Quantitative Videofluoroscopy
A New Evaluation Tool for Tracheoesophageal Voice Production

Corina J. van As, MSc; Bas M. R. Op de Coul, MD; Frank J. A. van den Hoogen, MD, PhD; Florien J. Koopmans–van Beinum, PhD; Frans J. M. Hilgers, MD, PhD

Objective: To develop a quantitative videofluoroscopy protocol using well-defined visual parameters and quantitative measures for the evaluation of anatomical and morphologic characteristics of the neoglottis in relation to perceptual evaluation of tracheoesophageal voice quality.

Design: A patient survey.

Setting: The Netherlands Cancer Institute, Amsterdam.

Patients: Thirty-nine individuals with laryngectomies, 30 with standard total laryngectomy and 9 with a partial or total pharynx reconstruction.

Interventions: Videofluoroscopy, speech recordings.

Main Outcome Measures: Well-defined visual parameters and quantitative measures based on videofluoroscopy images should improve the evaluation of neoglottic characteristics in relation to voice quality.

Results: Quantitative measures were significantly related to visual assessment outcomes. Tonicity ($P=0.02$) and presence of a neoglottic bar during phonation ($P=0.03$) were significantly related to voice quality, as were several quantitative measures, especially the minimal distance between the neoglottic bar and anterior esophageal wall at rest ($P<0.001$) and during phonation ($P=0.02$), and the index for the relative increase of the maximal subneoglottic distance from rest to phonation ($P=0.01$).

Conclusions: This new quantitative videofluoroscopy protocol is a useful tool for the study of the anatomy and morphology of the neoglottis. With this protocol, characteristics relevant to tracheoesophageal voice quality can be defined. The quantitative measures are promising for a more standardized evaluation of the neoglottis in individuals who have undergone laryngectomy.


Original Article

SINCE THE introduction of the first useful voice prosthesis 2 decades ago,1 tracheoesophageal puncture has become a widely accepted and successful method of voice restoration after total laryngectomy.2 The main advantage of this type of vocal rehabilitation compared with conventional esophageal speech is that it is pulmonary driven. Tracheoesophageal speech has been proven to be closer to normal laryngeal speech than esophageal speech regarding acoustical characteristics,3,6 perceptual characteristics,7 and intelligibility.8 Since tracheoesophageal voice rehabilitation has become the method of choice after total laryngectomy,2 the main issue in voice rehabilitation is no longer the ability to acquire speech: up to 90% of the patients acquire a fair to excellent voice,9 whereas with esophageal speech, this figure is much lower and more variable, as only 25% to 50% of these patients are able to develop functional esophageal speech.10-12 However, one should keep in mind that the tracheoesophageal voice is variable in quality.13

The new sound source, referred to as the neoglottis, pseudoglottis, or pharyngo-esophageal segment, is considered to play an important role in voice production. Throughout this article the term neoglottis will be used. One of the methods frequently used for investigation of the neoglottis is videofluoroscopy. Videofluoroscopic studies have been performed far more often for esophageal voice14-24 than for tracheoesophageal voice.25,26 Previous studies on esophageal voice were conducted mainly to gain insight on factors that influence the acquisition of speech and were also focused on discovering the site or source of vibration of the substitute esophageal voice.25,26

Several researchers found that sound originated at the level of the cricopharyngeus muscle.14,20,27 However, one should keep in mind that it was not vibration of the mucous membrane that caused the sound, but vibration of the accumulated mucus above the neoglottis.28 A number of researchers
PATIENTS AND METHODS

PATIENTS

Thirty-nine patients who underwent laryngectomy, all of whom used tracheoesophageal speech by means of the Provox2 voice prosthesis (Atos Medical AB, Hörby, Sweden), were selected from a group of 173 patients with laryngectomies in follow-up at the Netherlands Cancer Institute. Special care was taken to compose a sample of patients in which all variations normally encountered in this group were represented. Therefore, both male and female patients, patients with poorer voice quality, and patients with a reconstructed pharynx and/or esophagus were included in the study. Informed consent was obtained from all patients we asked to participate after they received written information about the purpose of the study. There were 29 men and 10 women. Ages varied from 47 to 82 years, with a mean of 67 years. The postoperative follow-up varied from 11 months to 18 years, with a mean of 6 years. Thirty patients had a standard wide-field total laryngectomy; these patients constitute the standard group. Nine patients had a pharyngeal reconstruction; these patients constitute the reconstruction group. Four patients in the reconstruction group underwent partial repair with a myotomy of the cricopharyngeus muscle combined with a neurectomy of the pharyngeal plexus. In 5 patients a myotomy of the cricopharyngeus muscle combined with a neurectomy of the pharyngeal plexus was performed, while in 12 patients only a neurectomy of the pharyngeal plexus. In 5 patients a myotomy of the cricopharyngeus muscle combined with a neurectomy of the pharyngeal plexus was performed, while in 12 patients only a neurectomy of the pharyngeal plexus was performed. A unilateral neck dissection was carried out in 13 patients and a bilateral neck dissection in 5 patients. Sixteen patients who underwent laryngectomy for a recurring tumor received primary radiotherapy to treat their laryngeal cancer; 21 patients received radiotherapy after their total laryngectomy; and 2 patients received no radiotherapy. Table 1 gives an overview of the clinical information of the patients.

