SUMMARIES OF PH.D. THESES DEFENDED IN 2000-2001

TRACHEOESOPHAGEAL SPEECH. A MULTIDIMENSIONAL ASSESSMENT OF VOICE QUALITY

author: Corina J. van As
promotor: Louis C.W. Pols
co-promotores: Frans J.M. Hilgers and Florien J. Koopmans-van Beinum
date of defence: September 26th, 2001

Summary

Annually, approximately 250 total laryngectomies are carried out in the Netherlands. During this surgical procedure, the entire larynx is removed, mostly because of laryngeal or hypopharyngeal cancer. With this removal of the larynx, also the vocal cords are removed, and consequently the natural voice is lost. During a total laryngectomy, caudally, the larynx is separated from the trachea and the remaining tracheal stump is sutured to the skin of the lower neck. This tracheal opening into the neck is called the tracheostoma or plain 'stoma'. Posterior, the larynx is detached from the pharynx. The remaining pharyngeal mucosa and musculature is sutured together in order to re-establish the digestive tract (see also Figures 2.2 and 2.3).

In the majority of the laryngectomized patients, voice rehabilitation is nowadays achieved by means of a voice prosthesis. This device is a one-way valve, placed into a fistula between the trachea and the esophagus, allowing air to pass from the trachea into the esophagus and preventing fluids/food from entering the trachea and lungs. The air insufflated into the esophagus causes vibrations of the pharyngoesophageal mucosa and the resulting sound enables tracheoesophageal speech production.

The anatomy and morphology of the new voice source (neoglottis) are highly variable: the neoglottis can be situated at different levels in the pharyngoesophageal segment (mostly at the level of the cricopharyngeus muscle and/or the middle and inferior constrictor pharyngeus muscles) and has no uniform anatomical size, shape, or location. The neoglottis also does not have the flexibility and volitional controllability of the glottis. Obviously, the neoglottis plays an important role in tracheoesophageal voice production and its characteristics must influence the quality of the voice. However, at present the knowledge about the relations between the neoglottic characteristics and tracheoesophageal voice quality is still limited.
In this thesis, tracheoesophageal voice quality is investigated both perceptually (subjective) and acoustically (objective), and the anatomical and morphologic characteristics of the neoglottis are investigated with videofluoroscopy and digital high-speed imaging. With this multidimensional study insight should be gained into the anatomical and morphologic characteristics of the neoglottis in relation to tracheoesophageal voice quality.

In chapter 1, a short introduction is given about total laryngectomy. The anatomical situation after total laryngectomy and its implications are described briefly. Then, the aims of the thesis and specific research questions are presented, followed by a description of the outline of this thesis.

In chapter 2, the different aspects of total laryngectomy are described in detail. First, classification, demographics, etiology, symptoms and treatment options of laryngeal cancer are described. Then, a historical overview about total laryngectomy is given, the surgical procedure itself is described, and additional surgical procedures, such as pharyngeal reconstruction and radical neck dissection are clarified. Furthermore, the three most widely used methods of voice rehabilitation after total laryngectomy (tracheoesophageal speech, esophageal speech, and the use of an electrolarynx) are explained. Then, the localization and characteristics of the new sound source, the surgical techniques used to influence this sound source, and the mechanism of tracheoesophageal voice production are described. The chapter ends with a section about other physical and psychosocial consequences of total laryngectomy.

In chapter 3, the patients participating in this study are described in detail. In a table, all sociodemographic and clinical patient characteristics are given. Furthermore, for each part of our study an overview of the clinical and sociodemographic factors is shown, since not all patients were able to participate in each part of the study. Additionally, the subgroups used for investigation of the sociodemographic and clinical factors are specified.

Chapter 4 reports about the results of the perceptual evaluations. These evaluations were carried out by 20 naive and 4 trained expert listeners (speech pathologists) on read-aloud text. For the naive listeners, 19 perceptual scales, and for the trained listeners 20 perceptual scales were used for the perceptual evaluations. In addition to the perceptual scale judgments, the trained expert listeners also gave an overall judgment of the voice quality (good, reasonable or poor). On the basis of this overall voice quality judgment, the voice quality was considered to be good in 14 patients, reasonable in 14, and poor in 12 patients.

