Tracheoesophageal Speech. A Multidimensional Assessment of Voice Quality
van As, C.J.

Citation for published version (APA):
CHAPTER 1

General introduction
ABSTRACT

The main aims of the studies presented in this thesis are (1) to develop standardized protocols for the investigation of various aspects of tracheoesophageal speech (voice quality as well as clinical aspects), (2) to investigate relationships between the perceptual characteristics of tracheoesophageal voice and the acoustic measures of voice quality, (3) to investigate relationships between the perceptual and acoustic characteristics of tracheoesophageal voice and the anatomical and morphologic characteristics of the neoglottis as studied by means of videofluoroscopy and digital high-speed imaging, and (4) to establish a basic subset of perceptual, acoustic, and imaging parameters needed for future, optimal assessment of tracheoesophageal voice quality.

In this introductory chapter a short background is given to the subject of the thesis, whereas also the aims of the thesis are presented. This chapter concludes with a section on the outline of the thesis.
1.1 INTRODUCTION

In the Netherlands, annually around 700 people are diagnosed with laryngeal cancer (Visser et al., 1998). When diagnosed at an early stage, those tumors can be treated with radiotherapy or laser surgery. However, in larger tumors and in case of recurrence of the disease, a total laryngectomy is often necessary.

In the Netherlands, every year approximately 250 patients are laryngectomized (Ackerstaff et al., 1990). During this surgical procedure the entire larynx is removed. Thereby, the upper airways (nose, pharynx) and the lower airways (lungs, trachea) are disconnected. After removal of the larynx the pharyngeal muscles and mucosa are sutured together, in order to re-establish the digestive tract, and the trachea is bowed forward and sutured in the skin at the base of the neck. This opening in the neck is called the (tracheo)stoma. The patient group reported about in this thesis uses tracheoesophageal speech; for this method of voice restoration a voice prosthesis is inserted into a surgically created fistula between the trachea and the esophagus. This voice prosthesis enables the use of exhaled air from the lungs for voice production. When the tracheostoma is closed, the exhaled pulmonary air is directed through the voice prosthesis into the esophagus. In the esophagus and pharynx the mucosa overlying the pharyngeal constrictor muscles and/or the cricopharyngeus muscle serves as the new voice source (neoglottis). The exhaled air that passes through the voice prosthesis is blown into the esophagus and there sets the neoglottis into vibration. The voice sound produced by the neoglottis is formed into speech by the normal resonators and articulators. The voice prosthesis itself does thus not generate any voice sound, but only allows the air from the lungs to enter the esophagus.

Due to the disconnection between the upper and lower airways the nasal functions during breathing (heating, moisturizing and filtering of the inhaled air) are lost. This is compensated for, by the use of a heat- and moisture exchanger (HME) that covers the tracheostoma (Hilgers et al., 1996). Due to this disconnection the sense of smell is also lost since the inhaled air enters the tracheostoma and thus does not pass the olfactory epithelium in the nose (Van Dam et al., 1999). In Figure 1.1 a schematic drawing of the anatomical situation after total laryngectomy is shown.

One of the most obvious implications of a total laryngectomy is the loss of the natural voice. Over the years, vocal rehabilitation after total laryngectomy has improved considerably. With the development of a useful voice prosthesis two decades ago, the use of tracheoesophageal speech was initiated (Singer and Blom, 1980). Nowadays it has become the most frequently used method of voice restorations after total laryngectomy. Over the years several voice prostheses have been developed. At the Netherlands Cancer Institute the Provox® voice prosthesis has been developed (Hilgers and Schouwenburg, 1990). This voice prosthesis was developed for retrograde insertion (via mouth and throat) into the tracheoesophageal fistula. In 1997 Hilgers et al. developed the Provox®2 voice prosthesis (Hilgers et al., 1997) which was designed for anterograde insertion (via the tracheostoma) into the tracheoesophageal fistula. Nowadays this second-generation voice prosthesis Provox®2 has become a frequently used voice prosthesis. The patients in this study all used a Provox®2 voice prosthesis. This prosthesis acts as a one-way valve, which enables air passing from the lungs into the esophagus, but prevents leakage of food or saliva from the esophagus into the lungs.

