that speech would be improved, potentially in relation to aerodynamic parameters and oral-nasal relationships for speech.

Gerden et al. (2003) have provided data that is quite comprehensive relative to speech. These researchers evaluated speech capacity via use of several methods of assessment for oronasal relationships for speech, along with self-
perceptions of speech using a questionnaire format. This work was carefully conducted and confirmed that issues of nasality (alterations in the nasal resonance of the speech signal) must be considered in those undergoing palatomaxillary resections with reconstruction. The importance of such work has been addressed in comprehensive fashion by Leeper, Gratton, Lapointe, and Armstrong (2005). However, various reconstructive surgeries have emerged as a means of restoring articulatory function, and in fact, seek to enable the vocal tract to mimic a somewhat normal system. Such changes, even if minor, hold the distinct and real possibility of creating dynamic deficits in the transmission pathway with a direct bearing on the acoustic characteristics of vowels, as well as other speech sounds. In a follow-up study, Gerden, Wallace, Okay, and Urken (2004) reported findings on 12 individuals who had undergone RFFF reconstruction for defects of the hard palate. Additionally, 8 individuals who had been fitted with an obturator also were evaluated. Data gathered indicated that those who had undergone flap reconstruction reported higher satisfaction for speech, comfort, and social interactions. However, individualized details of performance were not reported. In summary, it would appear clear that the variability are the data and that efforts to describe individual performance in greater detail would be of substantial value in understanding outcomes (Muldowney, Cohen, Porto, & Maisel, 1987).

It is important at this point to specify that our concerns relative to the potential acoustic changes are derived from two, overlapping questions secondary to surgical treatment for oropharyngeal cancer. First, the obvious issue of structural change due to the extent of tissue ablation always must be considered. The clinical and experimental literatures attest to the highly variable changes that exist as a consequence of surgery in the oropharyngeal region and its subsequent reconstruction. It is clear that no two surgical procedures are identical in this population. Further, despite occasional inferences of “similarity” for a given procedure, both the short- and long-term outcomes across many dimensions of “oral” functioning will be characterized by considerable variance; surgical groups are not homogenous. This variability also carries with it opportunity for idiosyncratic behavioral changes that may either positively augment oral function for speech (and perhaps other oral functions), or detract from it. However, we have been particularly concerned with the influence of reconstruction of extensive surgical defects, particularly those that comprise substantial surface areas in the oral and oropharyngeal regions. In such instances, our interest has been further expanded in situations where a relatively large free flap may be used to reconstruct the surgical defect that exists following tumor excision. Thus, the shape, size, configuration and interaction with other structures (e.g., the tongue) may be altered substantially. Based on these factors and an underlying desire to monitor how the postsurgical oral system and vocal tract functions under the demands of verbal communication, it has been the intent of our group to approach the problem in a stepwise manner that utilizes established, but relatively simple acoustic measures as evidence for describing the functional integrity of the postsurgical vocal tract.
The information to follow represents one small, and in fact what some might justifiably argue is a rather narrow, reductionist approach to understanding this population relative to speech. But we view it as a starting point and as such, this initial phase of the project has served the needs of a larger and longer-term ongoing program of research. The application of acoustic measurements certainly offer us the ability to systematically monitor changes in the postsurgical vocal tract under relative static conditions (i.e., during vowel production). Yet beyond acoustic measures, and although not covered in this discussion, we have also attempted to look ahead and consider the important area of auditory-perceptual evaluation as part of generating evidence on the success of flap reconstruction procedures. This approach would appear to offer considerable advantages; it not only has translational value in that it allows for longstanding and well-accepted “bench research” methods to be viewed in a clearly applied fashion, but more importantly, this work may ultimately provide a straightforward means for linking acoustics and the perceptual consequences of surgical treatment for cancers of the oral and oropharyngeal regions.

1.3. Monitoring and Indexing Change Secondary to Cancer Surgery

Postsurgical alteration of the oral and oropharyngeal cavities may be judged perceptually and/or by quantifying the acoustic character of the vocal tract via a number of acoustic measures. These acoustic measures cross frequency, intensity, and temporal domains. Acoustical measures of the composition and representation of energy concentrations in the resonated vocal signal allow for the direct analysis of characteristics of the vocal and speech signal (Baken, 1987; Kent, 1979; Kent, 1993; Kent & Read, 1992; Peterson & Barney, 1952). This type of assessment was initially reported in the seminal work reported by Peterson and Barney (1952). Peterson and Barney presented speakers with a list of 10 monosyllabic words each beginning with the glottal phoneme /h/ and ending with the voiced alveolar stop consonant /d/; the interconsonantal phoneme was a vowel. The variation in this structured production task was created by altering the vowel component within the /h_d/ stimulus structure (e.g., /hI_d/, /hæ_d/, etc.). From these stimuli, Peterson and Barney (1952) asked groups of adult men and women and children to produce stimulus tokens. Productions were then acoustically analyzed to determine the resonant frequencies of the vocal tract (i.e., formant frequencies) in association with the different vowels produced. The underlying premise of this concept is found in the fact that the tongue works to change the general cavity shape(s) and cross-sectional area of the vocal tract. It should be noted at this point, however, that two additional physiologic features are also critical to vocal tract shape and the specific resonance characteristics associated with any vowel, namely, lip-rounding, and the excursion of mandibular movement. Collectively, cooperative and highly integrated adjustments of the tongue, lips, and jaw permit the vocal tract to be modified into a variety of shapes for the purpose of speech production (Kent, 1993; Kent & Read, 1992).
Using Peterson and Barney’s (1953) method, the purpose of the present study was to quantitatively examine acoustic structure directly related to changes in vocal tract transmission (i.e., transfer function). The potential for changes in acoustic transmission would appear to be heightened following intraoral reconstruction via RFFF reconstruction in individuals who have undergone surgery for oral cancer (Meyerson, Johnson, & Weitzman, 1980). It was hypothesized that acoustic changes in postoperative speech would be directly related to surgical changes in the vocal tract. It was anticipated that vowel production would, at least to some degree, be degraded acoustically due to the extent of the resection, reconstruction, the development of scar tissue and decreased mobility of the tongue within the vocal tract. However, various reconstructive surgeries have been used as a means of restoring articulation via increased tongue mobility. Such changes should have direct bearing on the acoustic characteristics of vowels.

