CHAPTER 10

General discussion and conclusions
ABSTRACT

The four main aims of the present study, as presented in Chapter 1 were (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, (4) to establish a basic subset of perceptual, acoustic, and imaging parameters for future, optimal assessment of tracheoesophageal voice quality. The specific research questions corresponding to these aims are presented in section 1.2.

In this final chapter the results of the different parts of this study (perceptual evaluations, acoustic analyses, videofluoroscopy, and digital high-speed imaging) are summarized and discussed in relation to the aims presented in the first chapter of this thesis. Then, a proposal for a clinical investigation protocol, based on the results of the different parts of the study and some general conclusions are presented. The chapter ends with future perspectives and gives suggestions for further research on this subject.
10.1 INTRODUCTION

Voice rehabilitation after total laryngectomy is an issue that has been subject of concern already since the first total laryngectomy that was performed for laryngeal cancer by the Viennese surgeon Billroth in 1873 (Gussenbauer, 1874). Since the introduction of voice prostheses two decades ago, tracheoesophageal voice rehabilitation has become a well-established method of voice restoration after total laryngectomy (Hilgers et al., 1999). Tracheoesophageal speech shows a large number of differences in comparison to normal speech. Regarding voice production, the most obvious difference lies of course in the different sound source, i.e. the neoglottis, mostly formed by the cricopharyngeus muscle and/or the inferior and middle constrictor pharyngeus muscles together with the overlying mucosa, instead of the glottis, being the vocal folds in the larynx. The change of the voice sound is of course the most striking difference between tracheoesophageal and normal speech, but also intelligibility has been found to decrease and disturbing noises can be heard during speaking when stoma-occlusion is not completely airtight or incorrectly timed. The focus of the investigations in the present study was on the quality of tracheoesophageal voice in relation to the anatomical and morphologic characteristics of the neoglottis. First of all, it is of utmost importance that the patient receives proper voice training, in order for the voice prosthesis to function properly and for the speaker to acquire a good speaking technique with optimal airtight closure of the tracheostoma, by which the best results for voicing can be achieved. The most important factor in tracheoesophageal voice quality is however the functioning of the neoglottis. In the present study voice quality has been investigated perceptually and acoustically, and, additionally, anatomical and morphologic characteristics have been investigated by means of videofluoroscopy and digital high-speed imaging.

Although the patient sample in the present study was not primarily composed to study the influence of specific sociodemographic and clinical factors (except for sex and extent of surgery) on tracheoesophageal voice quality and anatomical and morphologic characteristics of the neoglottis, some relations nevertheless could be studied and are reported about.

In this chapter the results of all investigations are summarized and discussed and a proposal is made for an investigation protocol of tracheoesophageal speech in future clinical research/practice.

10.2 TRACHEOESOPHAGEAL VOICE QUALITY

In the present study tracheoesophageal quality was investigated by means of perceptual evaluations performed by naive and trained listeners (section 10.2.1), acoustic analyses (acoustic signal typing and acoustic measures) and maximum phonation time (section 10.2.2), videofluoroscopy (visual assessments and quantitative measures) (section 10.2.3), and digital high-speed imaging (visual assessments) (section 10.2.4). For each of these parts of the study specific aims were formulated and below the separate parts of the study will be described and discussed in relation to those aims.

10.2.1 PERCEPTUAL EVALUATIONS

Naive listeners have performed perceptual evaluations in order to gain insight in the perception of tracheoesophageal speech in daily communication, whereas trained expert listeners (speech-language pathologists) have performed perceptual evaluations to gain insight in the clinical point of view that could serve as a standard against which other clinical evaluation methods (acoustic analyses, videofluoroscopy and digital high-speed imaging) could be evaluated.
10.2.1.1 Naive raters

The perceptual evaluations of voice quality by the naive raters have shown that tracheoesophageal speech is still very deviant from normal speech, something that must have implications for the psychosocial functioning of these patients in their daily life. Although psychosocial functioning in relation to the quality of the voice was not a major issue in this thesis, it should be part of the clinical investigation protocol that we propose later in this chapter. The use of (an adapted version of) the Voice Handicap Index (Jacobsen et al., 1997) could probably be suggested for this. Future investigations are needed to study the usefulness of the VHI in laryngectomized patients.