VIDEOFLUOROSCOPY

The videofluoroscopy recordings were obtained with a Philips Diagnost 92 system (Philips Medical Systems, Eindhoven, the Netherlands) together with a Panasonic NV-HD650 video recorder (Matsushita Electric Industrial Co, Osaka, Japan). Videofluoroscopic recordings were made of all patients vocalizing 2 phonations of the sustained vowel /a/ at a comfortable pitch and loudness level. All x-ray film recordings were made in lateral view; patients were asked to swallow barium and phonate. A reference coin was stuck to the cheeks of the patients to enable the quantification of the different dimensions.

SPEECH RECORDINGS

Recordings for the perceptual evaluation were made of one fixed text that was read aloud. The recordings were made with use of the Computerized Speech Lab (CSL) (Kay Elemetrics, Lincoln Park, Nj). A standard headset microphone that came with the equipment was used, and through the hardware of this system, with use of the CSL software, the speech data were directly recorded on a digital audiotape (DAT) by means of a Sony TCD-8 DAT recorder (Sony Electronics Inc, Park Ridge, Nj). For the perceptual evaluation, the read-aloud texts of all speakers were randomly recorded on another DAT. Each speaker repeated the text until 2.5 minutes had been recorded on the tape.

VISUAL ASSESSMENT OF ANATOMICAL AND MORPHOLOGIC CHARACTERISTICS OF THE NEOGLOTTIS

In the pilot phase, 3 judges jointly viewed all videofluoroscopy recordings using the definitions of tonicity proposed by McIvor et al and Van Weissenbruch. These definitions include the tonicity not only during phonation, but also during swallowing and at rest. It was extremely difficult to reach consensus among the 3 judges using these definitions. A large number of our patients could not be categorized into 1 tonicity group because they did not meet all the criteria for one particular group or because they met the criteria for several groups.

Therefore, the assessment had to be adjusted considerably: the flattening of the neoglottic bar during swallowing and the appearance of a neoglottic bar at rest were judged separately from the tonicity of the neoglottis during phonation. Also, the presence or absence of regurgitation and stasis of barium contrast, as well as the level of the neoglottis relative to the cervical vertebrae, were added. With these adjustments, consensus was reached much more easily.

Apart from these more or less objective assessments, tonicity during phonation was judged subjectively using the following criteria: The tonicity of the neoglottis was judged normotonic when there was closure of the neoglottis, ie, complete or almost complete dynamic contact of the neoglottic bar with the anterior esophageal wall during phonation. The tonicity of the neoglottis was judged as hypotonic when there was no closure of the neoglottis during phonation and as hypertonic when the neoglottis was fully closed during phonation combined with considerable dilation of the esophagus below the neoglottis. Spasm was defined as complete

found that the ability to acquire esophageal speech was related to the anatomical and morphologic characteristics of the neoglottis, such as dilatability of the hypopharynx and shape of the neoglottis, form of the neoglottis, extent of surgery, and tonicity of the pharyngoesophageal segment. Others could not find any relationship between good voice quality and variations in the anatomy and morphology of the neoglottis. Some even thought that the acquisition of voice was merely related to psychological factors. The age of the patient was also found to be an important factor in the ability to learn esophageal speech. Videofluoroscopic studies regarding tracheoesophageal speech showed that the visual characteristics of the vibratory segment of tracheoesophageal and esophageal speakers were similar. Unfortunately, both these studies lacked the assessment of voice quality in relation to the observed characteristics of the neoglottis.
closure of the neoglottis with extreme dilation of the esophagus below the neoglottis during attempted phonation with no passage of air through the neoglottis. Stricture was defined as narrowing of the esophagus with no dynamic changes in any of the situations judged separately.