Secondary reconstructive surgeries that have been evaluated in the literature include the RFFF, microvascular tongue reconstruction, and tongue mobilizations to name a few (Salibian et al, 1993). Based on several reports in the literature, secondary reconstructive surgeries are suggested to be successful in increasing speech intelligibility (Dios et al, 1994; Michi et al, 1989; Michiwaki et al, 1990). Similarly, however, Fletcher (1988) reported that in spite of reports of individuals with partial glossectomy having “reasonably intelligible’ speech” (p.232), quantitative data (e.g., spectrographic analysis) that specifically defines altered articulatory parameters and patterns of articulation are lacking. Consequently, the current study evaluated quantitative acoustic measures of vowel production in a small, selected sample of individuals who had undergone surgical resection for oral cancer and subsequent reconstruction of the surgical defect. As such, these data may be used to ascertain changes following RFFF surgery on the resonant qualities of the oral cavity for vowel production. The specific questions addressed were: (1) What are the descriptive acoustic characteristics of F1, F2, and F3 for vowel productions by individuals with intraoral resections who have undergone reconstruction via the RFFF?, and (2) Do differences exist in the proportional relationship between formants (F1:F2) for vowels produced by experimental participants when compared to age-and gender-matched normal control participants?

2. Patients/Materials and Methods

Participants for this phase of a continuing project included 11 adults (3 females and 8 males (age 50-75 years, M = 59). All had undergone oral tumor excision and reconstructive surgery using the RFFF; marginal or segmental resection of the mandible was not an exclusionary criterion. Potential participants were excluded from consideration if they had undergone more than a single surgery for oral cancer, had received pre- or postoperative radiotherapy, or exhibited tumor recurrence. Eleven age- (+/- 3 years) and gender-matched normal control participants were also identified and evaluated. All control and experimental
participants met the following criteria for inclusion: no reported history of speech problems, head or neck surgery, hearing impairments that would affect speech production, or upper respiratory infection at the time of testing.

2.1. Procedures and Stimuli

Recordings were obtained in a sound-treated audiometric suite on research quality instrumentation. Each participant was asked to produce 3 trials each of all 10 vowels according to methods described by Peterson and Barney (1952). These stimuli were part of a larger recording protocol, but all were collected at the start of the experimental recording for all participants.

2.2. Data Analysis

The current data analysis centered solely on the inter-consonantal vowel contained in the stimuli. Initially, however, a listener assessment was carried out in order to confirm that vowel productions for each of the speakers who had undergone surgery for oral cancer were representative of the intended target. Forty samples of vowel productions were randomly selected to determine if the sample produced was representative of the intended vowel. Random selection was performed in order to control for fatigue and/or practice effects related to repeated productions of each vowel. From this set of 40 vowels, one naive listener was asked to identify the vowel by selecting it from a closed response set of all 10 target words (e.g., hid, had, head, etc.); these items were listed in normal English orthographies and the listener circled his choice following a single presentation. Listener responses were used to establish the acceptability of vowels for later acoustic analysis. The listener correctly identified 90% (36/40) of the samples presented. Of the four errors, two were attributed to a single participant (SOC 4), a 58 year old male with resection of the right lateral floor of mouth, segmental resection of the body of the mandible, and right radical neck dissection with a RFFF. It is worth noting that this gentleman presented with a slight, although clearly identifiable British accent that may have influenced errors in perception on the part of the listener. Overall, the high level of correct identification confirmed that the intended target vowels were produced appropriately and that acoustic analysis could be pursued. Based on the extent of reconstruction, we had anticipated that more of the samples would be incorrectly identified as part of the representativeness assessment. However, these results confirm that acoustic analyses could shed further light on formant structure and vocal tract acoustics.

2.2.1. Acoustic analysis

Acoustic measures of formant frequencies (F1, F2, F3) were obtained on all vowel samples for all participants in both groups. The acoustic analysis
techniques utilized are well established in the scientific literature (Kent, 1979; Kent & Read, 1992). Measures gathered were obtained using CSPEECH 4.0 (Milenkovic & Read, 1994). Initially, each stimulus (e.g., “hid”) was digitized at 20 kHz. Using a spectrographic subroutine of CSPEECH, the vowel segment of the stimuli was identified and segmented using cursors. The experimenter then listened to the segment to insure that only the vowel was present in the segment. Once confirmation occurred, an automatic acoustic analysis and formant tracking routine was initially performed on a broadband spectrogram. In instances where automatic tracking of the formants was not believed to be accurate based on visual evaluation (i.e., the visual identification of clear, distinct energy bands on the spectrographic display), the procedure was carried out by hand, with measures being taken from the midpoint of the vowel segment where energy bands were noted. Frequency values for F1, F2, and F3 were then gathered and recorded for each vowel.