On the basis of our principal component analyses on the perceptual evaluations of the 19 bipolar semantic 7-point scales by the naive raters, resulting in two perceptual dimensions (voice quality and pitch), a basic subset of four perceptual scales could be derived that is sufficient for perceptual evaluation of tracheoesophageal speech by naive raters. This basic subset consists of the scales ugly-beautiful, deviant-normal, low-high, and deep-shrill and covers the two perceptual dimensions that the naive listeners in the present study used to judge tracheoesophageal speech. Depending on the aim of the evaluation other perceptual scales can be added to this basic set. The use of judgments of naive listeners depends of course on the purpose of the study; they do not necessarily have to be included in the standard evaluation protocol for tracheoesophageal speech.

10.2.1.2 Trained expert raters

The perceptual evaluations that were performed by the trained expert listeners (speech-language pathologists who were trained for carrying out these perceptual evaluations) are more relevant with respect to clinical practice. The use of an overall judgment of voice quality as good, reasonable or poor already gives a rather good indication of the quality of the voice, but in addition to that more specific perceptual scales are needed to describe the different aspects of tracheoesophageal voice quality. Principal component analyses on the perceptual evaluations of the trained expert raters on the 17 out of the 20 bipolar semantic 7-point scales that were reliably judged, resulted in four perceptual dimensions (voice quality, tonicity, pitch and tempo). Based on those four perceptual dimensions, a basic subset of 8 perceptual scales could be composed that is sufficient for perceptual evaluation of tracheoesophageal speech by expert listeners (see below). The expert raters in the present study were trained before carrying out the perceptual evaluations. Regular training sessions could improve perceptual evaluations by different clinicians in order for the evaluations to become more consistent and reliable and for the different clinicians to judge the voice quality in a more comparable fashion.

An overall judgment of tracheoesophageal voice quality as good, reasonable or poor and a reduced subset of eight perceptual scales covering the four perceptual dimensions, is proposed for evaluation of tracheoesophageal speech by expert raters in clinical practice. This subset consists of the following scales: deviant-normal, ugly-beautiful, breathy-not breathy, hypotonic-not hypotonic, low-high, deep-shrill, slow-quick, and dragging-brisk. In this respect, it should however be kept in mind, that depending on the goal of the perceptual evaluations, also other scales that are reliably judged and are specifically valid for evaluation of tracheoesophageal speech, such as bubbly-not bubbly or hypertonic-hypotonic, can be useful to judge in addition to the subset with respect to certain anatomical or morphologic characteristics of the neoglottis.
10.2.1.3 Naive raters versus trained expert raters

Results showed that the trained expert raters judged the tracheoesophageal voices more positive than the naive raters and that they used the range of the perceptual scales better. For the naive raters only two perceptual dimensions could be extracted from the ratings whereas for the trained raters four perceptual dimensions were extracted. This indicates that the trained expert listeners differentiate more between various perceptual aspects of tracheoesophageal voice quality. With respect to neoglottic characteristics and clinical practice, the judgments of the trained expert listeners are clinically more relevant than the judgments of the naive listeners. The judgments of the trained raters are more useful with respect to the specific neoglottic characteristics leading to specific perceptual aspects of tracheoesophageal voice quality, since they differentiate more between the various perceptual aspects. Insight in the relations between neoglottic characteristics and anatomical and morphologic characteristics of the neoglottis is relevant with respect to voice and speech training after total laryngectomy as well as with respect to the effects of different surgical techniques. Also, it is clinically more relevant to use the judgments of the speech-language pathologist, who is giving the voice training, since the speech-pathologist is monitoring the results of the training and is trying, together with the patient, to achieve an optimal voice quality.

10.2.2 Acoustic analyses and maximum phonation time

Although perceptual evaluation is seen as the ‘gold standard’ for judgment of voice quality, its drawbacks (subjective, time-consuming) should not be underestimated. Reducing the set of scales and the use of an overall judgment of voice quality (good, reasonable, poor) is already a step towards less time-consuming evaluations. Nevertheless, it would be of advantage if specific acoustic analyses could be used to evaluate voice quality, since they would provide more objective and consistent results and be less time-consuming. In the present study, first of all an acoustic signal typing system has been developed for tracheoesophageal voice and, furthermore, a set of 7 acoustic measures were evaluated.

When recording the sustained /a/ for acoustic analysis, the investigator should always try to obtain a vowel that is representative for the voice quality of that patient. During the speech recordings for the present thesis, at least three vowels were recorded. A number of patients needed more than three attempts and proper instructions of the investigator (for instance to avoid a too loud onset or to keep the pitch as stable as possible) to achieve a representative vowel for further acoustic analyses.