QUANTITATIVE MEASUREMENTS OF THE NEOGLOTTIS

In addition to the visual assessment of the anatomical and morphologic characteristics, metrical measures were also used in the evaluation protocol. These quantitative measures of the neoglottis were obtained using a software program called Drawer (developed by M. B. van Herk, physicist at the Netherlands Cancer Institute/Antoni Leeuwenhoek Hospital, Amsterdam), which was initially designed to measure tumor volumes for the purposes of radiotherapy. Relevant frames of the neoglottis both at rest and during phonation were selected from the recordings (Figure 1), digitalized with a frame grabber, and saved as an image file. From these digitalized images, the quantitative measures were calculated. Distances were measured (in pixels) with a ruler-like tool. Areas were measured by indicating the region of interest with a paintbrush tool, after which the computer program counted the number of pixels in the region. Figure 2 shows the neoglottis, along with indications of the measures performed. All measures in pixels were converted to millimeters or square millimeters using a coin with a known diameter.

PARAMETERS MEASURED

The minimal distance (in millimeters) was measured as the distance between the neoglottic bar and the anterior wall of the esophagus (ie, the width of the neoglottis) at rest (MINREST) and during phonation (MINPHON). The maximal subneoglottic distance (in millimeters) was measured as the maximal width of the esophagus below the neoglottis at rest (MAXREST) and during phonation (MAXPHON). The line on which the measurements were made was placed perpendicular to the posterior wall. The surface area of the neoglottic bar in lateral view (in square millimeters) was measured at rest (SURREST) and during phonation (SURPHON) in the lateral view. The prominence of the neoglottic bar toward the anterior wall (in millimeters) was measured at rest (PROMREST) and during phonation (PROMPHON). The line on which the measurement was based was placed perpendicular to the posterior wall at the most prominent point of the neoglottic bar.

In addition to these quantitative measures, the MAXPHON-MAXREST index was also calculated in order to reflect the increase of the maximal subneoglottic distance during phonation. This index is thought to give an impression of the tension of the closure of the neoglottis—tighter closure of the neoglottis may reflect a larger increase in subneoglottic distance, although this increase may also be dependent on the rigidity of the subneoglottic tissues.

PERCEPTUAL EVALUATION OF VOICE QUALITY

Four speech and language pathologists experienced in the treatment of patients with laryngectomies were trained in the perceptual evaluation of this patient group. The evaluation involved 19 bipolar semantic 7-point scales and one overall judgment of voice quality in which the voice was judged as good, reasonable, or poor. A good voice was defined as “almost similar to a normal voice,” a poor voice was defined as “very deviant from a normal voice,” and a reasonable voice was defined as “somewhere in between both extremes.” At the time, the results of the extended perceptual evaluation of the semantic scales were under investigation. In the present study, only the results of the overall judgment were used. The interrater reliability calculated with Cronbach’s α was .88. In order to compare the videofluoroscopy recordings, the speakers were divided into 3 subgroups on the basis of overall voice quality. A voice was considered good or poor if at least 2 of the 4 listeners evaluated it as such. Voices that were judged good (or poor) by 1 listener and reasonable by 3 listeners were considered reasonable. In no instance was a voice judged good by one listener and poor by another.

STATISTICS

Statistical analysis was performed using the Statistical Package for Social Sciences, version 7.5 (SPSS Inc, Chicago, Ill). Paired t tests were used to compare the quantitative measures between rest and phonation, and χ² tests for linear-to-linear association were used to investigate the relations between voice quality, visual assessment, and clinical parameters, as well as the relations between visual assessment and voice quality. For the relation between voice quality and tonicity, an exact χ² test was used. Because of concerns regarding assumptions of normality for the quantitative measurements, nonparametric tests were used. Relations between the clinical parameters and the quantitative measurements and between the visual assessment (except for tonicity) and the quantitative measurements were investigated by means of either a t test for independent samples or a Mann-Whitney test, depending on the distribution of the observed values. Relations between tonicity (3 subgroups) and the quantitative measurements and between voice quality (3 subgroups) and the quantitative measurements were investigated by means of analysis of variance followed by post hoc Tukey tests or by means of Kruskal-Wallis tests followed by Mann-Whitney tests with a Bonferroni correction, depending on the distribution of the observed values. In the case of obvious differences between SDs according to the Levine test of equality of variances, a modified t test was used.