The dependent variables, F1, F2, and F3, were derived by analyzing all three samples of each vowel produced by participants who had undergone tumor resection and reconstructive RFFF surgery of the oral cavity and their age- and gender-matched speakers who had not undergone surgery. Thus, a total of 30 measures (3 samples x 10 vowels) were obtained from each of the participating subjects. This resulted in the analysis of 330 samples for each group (i.e., 3 samples x 10 vowels x 11 participants). Since three formant frequency values were obtained for each vowel sample, a total of 90 values were recorded for each participant (i.e., 3 productions x 10 vowels x 3 formant measures). Once obtained, the three measures recorded for each of the formants F1, F2, and F3 were averaged for each of the 10 vowels for comparison across subjects. Using the measures obtained from analysis of F1 and F2 for each of the three samples for all 10 vowels, the respective F1:F2 proportional values were calculated and a mean ratio for each speaker (by individual vowel) was obtained. The proportional values for F1:F2 and mean values for F1 and F2 were then statistically analyzed. A total of 33 measures for each vowel were obtained (1 vowel X 3 samples X 11 speakers) with a total of 330 measures obtained for each participant group.

2.2.2. Reliability – Formant Frequencies

Approximately 15% of vowel measures for F1, F2, and F3 were randomly selected and re-measured for reliability assessments. These samples were re-analyzed by an independent external judge trained in the same analysis procedures. Agreement was determined using a point-by-point method (i.e., # of agreements/total # of measurements x 100). As we did not anticipate identical measures to be demonstrated given the nature of the acoustic analyses, we judged measures that were within 10Hz of the original measures to be in “identical” agreement. Based on reliability analyses, measures for 40% of the surgical group and 37% for controls resulted in identical frequency values when compared to the original data. Of the remaining samples, differences were quantified in 10Hz increments. That is, when a difference was observed between
measures of a formant frequency, the difference was quantified according to whether the second measure was within a specified Hz level of the original (e.g., 11-20 Hz difference, etc.). Using this method of quantifying the degree of reliability for surgical participants, 44% of the remaining measures (n=34) were within 20Hz of the original measures; only 12% of these samples exceeded 51 Hz of the original measures. Therefore, inter-rater reliability of measures within 50 Hz of the original measures was 94% for the surgery group. For control participants, 76% were within 20 Hz of original measures, with 14% between 21-50 Hz; only 3 measures exceeded a difference of 51 Hz from the original. Hence, for the control group inter-rater reliability for measures within 50Hz of the original measures also was 94%.

When reliability measures were evaluated by formant frequency, the following findings were obtained. For surgery participants reliability measures, F1 = 55%, F2 =38%, and F3 =30% identical measures when compared to original measures. Inter-rater reliability for measures within 50 Hz of the original measures was 94%, 94% and 89% for F1, F2, and F3, respectively. For controls, F1= 50%, F2= 44%, F3= 17% identical measures compared to original measures. Inter-rater reliability for measures within 20 Hz of the original measures was 100% for F1, 100% within 30 Hz for F2, and 83% within 50 Hz for F3, respectively. Based on the combined results of this reliability assessment, it is believed that the acoustic measures obtained were gathered with good reliability across measures.

2.3. Statistical Analysis

From the acoustic measures of formant frequencies (F1, F2, and F3) obtained for each vowel, means, standard deviations, and the range were calculated. These data were maintained as individual data for the subjects in order to characterize individual variation, particularly in the oral cancer group. The mean values of F1 and F2 for each vowel produced by each speaker were then collapsed and used for statistical analysis between the two groups. From the absolute F1 and F2 values, a proportional (ratio) measure of formant structure (e.g., F1:F2) was calculated and a mean ratio for each speaker and each vowel was generated. Due to the independence of each of the vowel measures, statistical analyses were conducted using analysis of variance procedures with participant speakers serving as the independent variable and vowel measures the dependent variable. An a priori significance level of p < 0.05 was used.

3. Results

3.1. Formant Frequencies – Minima and Maxima Values

Mean measures for F1, F2, and F3 values for each vowel and speaker group (surgical vs. normal controls) are presented in Table 1 for comparison. As can be seen, surgical participants exhibited slightly greater mean frequency values
for both F1 and F2. For F1, these differences ranged from 6Hz (“head”) to 37Hz (“who’d” and “heard”). For F2, frequency differences ranged from 25Hz (“Had”) to 176Hz (“who’d”). For F3, control participants exhibited higher frequency values only for “had” and “who’d”, although these differences were not substantial. The current data for both surgical and control participants do compare favorably with normative data reported by Peterson and Barney (1952). However, it should be noted that all data for our speakers were collapsed across gender, whereas males and females were segmented by Peterson and Barney. Despite this difference, the similarity of our data provide support for the notion that even in the presence of significant reconstruction following oral tumor excision, physiologic demands for the accurate production of vowels has been met.