In addition to the objective acoustic measures of voice quality also an objective measure of vocal function was obtained: maximum phonation time. The value of these objective analysis methods is studied in relation to the perceptual evaluations of tracheoesophageal voice quality by the trained expert raters.

10.2.2.1 Acoustic signal typing

A narrow-band spectrogram of the sustained /a/ voice sample shows the presence or absence of the harmonic structure of the voice sound very well and gives a good visual representation of the acoustic characteristics of the voice. Although the creation of a narrow-band spectrogram is an easy procedure, the derived acoustic signal typing (I-IV), despite the clear definitions formulated, remains somewhat subjective. Nevertheless, we suggest including acoustic signal typing in the clinical investigation protocol, since it is (visually) very informative with respect to voice quality. Furthermore, information obtained from the acoustic signal typing can be of help in the evaluation of the acoustic measures. For instance, in type-IV signals, showing a noise-like spectrogram that is barely harmonic, calculation of
parameters based on pitch extraction (fundamental frequency, standard deviation of fundamental frequency, and jitter) cannot be carried out reliably, or even not at all.

10.2.2.2 Acoustic measures

In the present study, seven acoustic measures were calculated. Four out of these seven measures could be calculated reliably for the entire patient group (percentage of voiced, harmonics-to-noise ratio, glottal-to-noise ratio, and band energy difference). The other three measures are based on pitch extraction (median fundamental frequency, standard deviation of fundamental frequency, and jitter), and these could be reliably calculated for a majority (77%) of the voice samples. The variability of the acoustic measures among the patient sample was found to be large, showing that large differences exist in tracheoesophageal voice quality between patients.

An experienced investigator should carry out the acoustic analyses, since especially with these deviant, irregular voices pitch extraction errors may occur that can make the outcome unreliable. When performed by an experienced investigator the software program Praat, developed in the Institute of Phonetic Sciences of the University of Amsterdam, can be used best, since it allows adjustment of analysis settings to optimize the result of pitch extraction.

10.2.2.3 Maximum phonation time

Maximum phonation time is an objective measure of vocal function that can be obtained easily and reliably for all patients. Each patient was given at least three attempts and the best one was selected for further analysis. The maximum phonation times ranged from 3 s to 37 s with a mean of 13 s, while in normal speakers the average maximum phonation time is 26 s. Thus, although a number of patients reaches normal values for maximum phonation time, also extremely short phonation times were found.

10.2.2.4 Perceptual evaluations versus acoustic analyses

The correlations between the results of the acoustic analyses and the perceptual evaluations were moderate to strong, showing their relevance with respect to tracheoesophageal voice quality. However, not for all perceptual scales relevant acoustic measures with sufficient correlation can be found as a substitute, and also, not all acoustic parameters are related to tracheoesophageal voice quality, e.g. the acoustic parameter jitter did not show any significant relation with the perceptual scales.

When investigating those relations it should be kept in mind that the acoustic analyses are performed on the sustained vowel /a/ and that the perceptual evaluations are performed on read-aloud text. Acoustic analyses of voice quality in running speech could probably improve the relations between perceptual evaluations and acoustic analyses, since in this case they are performed on the same, more natural, type of speech material. However, acoustic analyses of sustained /a/ samples are already complicated enough for many of these irregular voices and more detailed studies are needed on the possibility of using running speech for acoustic analyses in these voices in a meaningful and reliable way. In this respect, one could think of performing the acoustic signal typing on one fixed and always equal part of running speech instead of a sustained vowel.

Maximum phonation time appears to be an important parameter. Relations between the perceptual evaluations and maximum phonation time showed that it was longer when the overall voice quality was better. Short maximum phonation times were related to less fluent and less intelligible tracheoesophageal speech.
10.2.3 Videofluoroscopy

Videofluoroscopy is a clinical evaluation method that is used in most clinics on a regular basis for diagnosis of problems with voice production or swallowing after total laryngectomy. First, we developed a standardized evaluation protocol for the evaluation of the videofluoroscopy recordings. Then, the results were related to the perceptual evaluations and acoustic analyses (including maximum phonation time) of tracheoesophageal voice quality in order to gain insight in the relation between the neoglottic characteristics and voice quality.