The results of the perceptual evaluations of the naive listeners showed that they judged the voices reliably and that they perceived tracheoesophageal speech as very deviant. They did not differentiate between the different perceptual scales regarding voice quality. Statistical analysis revealed only two perceptual dimensions for the judgments of the naive raters: pitch and voice quality.

The results of the perceptual evaluations of the trained expert listeners showed that they also judged tracheoesophageal speech reliably, but that they rated the perceptual scales more positively than the naive raters did. Statistical analysis revealed four perceptual dimensions for the trained raters: pitch, voice quality, tonicity and tempo.

Based on the factor analyses in this study, a reduced number of scales for perceptual evaluation of tracheoesophageal speech is proposed. For the naive listeners 4 scales are sufficient and for expert listeners 8 scales are sufficient as a minimal basic
subset that covers the underlying dimensions sufficiently. This reduction makes perceptual evaluations easier applicable in clinical practice.

Although perceptual judgments are considered to be the ‘gold standard’ for evaluation of voice quality, despite the possible simplifications proposed, it remains a subjective and time-consuming method of evaluation. Objective acoustic analyses, providing quick and reliable measures of voice quality, are therefore an interesting point of investigation in relation to the perceptual evaluations of voice quality.

In chapter 5 the results of acoustic analyses are described. First of all two packages for acoustic analyses were compared: MDVP (Multi Dimensional Voice Program, Kay Elemetrics, Lincoln Park, NJ, USA) and Praat (A System for doing Phonetics by Computer, www.praat.org). It was decided to use Praat, since it enabled reliable pitch extraction in a larger number of voice samples.

The acoustic analyses consisted of ‘acoustic signal typing’ into four different signal types based on a narrow band spectrogram, in which type I shows the best harmonic structure and type IV is barely harmonic. Furthermore, seven acoustic parameters were calculated.

Acoustic signal typing resulted in 7 type I signals, 13 type II signals, 11 type III signals and 8 type IV signals. The acoustic parameters based on pitch detection algorithms (fundamental frequency, standard deviation of fundamental frequency, and jitter) could be calculated for 30 out of the 39 voice samples. The remaining 9 voice samples contained no clear periodicity and thereby no fundamental frequency. However, the four acoustic parameters not based on pitch extraction (percentage of voiced, harmonics-to-noise ratio, glottal-to-noise excitation ratio, and band energy difference) could be calculated for the entire patient group.

Since acoustic analyses can only be considered valuable when they show relations with perceptual evaluations, this chapter also includes investigation of the relations between the acoustic analyses and the perceptual evaluations. It was found that the overall judgment of voice quality was significantly related to the acoustic signal typing: type IV signals were never perceived as good, while type I signals were never perceived as poor. Furthermore, moderate to strong correlations were found between the perceptual dimension pitch and the fundamental frequency; the perceptual dimension voice quality and the percentage of voiced, harmonics-to-noise ratio and band energy difference; and between the perceptual dimension tonicity and the glottal-to-noise excitation ratio. These relations suggest that acoustic parameters can be used to replace some of the perceptual evaluations.

In addition to the acoustic measures, also maximum phonation time was calculated and related to the perceptual evaluations. Results showed that the better the overall voice quality was, the longer the maximum phonation time was. Also, maximum phonation time was related to the perceptual scales jerking-fluent and unintelligible-intelligible, showing that a longer maximum phonation time was related to more fluent and better intelligible speech.

As a result of this study, acoustical analysis, complemented with maximum phonation time, has proven to be a valuable adjunct to perceptual evaluation, with the program Praat as the better choice for experienced investigators.

In chapter 6 the results of videofluoroscopy are described. The videofluoroscopy recordings were obtained in lateral view, during voice production of the sustained vowel /a/. The recordings were evaluated using a newly developed evaluation protocol, consisting of visual assessments and quantitative measures. In order to
obtain the quantitative measures, digitized images of a relevant lateral view of the neoglottis at rest and during phonation were used.