3.2. F1:F2 Proportional Values

The mean proportional relationship values (F1:F2) were calculated for both participant groups for each vowel. A summary of these results are also presented in Table 1. While most speaker pairs (i.e., surgery vs. control) exhibited some discrepancies in absolute F1 and F2 resonant frequencies across all 10 vowels, the mean proportional values of F1:F2 for the two groups remained quite similar. That is, those who had undergone surgery were able to code the resonant acoustic properties of vowels as well as their controls based on the ratio data generated. This finding was evident in some comparisons of individual speaker pairs where absolute frequencies differed, but where proportionally relationships for F1 and F2 were retained. Therefore, it is evident that the speakers who had undergone surgery for oral cancer and RFFF retained the ability to produce F1:F2 proportional resonances which consistently translated into well-coded vowels. This finding is consistent with results of the representativeness assessment.

4. Discussion and Conclusions

These data, though clearly preliminary in nature, offer some early insights into vowel production in those treated for oral cancer. As noted, our observation that surgical participants were able to effectively meet expectations for formant frequencies and the proportional relationships between F1 and F2 so critical for perception was unexpected. However, one observation that was of even greater surprise was that related to acoustic measures of particular vowels. For example, our data confirm that our surgical participants were able to meet the acoustic requirements of “point vowels” (/i, a, u/). These three vowels require the greatest degree of physiological excursion, therefore, provide information on the overall integrity of oral structures. It is evident that some speakers who had undergone surgery for oral cancer were more limited than others in their ability to achieve these extreme points of constriction. However, their ability to retain general proportional relationships for F1 and F2 that are so important for accurate perceptual identification was observed.
At a very basic level of production, these surgical participants were collectively able to meet the acoustical requirements of all 10 vowels investigated. Thus, despite what we would anticipate to be highly variable and individualized postsurgical vocal tract structure, the capacity to manipulate the system to achieve production of distinct vowels has been documented. While the external validity of these findings certainly must be addressed with caution, it does appear that one’s ability to generate vowels with high levels of accuracy could foster a framework for investigating more detailed aspects of consonant production in dynamic speech stimuli. Minimally, if an acoustic index of vowel structure can be identified for a given speaker, it may serve as a valuable point of reference for acoustic measures of other speech entities (e.g., when paired with consonants) with potential application to other measures of dynamic speech events (acoustic transitions, etc.) and additional methods of acoustic analysis. The present data provide one facet of a rich opportunity for monitoring the acoustic outcomes of surgical treatment for malignancies of the oropharyngeal region. The ability to measure the acoustic changes that exist following surgery may then be compared to a well-established body of normative data that has existed for many years. The value of these acoustic measures is that they may also serve as a foundation for auditory-perceptual evaluation of postsurgical speech proficiency.

4.1. Future Directions – Acquisition of Evidence

Efforts in our laboratory continue to address the complex problems of assessing vocal tract dynamics and its effect on speech production and perception despite the challenges that persist. Because of complexities associated with the acquisition of evidence to document outcomes and comprehensively understand the influence of surgical treatment on vocal tract functioning, several research areas are currently under investigation by our group. While individual and often narrow pathways of exploration related to voice and speech production remain at the forefront of this work, the ability to link these objective data to psychophysical measures obtained through auditory-perceptual evaluation is essential. In adopting this combined approach, in the past year we have adopted the concept of “acoustic signature” as a means of evaluating those who receive surgery for oropharyngeal cancers (Kent, Kent, Rosenbek, Weismer, Martin, Sufit, & Brooks, 1992; Weismer, Kent, Hodge, & Martin, 1988). Although acoustic signature research to date has typically been applied to intelligibility issues at the word or sentence level, we have attempted to apply this approach at the vowel level as well. In some situations, we have attempted to obtain acoustic information in the presurgical period in order to provide direct evidence of outcome on the vocal tract secondary to treatment. This approach, however, does have clear and real limitations, as well as some potential ethical considerations. Most importantly, and because of the presence of a mass lesion and the potential influence of such a lesion on the physiological integrity of the articulatory structures and vocal tract, the “baseline” that is often obtained is itself non-normal. Hence, comparative evaluations at multiple points in the post-treatment period are referenced to a potentially abnormal standard. In spite of this limitation, the ability to track
changes in the system over time can serve to reflect positive changes that occur in response to factors directly associated with healing such as decreased edema of vocal tract tissues, increased levels of structural mobility, etc. In this vein, we believe that evidence provided through acoustic tracking at the level of vowel and beyond may provide an array of measures that provide tangible evidence to the individual who has undergone surgery and must deal with multiple deficits, most notably in the early months of recovery. In this regard, we do concur with Fletcher's (1988) suggestion that the use of spectrographic analysis holds promise. In any case, whether simple or more elaborate measures are extracted from spectrographic representations, we also believe that use of acoustic measures provides a relatively simple method of addressing changes in the vocal tract that occur following radiation therapy and/or chemoradiation protocols.

Another area of recent exploration with those who undergo treatment for oropharyngeal malignancies has involved attempts to apply common acoustic measures in the frequency, intensity, and temporal domains in a combined manner so as to obtain information on relative contributions of particular features to the end product of speech. One of the concerns within this line of inquiry relates to changes that can be noted in the early period of recovery. For example, and not infrequently, as speech capabilities return to the individual in the postoperative period, it has been observed that vocal intensity may be reduced. It is suspected that these changes may reflect an easy and “careful” approach to using the speech production mechanism (both laryngeal and articulatory) to avoid discomfort or other related fears in the presence of healing. However, it is clear that the major impediment to outcomes in those treated for oropharyngeal cancers is likely found in the lack of a measurement standard. In this regard, it is increasingly apparent that objective, as well as subjective information is ultimately necessary to determine individual performance and capacity in this important clinical population. Finally, the desire to address such questions using the traditional group design may in fact be quite inadequate. Group designs may only serve to obscure variability in the data which is in reality the essence of the data we seek to obtain as a valid means of generating evidence on those treated for oropharyngeal cancer. Thus, we have begun to adopt single-subject experimental designs in an effort to provide more specific documentation of individual outcomes secondary to cancer treatment. In the months to come, we look forward to providing additional data that expands the application of acoustic measures to this important clinical population.