10.2.3.1 Evaluation protocol videofluoroscopy

An evaluation protocol has been developed that consists of structured visual assessments and quantitative measures of neoglottic characteristics at rest and during phonation. One important difference with assessments of videofluoroscopy, as found in the literature, was the judgment of tonicity. The tonicity is by most authors judged on the basis of the behavior of the neoglottis at rest, when swallowing, and during voicing together, whereas also voice quality is included in the judgment. The goal of this part of the present study, being the investigation of the relations between videofluoroscopy and voice quality, first of all led to the decision not to include voice quality in the judgment of tonicity of the neoglottis. Then, it appeared to be too difficult to reach consensus of opinion about the tonicity of the neoglottis, when judging the three conditions (rest, swallowing, phonation) together. Therefore, the neoglottis was judged for these three conditions separately. Using clear dichotomies (yes/no) and clear anatomical landmarks (cervical vertebrae), only one subjective parameter remained (tonicity). However, one of the remaining problems is that, even with well-defined criteria, the judgment of tonicity by clinicians might not always be consistent. The tonicity of the neoglottis can be seen as a continuous scale ranging from hypotonic to spastic. When trying to compose subgroups of tonicity, the boundaries between the subgroups might not be that clear. The use of quantitative measures is more useful in that respect. One of the problems remains, however, that especially hypertonicity is not so easily judged. In the majority of the patients, with hypertonicity the subneoglottic distance will enlarge during phonation, because of the pressure that is build up under the neoglottis, but when the tissues of the neck are rigid and fibrotic, this will not be the case. Other measuring methods, like manometry at the level of the neoglottis might be more useful in this respect. Overall, the standardized evaluation form that was developed appeared to be useful for the assessment of neoglottic characteristics. It enabled quick judgment and consensus was easily reached.

10.2.3.2 Videofluoroscopy versus perceptual evaluations

In relation to the perceptual evaluations of voice quality, the presence of a neoglottic bar during phonation, regurgitation of barium during phonation, and the tonicity of the neoglottis during phonation, appeared to be important characteristics to judge, and the quantitative measures minimal neoglottic distance during phonation and increase of the maximal sub-neoglottic distance from rest to phonation are important measures to obtain, in relation to voice quality.

10.2.3.3 Videofluoroscopy versus acoustic analyses and maximum phonation time

In relation to acoustic analyses the presence of a neoglottic bar, flattening of the neoglottic bar during swallowing, regurgitation of barium during phonation, and tonicity of the neoglottis during phonation are important characteristics to judge and quantitative measures of the minimal neoglottic distance during phonation and the increase of the maximal sub-neoglottic distance from rest to phonation are important measures to obtain.
Also, regarding maximum phonation time some relations were found with the neoglottic characteristics, indicating that this simple objective measure is also valuable in this respect.

10.2.4 DIGITAL HIGH-SPEED IMAGING

Although it is very common in normal laryngeal voices to use stroboscopy as a diagnostic tool to obtain a (virtual) slow-motion image of the vibrating vocal folds from above, this is not the case in tracheoesophageal voice. This is mainly due to the irregular vibration of the neoglottis, which causes mis-triggering of the stroboscopic light pulses, making this method useless for a substantial portion of the tracheoesophageal voices. With endoscopic digital high-speed imaging, a research tool enabling recording with a very high-frequency that provides a slow-motion image of the real vibration and is developing rapidly into a clinically useful method, it became nevertheless possible for us to get a 'birds-eye' view of the movements of the neoglottis during tracheoesophageal speech.

First, a standardized evaluation protocol was developed for the evaluation of the digital high-speed recordings. Then, the results were related to the perceptual evaluations and acoustic analyses (including maximum phonation time) of tracheoesophageal voice quality in order to gain insight in the relation between the neoglottic characteristics and voice quality. Furthermore, the visual assessments of the view from above seen in the digital high-speed recordings are related to the visual assessments of the lateral view seen in the videofluoroscopy recordings in order to gain insight in the additional value of digital high-speed imaging to the videofluoroscopy recordings.

10.2.4.1 Evaluation protocol for digital high-speed imaging of the neoglottis

Since this was the first time that neoglottic characteristics in digital high-speed recordings were described, no literature was available on that subject. The evaluation protocol was based on existing protocols for stroboscopy in normal vocal folds. First, the recordings were judged by 6 raters in order to compose a set of characteristics that can be judged in these recordings, then the final evaluation protocol was composed and consensus judgments were performed.