Both the visual assessments and the quantitative measures showed a large variability in the anatomical and morphologic characteristics among the patients. Comparison of the quantitative measures at rest and during phonation showed that in patients after standard total laryngectomy, the maximal subneoglottic distance at rest and the prominence of the neoglottic bar at rest became significantly larger during phonation, indicating a dynamic change from rest to phonation. Also, the cervical level of the neoglottis tended to shift up one half vertebrae from rest to phonation.

Relations between the visual assessments and the quantitative measures showed that part of the visual assessments could be replaced by the more objective and consistent quantitative measures.

Overall, it can be stated, that the standardized evaluation form that was developed (using clear dichotomies and anatomical landmarks for the visual assessments of the situation at rest, when swallowing and during phonation separately), facilitates consensus judgment considerably, and that some of the descriptive visual assessments can be replaced by objective quantitative measures of the neoglottis.

Chapter 7 reports about the results of videofluoroscopy, in relation to the results of the perceptual evaluations of the trained expert raters, the results of the acoustic analyses, and the sociodemographic and clinical factors.

A good overall voice quality and a positive judgment on the perceptual dimension voice quality were related to the presence of a neoglottic bar during phonation and to a normotonic or slightly hypotonic neoglottis. Some of the more specific perceptual scales were related to stasis of barium on the neoglottis during phonation and regurgitation of barium during phonation. Also, good voice quality was related to a good neoglottic closure as indicated by some of the quantitative measures of the neoglottis.

The acoustic signal type was better and the percentage of voiced was higher when a neoglottic bar was present. Regurgitation of barium during phonation and hypotonicity of the neoglottis were related to a lower fundamental frequency. Furthermore, the acoustic signal type was better when the neoglottic closure was better according to the quantitative measures of the neoglottis. Also, moderate to strong correlations were found between the quantitative measures and the acoustic measures.

Regarding the sociodemographic and clinical factors, relations were found for the extent of surgery (better neoglottic characteristics in the group after standard total laryngectomy in comparison to the group after partial or full pharyngeal reconstruction), and for neck dissection and age (more favorable neoglottic characteristics in the group that did not undergo radical neck dissection and in the younger age group).

Overall, it can be stated that the developed protocol is useful for the evaluation of anatomical and morphologic characteristics of the neoglottis in relation to voice quality (perceptual evaluation and acoustic analyses) and that the use of quantitative measures is promising towards a more standardized evaluation of videofluoroscopy recordings in tracheoesophageal speech.

Chapter 8 is a methodological and descriptive chapter on the use of digital high-speed imaging for the evaluation of anatomical and morphologic characteristics of the neoglottis as seen from above (“bird’s-eye view”). In normal laryngeal voices the use
of stroboscopy in order to obtain a slow-motion image of vocal fold vibration is rather common in clinical practice. Digital high-speed imaging has the clear advantage over the more widely applied stroboscopy, that it does not require regular vibrations of the voice source in order to obtain reliable dynamic recordings of these vibrations. Since the vibrations of the neoglottis are often irregular, we chose to use digital high-speed imaging with 2000 frames per second, instead of stroboscopy. Digital high-speed imaging is at present still more used as a research tool, but is developing towards a clinically useful instrument rapidly. To the best of our knowledge, the present study was the first one to investigate the usefulness of digital high-speed imaging for investigation of the vibratory behavior of the neoglottis.

The recordings were obtained with a rigid endoscope transorally, while the patient was phonating a sustained vowel /a/. The recordings were evaluated by means of a newly developed assessment form.

The results of the visual assessments showed a wide variability in all anatomical and morphologic characteristics of the neoglottis studied. Various different shapes of the neoglottis and locations of the vibration were observed, the amount of saliva seen was variable and differed among the patients, the closure phase differed, and a mucosal wave was strong in some cases, but weak or absent in others. The vibration of the neoglottis was irregular in 66% of the patients.

Overall, digital high-speed imaging appeared to be a useful tool for obtaining more insight in the anatomical and morphologic characteristics of the neoglottis, not only in those with a regular vibration, but also in those with an irregular vibration. Useful recordings could be obtained for the majority of the patients.