5. Acknowledgments

The author would like to thank colleagues Drs. Kevin Fung, John Yoo, and Howard Lampe from the Department of Otolaryngology and Reconstructive Surgery University of Western Ontario, Dr. Wayne Matthews from the Department of Otolaryngology at the University of Calgary, and members of the Voice Production and Perception Laboratory and Postlaryngectomy Research Laboratory at the University of Western Ontario.
6. References


Table 1. Means Values (in Hz) for F1, F2, F3, and F1:F2 Proportional Values for Control and Surgical Groups.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Group</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1:F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>heed</td>
<td>Control M</td>
<td>300</td>
<td>2301</td>
<td>3035</td>
<td>.130</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>323</td>
<td>2413</td>
<td>3109</td>
<td>.134</td>
</tr>
<tr>
<td>hid</td>
<td>Control M</td>
<td>440</td>
<td>1910</td>
<td>2610</td>
<td>.230</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>470</td>
<td>1960</td>
<td>2680</td>
<td>.240</td>
</tr>
<tr>
<td>head</td>
<td>Control M</td>
<td>539</td>
<td>1895</td>
<td>2601</td>
<td>.284</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>545</td>
<td>1986</td>
<td>2678</td>
<td>.274</td>
</tr>
<tr>
<td>had</td>
<td>Control M</td>
<td>674</td>
<td>1735</td>
<td>2492</td>
<td>.388</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>691</td>
<td>1760</td>
<td>2489</td>
<td>.392</td>
</tr>
<tr>
<td>hod</td>
<td>Control M</td>
<td>693</td>
<td>1131</td>
<td>2451</td>
<td>.612</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>705</td>
<td>1201</td>
<td>2476</td>
<td>.587</td>
</tr>
<tr>
<td>hawed</td>
<td>Control M</td>
<td>659</td>
<td>1092</td>
<td>2412</td>
<td>.603</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>669</td>
<td>1169</td>
<td>2416</td>
<td>.572</td>
</tr>
<tr>
<td>hood</td>
<td>Control M</td>
<td>465</td>
<td>1301</td>
<td>2353</td>
<td>.357</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>478</td>
<td>1353</td>
<td>2397</td>
<td>.353</td>
</tr>
<tr>
<td>who’d</td>
<td>Control M</td>
<td>329</td>
<td>1007</td>
<td>2301</td>
<td>.326</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>366</td>
<td>1183</td>
<td>2228</td>
<td>.309</td>
</tr>
<tr>
<td>hud</td>
<td>Control M</td>
<td>610</td>
<td>1427</td>
<td>2426</td>
<td>.427</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>626</td>
<td>1486</td>
<td>2511</td>
<td>.421</td>
</tr>
<tr>
<td>heard</td>
<td>Control M</td>
<td>424</td>
<td>1407</td>
<td>1895</td>
<td>.301</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>461</td>
<td>1458</td>
<td>1937</td>
<td>.316</td>
</tr>
</tbody>
</table>
Figure 1. Radial forearm free flap (RFFF) of oral cavity.
Photograph compliments of K. Fung, M.D., Department of Otolaryngology, Head and Neck Oncology and Reconstructive Surgery, University of Western Ontario, London, ON, Canada.
Tracheoesophageal speech with manual versus automatic stoma occlusion: a multidimensional comparison

Annelies Labaere 1, Jan Vanderwegen 2, Frans Debruyne 1

1 Department of Otorhinolaryngology, Head and Neck Surgery, University Hospitals Leuven, Belgium; 2 Antwerp University Hospital, Department of Otorhinolaryngology, Antwerp, Belgium
Annelies.Labaere@uz.kuleuven.ac.be

Abstract

Objectives: The objective of this study was to make a qualitative comparison between tracheoesophageal speech with manual stoma occlusion and handsfree tracheoesophageal speech with the Provox FreeHands device.

Patients/Materials and Methods: Both manual and handsfree tracheoesophageal speech were compared in 13 patients who were regular users of Provox FreeHands for at least four months. Evaluation of speech material consisted of objective analysis using the KayPentax CSL, perceptual ratings by a group of experienced listeners and patients own perceptual judgment. In addition, data concerning user-friendliness, additional values and inconveniences of the FreeHands device were gathered using questionnaires. The Voice Handicap Index (VHI) was used to evaluate the relation between voice related QOL and method of stoma occlusion. Therefore a control group of TE-speakers, who did not use handsfree speech, was selected.