The high-speed recordings in the present thesis had a resolution of 128 x 128 pixels. Meanwhile higher resolutions are already possible. Future developments of these high-speed imaging techniques (larger resolution, color images, objective analysis methods of the images) might improve their usefulness even more and are promising towards a better understanding of neoglottic vibrations. Digital high-speed imaging could, for instance, be used for the investigation of the production of voiced-unvoiced transitions, production of different fundamental frequencies and loudness levels, or the ability to whisper.

10.2.4.2 Digital high-speed imaging versus perceptual evaluations

Relations with the perceptual evaluations showed that the visibility of the origin of the neoglottis, the amount of saliva interfering with the vibration, the shape of the neoglottis and the regularity of the vibration appeared to be the most relevant characteristics in relation to the voice quality. Not all relations can be explained at this point, especially the relation between the shape of the neoglottis and the perceptual characteristics require further study.

10.2.4.3 Digital high-speed imaging versus acoustic analyses

Regarding the relations between the visual assessments of the digital high-speed imaging recordings and the acoustic analyses, the same neoglottic characteristics (as for the perceptual evaluations) appeared to be relevant and no additional relations with voice quality were found.
10.2.4.4 Digital high-speed imaging versus videofluoroscopy

Two parameters appeared to show some overlap: (1) the judgment of the visibility of the origin of the neoglottis in the digital high-speed imaging recordings and the judgment of the presence of a neoglottic bar in the videofluoroscopy recordings, and (2) the judgment of the amount of saliva in the digital high-speed recordings and regurgitation of barium in the videofluoroscopy recordings. Regarding the visibility of the origin of the neoglottis, videofluoroscopy seems to be the best choice since in the lateral view also a deeper situated neoglottis can be seen. Regarding regurgitation, digital high-speed imaging might be a better representative for daily life since it requires no swallowing of barium before phonation, which might increase regurgitation. The remaining assessments did not show any overlap, indicating that digital high-speed imaging can indeed provide valuable extra information complementary to videofluoroscopy.

At present more is known about the relation between the neoglottic characteristics as seen in videofluoroscopy and tracheoesophageal voice quality, due to the longer history and clinical availability of videofluoroscopy. The use of digital high-speed imaging is new and in the present study the first relations between the characteristics of the neoglottis that can be seen with it are described. Future research into this subject might increase our insight in these relationships and might provide explanations for some of the relations found.

10.2.5 Sociodemographic and clinical factors

Although the patient sample in the present study was not composed to study the influence of specific sociodemographic and clinical factors on tracheoesophageal voice quality and anatomical and morphologic characteristics of voice quality, whenever possible, influences were studied. The factors studied were sex, age, postoperative follow-up, reconstruction (standard total laryngectomy versus partial or full reconstruction of the pharynx), myotomy, neurectomy, radical neck dissection, and radiotherapy.

Regarding the perceptual evaluations and acoustic analyses of tracheoesophageal voice quality, an influence was found of reconstruction, radiotherapy, and radical neck dissection. Results are pointing to a less good voice in the reconstructed group, a less good voice in patients who received postoperative radiotherapy compared to the patients who receive radiotherapy as a primary treatment and a slower speaking rate and shorter maximum phonation time in the patient group that underwent a radical neck dissection.

Regarding the anatomical and morphologic characteristics of the neoglottis, for videofluoroscopy a relation was found with reconstruction, radical neck dissection, and age. Results showed less favorable neoglottic characteristics in the reconstruction group, something that can be expected due to the larger extent of surgery. Also, neoglottic characteristics were less favorable in the patient group that had undergone a radical neck dissection and in patients of older age. For digital high-speed imaging a relation was found for reconstruction, sex, and myotomy. As for the videofluoroscopy recordings, the influence of reconstruction is easy to explain, the other aspects require however further study in patient populations that are composed especially for that purpose.

The investigation of the influence of sex confirmed again the well-known fact that there are no consistent differences between male and female tracheoesophageal speech. Although this is an already well-established and clinically generally accepted phenomenon, it should be stressed here that the impact of such a male-sounding voice does give problems in daily life for the female laryngectomized patients. The need for developments towards a higher pitched voice for the female patients should receive more attention.
10.3 Discussion

Of course, one should strive towards a 'good' voice for all tracheoesophageal speakers. The investigations in this thesis give an indication for those neoglottic characteristics that have to be present to achieve good voice quality. Results showed large differences among the patients for all parts of the study. Thus, large differences in anatomy and morphology of the neoglottis are seen, and consequently, large differences in voice quality according to the perceptual evaluations and acoustic analyses exist.