In chapter 9 the results of the visual assessments of high-speed digital imaging, are related to the results of the perceptual evaluations of the trained expert raters, the results of the acoustic analyses, the results of videofluoroscopy, and the sociodemographic and clinical factors.

A good overall voice quality was related to the visibility of the origin of the neoglottis. In relation to the more specific perceptual scales, the visual assessments of the amount of saliva, the origin of the neoglottis, the shape of the neoglottis and the regularity of the vibration appeared to be of importance.

The acoustic signal type was better when the amount of saliva was smaller and when the origin of the neoglottis was visible. In relation to the acoustic measures, the amount of saliva, the shape of the neoglottis, and the regularity of the vibration were found to have some influence.

The presence of a neoglottic bar at rest, regurgitation of barium during phonation, and the tonicity of the neoglottis judged in the videofluoroscopy recordings were found to be related to the amount of saliva judged in digital high-speed recordings. The presence of a neoglottic bar at rest judged in videofluoroscopy recordings was related to the visibility of the origin of the neoglottis and to the shape of the neoglottis judged in digital high-speed imaging. Furthermore, some quantitative measures obtained from the videofluoroscopy recordings were related to the amount of saliva judged in the digital high-speed imaging recordings.

Regarding the sociodemographic and clinical factors, relations were found for the extent of surgery (better neoglottic characteristics in the group after standard total laryngectomy in comparison to the group after partial or full pharyngeal reconstruction), and whether or not a myotomy has been carried out.

This study shows that digital high-speed imaging can provide valuable extra information about the neoglottic characteristics in relation to tracheoesophageal voice
quality. In relation to videofluoroscopy it can be seen partly as additional information and partly as an overlap. Future possibilities for automated objective analyses of the high-speed recordings will increase their usefulness for research and in clinical practice even more.

This thesis ends with chapter 10 in which the results are discussed in relation to the aims presented in chapter 1, a proposal is formulated for a clinical evaluation protocol and the general conclusions are presented. The chapter ends with future perspectives and suggestions for further research into this subject.

The clinical evaluation protocol that is formulated, is based on the studies described in this thesis and consists of a basic subset of perceptual scales, some acoustic parameters, and videofluoroscopy recordings. Depending on the goal of the evaluations, other methods of investigation, such as manometry or the Voice Handicap Index, can be added, or the number of perceptual scales or acoustic parameters studied can be enlarged. Although digital high-speed imaging has proven to give valuable information complementary to videofluoroscopy, it is not included in the basis protocol, since it is not yet widely used as a clinical diagnostic tool and still under development towards that.

In general, it can be concluded that, although tracheoesophageal speech is the best possible and generally well functional communication method presently available, it is still considerably deviant from normal speech. Especially perceptual judgments of naive raters show this deviancy. The perceptual judgments of the trained speech-language pathologists were more positive than those of the naive listeners. This could of course be expected since the speech-language pathologists are used to this type of speech and they have an extensive knowledge about the anatomy and morphology of the new voice source. The judgments of the speech-language pathologists are therefore more relevant in relation to clinical practice and more useful to serve as a standard for the other evaluation methods used in the present study.

Acoustic analyses confirmed the deviant voice quality found by means of the perceptual evaluations. The moderate to strong correlations between the perceptual evaluations and the acoustic measures show that part of the acoustic measures is suitable to replace the more subjective and time-consuming perceptual evaluations. An acoustic signal typing system as proposed in this thesis also showed to be relevant in relation to voice quality.

Furthermore, a number of parameters were found that related voice quality in some way to the anatomy and morphology of the neoglottis. These parameters can serve as a basis for future research and for evaluation of tracheoesophageal speech in clinical practice. The improved knowledge about neoglottic characteristics that are important for voice quality, might, in future, lead to adjustments of surgical techniques at the time of surgery, in order to achieve optimal and predictable voice quality. Furthermore, at a secondary stage it might be possible to develop surgical techniques based on the knowledge obtained in this study.