Results: Objective analysis revealed significant differences (p< 0.05) for parameters fluency (i.e. the number of syllables produced on one intake of breath) and maximal phonation time, to the detriment of handsfree speech. There were no significant differences found for the dynamic characteristics Perceptual judgments of running speech were rated higher in the manual occlusion condition for most patients, but there were interrater differences. Subjective impressions showed that the majority of patients preferred handsfree speech to manual stoma occlusion, noticeably for voice quality, fatigability, attractiveness and feelings of self-confidence. Major inconveniences of handsfree speech were a significantly decreased duration of sticker adhesion, the occurrence of disturbing noises and the need for continued effort. The VHI scores were significantly higher in the group of handsfree speakers compared to the control group.
Conclusions: This clinical study showed that in selected patients the Provox FreeHands shows important subjective benefits. Improvements however are necessary to make the device a useful rehabilitation device for a larger group of laryngectomized patients.

Keywords: tracheoesophageal speech, handsfree speech, automatic tracheostoma valve

1. Introduction

Loss of voice is considered as being the major handicap after a total laryngectomy. Since the first laryngectomies in 1866 and 1873 there has been an evolution in developing techniques for speech rehabilitation. Although in general survival chances have not changed a lot in the last 25 years, a growing attention for speech rehabilitation after surgery has lead to an important improvement of quality of life (Ackerstaff & Hilgers, 1996).

Tracheoesophageal speech is considered as being the most effective way of alaryngeal speech. Advantages are the shortened revalidation time, the use of the normal respiratory system and the better quality of speech (Robbins et al., 1984; Fujimoto et al., 1991.

A frequently reported disadvantage of this way of voice production is the necessity to use one hand to close the tracheostoma. This impedes doing bimanual tasks and attracts the attention to the impairment. (Blom et al., 1982; Singh, 1987; Verkerke et al., 2002).

To overcome this major disadvantage Blom et al. introduced handsfree tracheoesophageal speech using an automatic stoma valve(Blom et al., 1982). Since then different types of tracheostoma valves have come onto the market. In our study Provox FreeHands HME (ATOS Medical) was used (Hilgers et.al 2003). Provox FreeHands HME (fig 1) is an automatic tracheostoma valve with an integrated filter cassette. In contrast to most other types, the openings for breathing are on the side of the device. By increasing exhalation airflow, a membrane will unroll which closes the side openings and gives the patient the ability to speak. The valve in front makes coughing possible without loosing the device.

Fig. 1 Provox FreeHands with HME-cassette. The coughing valve is opened, the membrane is rolled.
2. Patients/Materials and Methods

To compare manual and automatic stoma occlusion 13 patients who use the automatic stoma valve on a regular base were selected. The test group existed of 12 men and 1 woman. Their age varied between 40 and 78 years (mean 57 years). To compare quality of life a control group was composed. Therefore we contacted the patients who had tried out the FreeHands system, but had not bought it for different reasons. 11 patients granted their cooperation. All patients of both study group and control group completed an identification form, the Voice Handicap Index and a questionnaire in which was asked for user-friendliness of the device and patient satisfaction.

The 13 patients of the test group had to read a standard text at a comfortable loudness and sustain the vowel /a/, first at a comfortable loudness and then as long, as loud and as soft as possible, both with manual and with automatic stoma occlusion. In this way 26 speech samples were collected and registered on DAT-recorder.

Following objective parameters were selected: speech rate (i.e. the number of syllables per minute), fluency (i.e. the number of syllables produced on one intake of breath), maximal phonation time (MPT), minimal and maximal intensity and dynamic range. Fundamental frequency measurements turned out to be impossible in 22 of the 26 samples since the vibration pattern was too deviant. For these dynamic measurements we used the Computerised Speech Lab (CSL Model 47300B, Kay Elemetrics).

The perceptual judgments were performed by 6 blinded experienced listeners who were asked to compare, for person 1 till 13, both speech samples. For the parameters naturalness, intelligibility, tension, disturbing noises, intonation/variation and general voice quality, they had to indicate which of both samples they found the best. The possibility ‘both samples are equal’ was also added.

To examine the influence of handsfree speech on the quality of life of our patients, the VHI’s of the test group and the control group were compared.

For statistical processing of the data we used SPSS 9.0

3. Results

3.1 Objective analysis

There was a significant difference (p< 0.05) between automatic and manual stoma occlusion for the parameters MPT and fluency. In the manual stoma occlusion condition patients could sustain the vowel /a/ on average 4 seconds longer than in the handsfree condition. With manual occlusion they could
produce 11 syllables on one intake of breath, in contrast to 8 with automatic occlusion. For the parameters tempo of speech and the dynamic characteristics no significant differences were found.

### 3.2 Perceptual evaluation

When we focus on the results per patient, we can see that the speech of 3 persons was judged better when they were using FreeHands. For 3 other patients, the listeners reported no difference between both conditions. The handsfree speech of 7 patients was found less good than their TE-speech with manual stoma occlusion, by the majority of the listeners. On average, TE-speech with manual stoma occlusion is judged better than handsfree speech for the parameters naturalness, intelligibility, tension and general quality of speech. Only for the parameter intonation/variation no difference was reported in most of the cases (table 1).

Table 1: Subjective comparison (6 listeners) of handsfree and manual stoma occlusion on 5 parameters (n= 13)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Handsfree speech is better</th>
<th>Manual stoma occlusion is better</th>
<th>No difference between both conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naturalness</td>
<td>16</td>
<td>44</td>
<td>18</td>
</tr>
<tr>
<td>Intelligibility</td>
<td>12</td>
<td>37</td>
<td>29</td>
</tr>
<tr>
<td>Tension</td>
<td>17</td>
<td>39</td>
<td>22</td>
</tr>
<tr>
<td>Intonation/variation</td>
<td>11</td>
<td>27</td>
<td>40</td>
</tr>
<tr>
<td>General quality of speech</td>
<td>18</td>
<td>49</td>
<td>11</td>
</tr>
</tbody>
</table>

In handsfree condition, disturbing noise is reported in almost all cases, whereas in half of the speech samples with manual occlusion no noise is heard by the listeners (table 2).