The main goal of the present study was to find relations between anatomical and morphologic characteristics of the neoglottis and tracheoesophageal voice quality. The next step is of course to give some indications towards surgical techniques that can influence these neoglottic characteristics in a positive way, both at the time of surgery and as a secondary procedure ('phonosurgery' of the neoglottis in analogy to 'phonosurgery' of the glottis) in case of insufficient voice quality after surgery. In this thesis an attempt was made to relate all of the different parts of the study to the sociodemographic and clinical factors that were known about these patients. Although some relations were found, it should however be kept in mind that this study was not designed for this purpose and that additional studies are needed to study the specific influence of some surgical techniques and other factors like radiotherapy and radical neck dissection.

It should also be kept in mind that other clinical investigation methods, such as electromyography, fiberoptic studies, videokymography, manometry and aerodynamic investigations were not used in the present study, but that they could of course also provide insight in the relations between voice quality and characteristics of the neoglottis. The recommendations for a clinical evaluation protocol for tracheo-esophageal voice presented below are based on the evaluation methods used in this study, which does not automatically exclude the use of other investigation methods. At this point we cannot make any specific suggestions for parameters of other clinical investigation methods that could give valuable information, but we do suggest including a psychosocial questionnaire, such as the Voice Handicap Index (Jacobsen et al., 1997), in the clinical investigation protocol as well.

10.4 Clinical Investigation Protocol

Based on the four parts that the present study consisted of (perceptual evaluations, acoustic analyses, videofluoroscopy, and digital high-speed imaging) a proposal is made for a clinical evaluation protocol. This protocol contains parameters of all four parts of the study that appeared to be useful in relation to the voice quality. The parameters suggested here are supposed to be seen as a minimal set of evaluation parameters. Depending on the goal of the evaluations or specific research questions, also other parameters can of course be added.

To summarize, a clinical investigation protocol should contain:

1. a perceptual judgment of a trained speech-pathologist of the overall voice quality (good-reasonable-poor) and a judgment of the subset of the 8 perceptual semantic bipolar 7-point scales (deviant-normal, ugly-beautiful, breathy-not breathy, hypotonic-not hypotonic, low-high, deep-shrill, slow-quick, and dragging-brisk) based on read-aloud text;
2. acoustic signal typing based on a narrow-band spectrogram (100 ms analysis window) of a 2 s voice sample of a sustained /a/ at comfortable pitch and loudness level;
3. calculation by an experienced investigator of the following 6 acoustic parameters: median fundamental frequency, standard deviation of fundamental frequency, percentage of voiced, harmonics-to-noise ratio, glottal-to-noise excitation ratio, and
band energy difference, on the same 2 s of a sustained /a/ at comfortable pitch and loudness;

(4) measuring of maximum phonation time for a sustained /a/;

(5) visual assessment by a clinician (ENT-specialist, radiologist, or speech-language pathologist) of the presence of a neoglottic bar during phonation, regurgitation of barium during phonation, and tonicity of the neoglottis during phonation in videofluoroscopy recordings and quantitative measurement of the minimal neoglottic distance at rest and during phonation, the surface area of the neoglottic bar at rest and during phonation, the prominence of the neoglottic bar at rest and during phonation and the increase of the maximal sub-neoglottic distance from rest to phonation in two representative digitalized images (rest and phonation) of videofluoroscopy recordings;

and when available:

(6) assessment of the visibility of the neoglottis, the visibility of the origin of the neoglottis, the amount of saliva interfering with neoglottic vibration, the shape of the neoglottis and the regularity of the vibration in digital high-speed imaging recordings.

10.5 GENERAL CONCLUSIONS

In the present study protocols for standardized evaluation of the various aspects of tracheoesophageal voice quality and voice production are presented and used for investigation. Perceptual evaluations have shown that tracheoesophageal voice can still be considered as very deviant from a normal voice as reflected in the judgments of the naive listeners. Trained expert listeners judged the tracheoesophageal voices more positive and differentiated more between the several perceptual aspects.

Relations with the acoustic analyses have shown that with the proposed acoustic signal typing system already a good impression of the voice quality can be obtained. Moderate to strong correlations were found between the acoustic measures and the perceptual evaluations, indicating that a number of acoustic measures are valuable towards a more objective evaluation of tracheoesophageal voice quality.