Older methods of voice rehabilitation after laryngectomy that are still used nowadays are esophageal speech and speech by use of an electrolarynx. In esophageal speech air from the mouth is trapped into the esophagus and then forced upwards into the mouth again. This method is sometimes compared with belching, although the air needed for speech is only
forced into the esophagus and not into the stomach, as is the case with belching. The upward movement of air causes vibration of the neoglottis and so the voice sound is produced. The maximum phonation time of this type of speech is very short since only small amounts of air (60-80 cc) can be used for voice production. An electrolarynx is a hand-held mechanical device. A vibrating membrane is held against the floor of the mouth. The vibrations of the membrane are transmitted through the tissues and the user modulates these vibrations with normal articulation.

Several studies have shown that tracheoesophageal voice rehabilitation gives the best voice results compared to the other methods (Williams and Barber Watson, 1987; Pindzola and Cain, 1988; Nieboer et al., 1988; Debruyne et al., 1994; Bertino et al., 1996; Max et al., 1996). However, the voice quality of tracheoesophageal speech is still very deviant from normal voices and highly variable amongst speakers (Baggs and Pine, 1983; Robbins, 1984; Robbins et al., 1984a; Robbins et al., 1984b; Van As et al., 1998). Furthermore, apart from the deviant voice quality, also the intelligibility is decreased (Cullinan et al., 1986; Doyle et al., 1988; Hammarberg et al., 1990; Hammarberg et al., 1992; Max et al., 1997; Nord et al., 1992).

The new voice source (neoglottis) plays an important role in the quality of the voice. Several methods of investigation can be used to study the characteristics of the neoglottis and the mechanism of tracheoesophageal speech production. One of the most frequently used methods is videofluoroscopy recording (video recording of X-ray examination) of the
neoglottis (Lindsay et al., 1944; Robe et al., 1956; Vrticka and Svoboda, 1961; Kirchner et al., 1963; Smith et al., 1966; Damsté and Lerman, 1969; Bentzen et al., 1976; Richardson, 1981; Wetmore et al., 1985; Melvor et al., 1990; Sloane et al., 1991; Daou et al., 1992; Isman and O’Brien, 1992). Other methods that are used are fiberoptic studies (Brewer et al., 1975), stroboscopy (Omori et al., 1994), and digital high-speed recordings (Van As et al., 1999), which are used to provide a view from above, electromyographical (EMG) studies, which provide information on muscle activity (Mohri et al., 1994), and aerodynamic studies (Moon and Weinberg, 1987; Deschler et al., 1999), which provide information on pressure, airflow and resistance parameters of the voice production. Anatomical and morphologic characteristics of the neoglottis have been more extensively studied for esophageal speech than for tracheoesophageal speech, most probably because of the longer history of esophageal speech production. However, results obtained from investigations of esophageal speech cannot automatically be applied to tracheoesophageal speech. Comparisons between speech results have already shown large differences between these two voice rehabilitation methods (Nieboer et al., 1988; Robbins et al., 1984a; Debruyne et al., 1994). Especially, the underlying mechanisms of esophageal speech and tracheoesophageal speech are different, i.e. the use of relatively small amounts of air from the mouth as a driving source in esophageal speech versus the use of pulmonary air directed into the esophagus by means of a voice prosthesis in tracheoesophageal speech. Therefore, there is a need for basic investigation of the characteristics of the neoglottis in tracheoesophageal speech and the relation between those characteristics and the voice quality in tracheoesophageal speech. Investigation of those relations could also give insight in the causes of the variability in quality that has been found between the tracheoesophageal voices. Throughout this thesis the term voice quality is used, although one should of course realize that voice quality is a property of speech and can thus not always be seen as an independent aspect. For instance, when listening to voice quality, the dynamic aspects of speech might also influence the impression of it.

In this study, perceptual evaluations of tracheoesophageal voice quality will be performed. Since speech is used for oral communication, the listener plays an important role in this. Relatives, family, and people in the street and in shops are daily communicative partners for the patients and their perceptual judgments are thus relevant in that respect; speech-language pathologists base their clinical judgments and therapy strategies partly on their perception and their judgments are thus relevant in that respect. The purpose of the perceptual evaluations in this thesis is thus twofold. On the one hand, perceptual evaluations of speech-language pathologists treating laryngectomized patients will serve as a standard against which other measures are evaluated. On the other hand, perceptual evaluations by naive listeners give insight in the perception of tracheoesophageal speech in the normal communicative situations of the laryngectomized speakers.