Table 2. Subjective comparison (6 listeners) of handsfree and manual stoma occlusion on parameter ‘disturbing noises’ (n= 13).

<table>
<thead>
<tr>
<th></th>
<th>No disturbing noises</th>
<th>Slightly disturbing noises</th>
<th>Strongly disturbing noises</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free hands condition</td>
<td>4</td>
<td>39</td>
<td>35</td>
</tr>
<tr>
<td>Manual condition</td>
<td>39</td>
<td>31</td>
<td>8</td>
</tr>
</tbody>
</table>
3.3 Subjective patient judgment

The majority of the patients of our test group preferred handsfree speech, concerning quality of voice, fatigability, attractiveness, and feelings of self confidence (table 3).

Table 3. Comparison of both ways of stoma occlusion by means of 4 parameters. Judgements from 13 patients.

<table>
<thead>
<tr>
<th></th>
<th>Voice quality</th>
<th>Fatigability</th>
<th>Attractiveness</th>
<th>Self confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handsfree better than or similar to manual occlusion</td>
<td>9</td>
<td>8</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Handsfree worse than manual occlusion</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Almost all patients reported that the possibility to use both hands was the main advantage of the system. In addition a minority mentioned that their handicap was less remarkable with the free hands device. As main disadvantage patients reported the decreased adhesion of the stoma stickers when using FreeHands. The difference in duration of adhesion was significant ($p < 0.05$). A second frequently reported disadvantage is the fact that handsfree speaking is far more fatiguing than normal TE speech. Furthermore the presence of disturbing noises during speech was experienced as very adverse. Noises were specified as audible hissing noise during inhalation, whistling during speech (caused by the adhesives coming loose) and an audible flop when the membrane closes. The majority of the control group mentioned the higher cost of the device and the fact that handsfree speech is more fatiguing as the main reasons to decide not to use FreeHands any more. Other motivations were the decreased adhesion, disturbing noises and decreased speaking volume.

3.3 Quality of life

There was a significant difference between the FreeHands users and the control group for the VHI total score and the F-score (i.e. the functional component). The FreeHands group has higher scores, what corresponds with a lower quality of life.

4. Discussion and conclusions

The objective measurements showed worse results for handsfree speech for the parameters MPT and fluency. These findings can be explained by the theory that the air consumption is larger when using the automatic valve. There are different causes for this increased air consumption. Firstly, handsfree speaking requires a higher expiration force than manual stoma occlusion, both to unroll the
membrane and to maintain this situation while speaking. Another reason is that there is a certain delay in complete closure of the valve: during the unrolling of the membrane, a small amount of the expiration air can escape by the side openings. Figures 2a and 2b show these findings on the spectrograms of a sustained /a/, the first with manual occlusion, and the second with automatic valve occlusion. On figure 2b we can see a burst of multifrequent energy (white arrow), followed by a smaller low frequency component (black arrow). Then a short break occurs, followed by the spectrum of the vowel /a/. These spectral characteristics correspond with the moment on which the membrane unrolls (white arrow) and the valve closes (black arrow), and do not occur when the stoma is closed manually. They are found to a smaller or larger degree in the handsfree speech of all patients.

This phenomenon does not only cause a larger air consumption during speech, but also leads to the presence of disturbing noise. In perceptual evaluation the occurrence of disturbing noises is reported in handsfree speech in almost all patients, whereas this is much less the case in manual TE-speech. Other authors confirm this finding (Williams et al., 1990; Fujimoto et al., 1991). By both the Free Hands users and the control group, the occurrence of noise, hissing sounds, flops and whistling was mentioned as a disadvantage of the system.

An important issue is the adhesion of the FreeHands device. Looking at individual findings in the test group, we see a group of patients in which the duration of adhesion does not differ according to the way of stoma occlusion. Nevertheless, in the other group the difference in duration is very large, from several days when using manual closure to some hours when using FreeHands.

These patients can only use the automatic valve occasionally during a short period and therefore mention the limited duration of adhesion as an important disadvantage of the system.
Previous studies also reported the maintenance of an airtight seal between stoma valve and peristomal skin as the main problem of handsfree TE-speech (Zanoff et al., 1996; van den Hoogen et al., 1996; Lewin et al., 1999, Op de Coul et al., 2005).

Despite poorer results for handsfree speech, both in objective and perceptual evaluation, and the reported disadvantages by FreeHands users themselves, patients prefer the FreeHands condition. Similar results are found by Op de Coul et al. (2005).

We can conclude that in selected patients the Provox FreeHands shows important subjective benefits. Improvements however are desirable to make the device a useful rehabilitation device for a larger group of laryngectomized patients.