The visual assessment of the presence of a neoglottic bar, the tonicity of the neoglottis, and regurgitation of barium during phonation and the quantitative measures minimal distance during phonation and at rest and the increase of the maximal sub-neoglottic distance from rest to phonation have shown to be important anatomical and morphologic correlates of tracheoesophageal voice quality that can be studied with videofluoroscopy.

The visibility of the origin of the neoglottis, the amount of saliva, the shape of the neoglottis and the regularity of the vibration have shown to be important anatomical and morphologic correlates of tracheoesophageal voice quality that can be studied with digital high-speed imaging.

Based on the results of the investigations in the different parts of this study, a clinical evaluation protocol was developed, that consists of a set of parameters for the evaluation of tracheoesophageal speech in clinical practice. The proposed evaluation protocol now requires further testing, and studies of the relevance of other evaluation methods in relation to tracheoesophageal voice quality are needed to complement the protocol.

10.6 FUTURE PERSPECTIVES

In general, it can be stated that for a good tracheoesophageal voice, optimal closure of the neoglottis together with an optimal tonicity (not hypotonic, not extremely hypertonic, but preferably normotonic or slightly hypertonic) are the most important determinants. Efforts
should be made to achieve these optimal neoglottic determinants in the majority of the laryngectomized patients. Obtaining quantitative measures of neoglottic characteristics in videofluoroscopy recordings provides the most objective results in order to investigate these characteristics. However, tonicity remains a characteristic that is difficult to judge on a visual basis only. Tonicity can be seen as a continuous scale ranging from hypotonicity to normotonicity, hypertonicity and in the extreme case spasm. Especially to define the boundary between normotonicity and hypertonicity, and to decide when the hypertonicity starts to negatively influence voice quality is a difficult aspect to judge. Other investigation methods such as manometry at the level of the neoglottis and aerodynamic measures might improve the possibilities of judging tonicity.

Although some attempt were made to study the influence of clinical factors on the neoglottic characteristics, no strong specific indications were found towards surgical techniques leading to those optimal characteristics in order to achieve a good voice. The present study was not aimed at the investigation of these clinical factors, but the need for such studies is obvious. The next step could be to select a patient sample specifically for comparison of surgical tonicity-intervention techniques, such as myotomy and neurectomy, using the clinical evaluation protocol presented in section 10.4, probably complemented with some other research methods such as manometry and quality-of-life questionnaires. For this, most probably prospective randomized studies are needed.

From the present study, already some thoughts about 'phonosurgery' of the neoglottis arose. Knowing which neoglottic characteristics are important, may eventually lead to some surgical techniques that can be used as a secondary procedure after total laryngectomy, when the voice quality is poor. In this respect it is, however, important to realize that optimizing the neoglottic characteristics for voice quality, might lead to swallowing problems, since the neoglottis/PE-segment is used for both functions. What is 'good for voicing' might not necessarily be 'good for swallowing'. Possible techniques and thoughts about 'phonosurgery' of the neoglottis should, therefore, always take both speaking and swallowing into account.

Furthermore, the evaluation protocol that was proposed in section 10.4 could be used to investigate tracheoesophageal voice quality in a longitudinal study. For instance, for investigating the development of the voice quality during voice training in the postoperative rehabilitation period, or for investigation of voice quality changes during radiotherapy.

Not only investigations in order to achieve a good voice, but also investigations towards achieving a more female sounding voice for the female patients are necessary. The present study once again showed that there are no differences between male and female voices after total laryngectomy. Although some of the female patients can probably be identified as such by their voice, their majority has a male sounding voice. Recently, a voice-producing element that can be placed into the shaft of the voice prosthesis has been described (Van der Torn et al., 2001). With this element a higher, although relative monotonous, pitch can be achieved. It is however only useful for the female patients that are aphony due to a hypotonic neoglottis. The device is presently useless in patients with a vibrating neoglottis with its own fundamental frequency, then the two frequencies interfere and the voice sounds diplophonic. In the present study it was already shown that fundamental frequency was related to the tonicity of the neoglottis as well as to the surface area of the neoglottis. Extending the knowledge about the neoglottis in relation to fundamental frequency, as well as detailed investigation of tracheoesophageal speech of female laryngectomized patients that indeed have a female sounding voice (in order to extract other characteristics apart from fundamental frequency that influence the perception of female tracheoesophageal speech), and thereby probably formulating possible solutions for this problem, is necessary to improve quality-of-life after total laryngectomy in female patients.
In the present study all acoustic analyses were performed on a sustained /a/ while the perceptual evaluations were performed on read-aloud text. The use of read-aloud text is more natural and most probably reflects the voice quality better than a sustained /a/. Acoustic analyses on sustained /a/ in those irregular voices, is however already something that should be done with some care and using running speech would even make it more troublesome. It is however an aspect that deserves some attention in future, since acoustic analysis on running speech would most probably improve the relation between the perceptual evaluations and the acoustic parameters. Translating the criteria used for acoustics signal typing in sustained /a/ into useful criteria for running speech might be a first start to incorporate the use of running speech in acoustic analysis of tracheoesophageal speech into an evaluation protocol.