Apart from these perceptual analyses, acoustic parameters (fundamental frequency parameters, pitch perturbation parameters, loudness perturbation parameters, harmonics-to-noise ratio) of voice quality are investigated. These acoustic parameters can be obtained in a more objective manner and could therefore be of clinical importance, since they are more objective and less time-consuming than perceptual evaluations. However, acoustic analysis in tracheoesophageal voices is troublesome due to high irregularity of the voice (Van As et al., 1998). With the acoustic measures traditionally used for normal voices, only the better tracheoesophageal voices can be analyzed. It is, however, more interesting to investigate the entire range of voice qualities. The first purpose will thus be to find acoustic parameters that can be obtained reliably and that make sense for the entire range of voice qualities. Secondly, these parameters will be related to results of the perceptual evaluations of the expert listeners to investigate their relevance for the quality of the voice.
Although EMG and aerodynamic studies could provide interesting information on the tracheoesophageal voice production mechanism, in the present study the choice was made to study the visual aspects of anatomy and morphology of the neoglottis by means of videofluoroscopy recordings and digital high-speed recordings. Videofluoroscopy, on the one hand, is an investigation method that is available in most clinics and is often used for investigation of the neoglottis. Digital high-speed imaging, on the other hand, is a relatively new evaluation method used for research purposes that enables visualization of the vibration of the neoglottis during phonation.

Many videofluoroscopic studies have been performed in esophageal speech, but similar studies regarding tracheoesophageal speech are rare. As already said this is most probably due to the shorter history of tracheoesophageal speech. In this study, videofluoroscopy is used to enable a lateral view of the neoglottis. Also, a new protocol will be developed in this study to evaluate the anatomical characteristics of the neoglottis in a more quantitative manner. The information on the anatomical and morphological characteristics of the neoglottis, as obtained by videofluoroscopy, will be related to the results of the perceptual and acoustic evaluations and to sociodemographic and clinical factors.

The lateral view of the neoglottis, as provided by videofluoroscopy, gives only part of the information needed for investigation of anatomical and morphologic characteristics of the neoglottis. Especially, characteristics on the vibratory behavior and other anatomical and morphologic characteristics, which can only be observed from above, might give additional and valuable information. The standard method used for investigation of laryngeal voices (stroboscopy) cannot be used in the majority of tracheoesophageal voices because of the lack of regular periodicity of the tracheoesophageal voice source. Furthermore, the electrodes for electroglottography or the microphone at the neck used for tracing the fundamental frequency may cause deformation of the neoglottis. Therefore, specialized digital high-speed imaging equipment is used. This is a relatively new method that is currently only available for research purposes. The digital high-speed imaging equipment allows recording of the neoglottis during phonation with 2000 or even more frames per second. This method has not been used until recently for studying the neoglottis after total laryngectomy and is expected to give new insights in the anatomical and morphological characteristics of the neoglottis. First, a protocol for the visual assessment of the characteristics of the neoglottis is developed and then the recordings are assessed on the basis of this protocol. Results of these assessments are related to perceptual and acoustic evaluations of voice quality and relations with sociodemographic and clinical factors are investigated.

With this study insight might be gained in the most favorable characteristics of the neoglottis in tracheoesophageal speech. Although phonosurgery of the neoglottis, in contrast to the vocal cords, has not been performed yet, in future this might become a treatment option for optimizing neoglottic characteristics and thereby improving tracheoesophageal voice quality. In this respect, however, it should be kept in mind that in oncology, radical removal of the tumor is always the first goal of surgery, which implies that sometimes concessions have to be made as to the optimal (re)construction of the neoglottis. However, better insight in the parameters that determine voice quality might lead to more evidence-based choices in the future.

In this study, mainly male patients after standard total laryngectomy are studied. Although most laryngectomized patients are males, the number of laryngectomized females is increasing, due to a change in smoking and drinking habits. Therefore, some female patients were included in the study as well.

With a voice prosthesis, voice production in patients after total laryngectomy combined with pharyngeal and/or esophageal reconstruction is enabled as well. The resulting voice quality after such extensive surgery is reported to be less satisfactory than the results after
standard total laryngectomy with primary closure of the pharynx. In this study, also some laryngectomized speakers with pharyngeal and/or esophageal reconstruction are included.

1.2 AIMS OF THE STUDY

As stated above many questions on tracheoesophageal voice quality are still unanswered. The investigations presented in this thesis try to answer part of them.