5. References


Williams SE, Scanio TS, Ritterman SI. Perceptual characteristics of tracheoesophageal voice produced using four prosthetic/occlusion combinations. Laryngoscope. 1990;100(3):290

Early Rehabilitation of Head and Neck Cancer Patients at Karolinska University Hospital

Therese Engström 1, Alexander Ahlberg 2, Polymnia Nikolaidis 3, Karin Lundgren Gunnarsson 4, Lena Sharp 4, Göran Laurell 5

1 Dept. of Speech Pathology Karolinska University Hospital, Stockholm, Sweden; 2 Dept. of ENT Karolinska University Hospital, Stockholm, Sweden; 3 Dept. of Physiotherapy, Karolinska University Hospital, Stockholm, Sweden; 4 Dept. Of Oncology, Karolinska University Hospital, Stockholm, Sweden; 5 Dept. of Clinical Sciences, Umeå University, Umeå, Sweden

therese.engstrom@karolinska.se, Alexander.ahlberg@karolinska.se
polymnia.nikolaidis@karolinska.se, karin.lundgren-gunnarsson@karolinska.se
lena.sharp@karolinska.se, goran.laurell@ent.umu.se

Abstract

Objectives: To describe a clinical pathway for early rehabilitation of head and neck cancer patients.

Method: A clinical pathway has been developed at Karolinska University Hospital to improve rehabilitation for patients with head and neck cancer receiving radiotherapy (RT) or a combination of RT and surgery. For many years the clinical pathway for these patients has included consultations with dentists, dental hygienists, dieticians, social workers, nurses and physicians before, during and after RT. Since three years, the patients are also seen by a speech language pathologist and physiotherapist for assessment and information before and after RT. The aim is to start the rehabilitation process early, before the initiation of RT and with that to reduce the risk of complications such as dysphagia, trismus and voice and speech impairment. Emphasis is put on giving patients information about treatment side-effects, improving patient involvement and self-care. The patients receive a physical training program with exercises aiming to maintain tongue and laryngeal motility, jaw opening, swallowing and shoulder function. These exercises should be performed on a daily basis during and after the RT-period. Altogether 257 patients have been included in the intervention group. In an on-going study data is collected from medical hospital files, individual assessments and patient questionnaires (EORTC QLQ-C 30, EORTC QLQ-H&N 35, HADS and a study specific questionnaire) at baseline and at six-month follow-up. Data from the intervention group will be compared with data from a control group.

Keywords: rehabilitation
Introduction

Approximately 300 patients are treated for head and neck (H&N) cancer in Stockholm every year. Many patients suffer from side effects after treatment such as dysphagia, voice-and speech impairment as well as trismus and reduced head/shoulder motility. For many years the clinical pathway for these patients have included consultations with dentists, dental hygienists, dieticians, social workers, nurses and physicians before, during and after RT. The patients were seldom referred to a speech language pathologist or physiotherapist. And if a referral was made, it was always initiated by persistent treatment induced side effects and in general a long time after end of treatment. Earlier studies have shown inadequate rehabilitation for this patient group. During 2004-2007 the Swedish Cancer Society supported a project aiming to develop a clinical pathway to improve rehabilitation for head and neck cancer patients. The goal has been to start the rehabilitation process early, at time of diagnosis, and if possible prevent some of the long-term side effects and to increase patient’s involvement.

Patients/Materials and Methods

2.1 Patients

257 consecutive patients with H&N cancer who received radiotherapy or combination treatment (RT+ surgery) with curative intent between January 2004-June 2007 at Department of Oncology (unit for radiotherapy at South Hospital), Karolinska University Hospital were included in the program.

2.2 Methods

The patients were invited by the coordination nurse for participation in a rehabilitation program including regular visits to a speech language pathologist (SLP) and physiotherapist (PT) for assessment, information about side effects and treatment if needed.

All patients were seen by a SLP before the start of RT and three months after completed treatment. At each visit the patients chewing and swallowing functions of four consistencies were examined by a clinical examination. The tongue motility was examined with oralmotor exercises. The voice and speech functions were assessed perceptually by the SLP. All patients were given prophylactic exercises to maintain motility in tongue and muscles involved in swallowing. The patients were informed to perform tongue motility exercises and a laryngeal motility exercise (Mendelson maneuver) daily during ongoing radiotherapy and preferably two-three months after end of treatment.

The patients also met with a PT before RT start and instructions for exercises were given for preventing trismus and head and neck motility. Follow ups were done 2, 6 and 12 months after RT. The head and neck mobility and interincisial distance was measured before RT and at follow-up. The patients were informed
to perform prophylactic trismus-exercises daily with a so called Engströms-klämma, a large wooden cloth pin that the patient put between their teeth and then pressed the jaws apart according to a certain schedule.

The patients also answered a study specific questionnaire about their swallowing and chewing capacity, speech, voice, dryness of mouth and sensation of taste before and three months after RT. In addition EORTC QLQ-C 30, EORTC QLQ-H&N35 and HADS were completed by the patients at baseline and a six-month follow-up. Data from the intervention group will be compared with data from a control group.

**Results**

This is an ongoing study but the preliminary results show that jaw motility is reduced in almost all H&N patients after RT. Head and shoulder motility is reduced in patients receiving combination therapy (surgery and RT) but not in patients with single modality treatment.

**Discussion and Conclusions**

Our experience from this four year long project is that multi-professional collaboration is educational and essential in high quality cancer care and that early intervention of SLP and PT should be a part of the cancer treatment for H&N patients.

Acknowledgements
The authors like to acknowledge The Swedish Cancer Society for financial support.
Sponsors

Amsterdam Center of language and Communication, University of Amsterdam
Atos Medical AB, Sweden
Atos Medical BV, The Netherlands
Atos Medical GMBH, Germany
Institute of Phonetic Sciences, University of Amsterdam
The Netherlands Cancer Institute, Amsterdam