The investigations in the present study were limited to voice quality in tracheoesophageal speech. It is, however, known that also the intelligibility decreases. Therefore, apart from this thesis, also some related studies on intelligibility were performed. Polak and Roeleven (1999) conducted a pilot study on the intelligibility of Dutch consonants and found that especially the feature voiced-voiceless was disturbed, something that might be expected in relation to the function of the neoglottis. Some studies have shown that the aerodynamic-myoelastic theory for normal voice production (Van den Berg, 1958) cannot be applied directly to tracheoesophageal voice production. The aerodynamic properties of the pulmonary airflow are of course present in all patients, but it seems that the active myoelastic properties of the neoglottis are only present in part of the patients (Deschler et al., 1999; Moon and Weinberg, 1987). Such an active myoelastic function of the neoglottis is expected to influence the intelligibility as well as the voice quality in a positive way. Boon-Kamma (2001), investigating the intelligibility of consonants in the same patient group used in the present study, also found problems with the voiced–voiceless distinction, as well as with the production of the /h/ sound. The pulmonary air, passing the neoglottis will cause its vibration, and only some of the patients seem to have the ability of adjusting the neoglottis in order to produce voiced or voiceless consonants properly. The fact that not only voiced sounds are produced as voiceless, but also vice versa, might be explained by the fact that when only aerodynamic forces are assumed, this should be related to the closure of the neoglottis. It might be that in the patients with a closed neoglottis, it is more likely that voiceless consonants are produced as their voiced counterparts, while in the patients with an incomplete neoglottic closure it is more likely that the voiced consonants are produced as their voiceless counterparts. In the same study of Boon-Kamma (2001) a relation was found between the intelligibility and the voice quality as studied in the present study. Apparently, with a well-functioning neoglottis (aerodynamic-myoelastic), not only a good voice quality can be achieved, but also a better intelligibility. In the pilot study by Polak and Roeleven (1999), initially conducted to gain insight in problems with the voiced–voiceless distinction in Dutch, also some problems with vowel intelligibility were noted. This might sound somewhat surprising, since it cannot be explained by the change of the voice source. However, with the removal of the larynx, also the vocal tract changes, the new voice source might be situated at a higher or lower level than the vocal cords and the shape of the pharynx has changed. To get more insight in the problems with vowel intelligibility, Oubrie (1999) studied the vowel intelligibility of the patients participating in the present study, and found indeed a change in formant frequencies and a decreased vowel intelligibility.

The results of these three studies show once more, that with the removal of the larynx and thus the vocal cords, not only the voice quality is affected, but also the intelligibility. More detailed studies into the myoelastic function of the neoglottis and investigation of the behavior of the neoglottis in comparison to the glottis (by means of digital high-speed imaging) are needed. In this respect one can think of investigation of the neoglottis when
producing voiced or voiceless sounds, voiced-voiceless transitions, and gradual changes from voiceless to voice (/h/ sound), or, for instance, when whispering.

Another question, yet unanswered, regarding intelligibility is the usefulness of articulatory training. A study should be conducted to investigate, whether articulatory training positively influences intelligibility. The investigation of the value of training is necessary, since it might as well be that in those patients, where the neoglottis only plays a passive, aerodynamic role, articulation exercises will not improve the ability to distinguish between voiced and voiceless sounds.

One final aspect that was not addressed in the present thesis and that also deserves more attention is the investigation of the relations between tracheoesophageal voice and speech quality and quality-of-life. Future investigations into this subject might provide additional information relevant for rehabilitation after total laryngectomy.

REFERENCES