The specific aims of the investigations described in this thesis are:

- to investigate the perceptual characteristics of tracheoesophageal speech as perceived by naive listeners (i.e. the population that laryngectomized patients meet during their normal daily communicative situations);
- to investigate the perceptual characteristics of tracheoesophageal speech as perceived by trained expert listeners (i.e. the speech-language pathologist involved in their vocal rehabilitation);
- to investigate the differences between the judgments of naive listeners and those of trained expert listeners;
- to develop an acoustic signal typing system for tracheoesophageal voice quality and to search for acoustic measures that can be reliably applied on the entire range of tracheoesophageal voice qualities;
- to investigate the relations between the results of the perceptual evaluations of the trained experts and the results of acoustic analyses;
- to develop a protocol for standardized and quantitative evaluation of the anatomical and morphologic characteristics of the neoglottis in videofluoroscopy recordings of tracheoesophageal phonation;
- to investigate the relations between the results of the perceptual evaluations of the trained experts, the acoustic parameters, and the anatomical and morphological characteristics of the neoglottis studied in videofluoroscopy recordings by means of the newly developed evaluation protocol;
- to develop a protocol for standardized and quantitative evaluation of the anatomical and morphologic characteristics of the neoglottis in digital high-speed imaging of tracheoesophageal phonation;
- to investigate the relations between the results of the perceptual evaluations of the trained experts, the acoustic parameters, and the anatomical and morphologic characteristics of the neoglottis as studied in digital high-speed imaging of phonation by means of the newly developed evaluation protocol;
- to investigate the relation between the anatomical and morphologic characteristics of the neoglottis studied in videofluoroscopy and those studied in digital high-speed imaging;
- to investigate what clinical and sociodemographic factors influence tracheoesophageal voice quality;
- to investigate what clinical and sociodemographic factors are related to anatomical and morphologic characteristics of the neoglottis;
- to establish a basic subset of the perceptual, acoustic, and imaging parameters that are used in this study for future optimal assessment of tracheoesophageal voice quality.

1.3 OUTLINE OF THIS THESIS

The perceptual evaluations of tracheoesophageal voice quality that are performed by the trained expert raters serve as a basis against which the acoustic measures of tracheoesophageal voice quality and the anatomical and morphologic characteristics of the neoglottis as studied by videofluoroscopy and digital high-speed imaging will be evaluated.
The thesis will start with a chapter on total laryngectomy, Chapter 2. That chapter is meant to give the reader who is inexperienced regarding this topic, as well as the experienced reader, insight in the various aspects of total laryngectomy: laryngeal cancer, demographics, etiology, surgical approaches, types of voice rehabilitation and physical and psychosocial consequences are described. Chapter 3 describes the group of patients participating in the study. Chapter 4 reports about the perceptual evaluations. The investigations of the perceptual scale judgments, performed by the trained expert listeners and the naive listeners, and the overall judgment of the voice quality performed by the trained expert listeners are reported. Factor analyses show the underlying perceptual dimensions, and discriminant analyses show the discriminating perceptual scales between good, reasonable and poor voices. Also relations with sociodemographic and clinical factors are described. The chapter concludes with a recommendation on the perceptual scales that can be used for perceptual evaluation of tracheoesophageal voices. Chapter 5 describes the results of acoustic analysis (fundamental frequency parameters, periodicity analyses, harmonicity analyses, and acoustic signal typing) of the voice quality and the relation between the acoustic parameters and the perceptual scales. Also relations with the sociodemographic and clinical factors are described. Chapter 6 reports about the results of visual assessment and quantitative measures of the anatomy and morphology of the neoglottis in videofluoroscopy recordings, and about the relation between these results and the overall judgment of voice quality. The relations between the results of videofluoroscopy and the sociodemographic and clinical factors are described as well. In Chapter 7 the results of videofluoroscopy, as described in Chapter 6, are related to the results of the perceptual and acoustic evaluation of voice quality as described in Chapter 4 and Chapter 5. Chapter 8 describes the anatomical and morphologic characteristics of the neoglottis that are found by means of digital high-speed imaging. In Chapter 9 the results of digital high-speed imaging, as described in Chapter 8, are related to the results of the perceptual and acoustic evaluation of voice quality described in Chapter 4 and Chapter 5 and to the results of videofluoroscopy described in Chapter 7. Chapter 10 contains a general discussion and the conclusions.

The thesis is concluded with summaries in English and in Dutch.

**REFERENCES**


CHAPTER 1