Visual fluoroscopy would gain in value if the visual assessment of anatomical and morphologic characteristics of the neoglottis were standardized and if these standards could be combined with quantitative measurements of the different dimensions of the neoglottis. This would make easier the establishment of a relationship between form and voice quality. Therefore, a novel assessment protocol for quantitative videofluoroscopy was developed and evaluated in relation to the perceptual evaluation of voice quality.

Table 2 gives the results of the visual assessment of the anatomical and morphologic characteristics of the neoglottis for both speaker groups. At rest, 19 patients showed 1 neoglottic bar and 4 patients showed 2 neoglottic bars. During phonation, 21 patients showed 1 neoglottic bar and...
3 patients showed 2 neoglottic bars. The number of patients with 2 neoglottic bars at rest (n=4) or during phonation (n=3) was too small for a separate evaluation in statistical analysis. Since the voice quality of the groups with 1 and 2 neoglottic bars was thought to be comparable, these groups were taken together in further analyses. The 1 patient with stricture was left out of the analysis regarding tonicity, since no meaningful statistical analysis is possible with only one patient in a subgroup.

In the standard group, the neoglottis at rest was mostly situated around cervical vertebrae C4 and C5. During phonation, however, the neoglottis was situated somewhat higher in most speakers, with an upward shift from C4-5 to C3-4 (Table 3).

### QUANTITATIVE MEASUREMENTS OF THE NEOGLOTTIS

Quantitative measurements are given in Table 4. Paired t tests between the measures at rest and during phonation showed statistically significant differences in the standard group for 2 measurements: MAXPHON was larger than MAXREST (P<.001), and PROMPHON was larger than PROMREST (P<.001). No differences were found in the reconstruction group.

### VOICE QUALITY

Voice quality could be judged for 38 of the 39 patients, since 1 patient died before the speech sample could be obtained. Voice quality was judged as good for 13 patients, reasonable for 14, and poor for 11.

### RELATION BETWEEN CLINICAL PARAMETERS AND VISUAL ASSESSMENT OF THE NEOGLOTTIS

The neoglottic bar at rest was more often visible in the standard group than in the reconstruction group (P=.003). Furthermore, patients with a standard total laryngectomy more often had a normotonic or hypertonic neoglottis during phonation than patients with a reconstructed pharynx (P=.02).

Within the standard group, there was no relation between visual assessment of the neoglottis and the clinical parameters myotomy, neurectomy, radiotherapy, neck dissection, age, postoperative follow-up, and sex. Relations with visual assessments were investigated only within the standard group; either the reconstruction group was too small or the parameters were invalid.

### RELATION BETWEEN VISUAL ASSESSMENT AND QUANTITATIVE MEASURES OF THE NEOGLOTTIS

The SURREST and SURPHON measurements were larger in the standard group than in the reconstruction group (SURREST, P=.02; SURPHON, P=.01), and the PROMREST measurements were smaller in the reconstruction group than in the standard group (P=.01). The MINREST and MINPHON measurements were smaller in the standard group (MINREST, P=.001; MINPHON, P=.01) than in the reconstruction group.

Results within the standard group also revealed a relationship between clinical parameters and quantitative measures. Whether the patient had a radical neck dissection appeared to influence some measurements of the neoglottis. The MINPHON measurement was smaller in the subgroup without neck dissection (P=.04), indicating that this group had a more closed neoglottis. Age was another factor that was associated with differences in the measurements; the MINREST measurement appeared smaller in the younger patient group (<70 years) (P=.048), indicating a narrower neoglottis in the younger patient group and a looser neoglottis in the older patient group. The clinical parameters myotomy, neurectomy, postoperative follow-up, radiotherapy, and sex were not associated with any differences in the quantitative measurements.

### RELATION BETWEEN CLINICAL PARAMETERS AND VOICE QUALITY

The χ² tests for linear-to-linear association did not reveal any relation between voice quality and the clinical parameters reconstruction, myotomy, neurectomy, radiotherapy, neck dissection, age, postoperative follow-up, and sex.

### RELATION BETWEEN VISUAL ASSESSMENT AND QUANTITATIVE MEASURES OF THE NEOGLOTTIS

Relations between visual assessment and quantitative measures of the neoglottis were based on the results for all of the speaker groups, since the type of surgery was irrelevant. In the subgroup with the appearance of a neoglottic bar at rest, SURREST and SURPHON, PROMREST and PROMPHON, and MAXPHON were larger (SURREST, SURPHON, and PROMREST, P<.001; MINPHON, P=.003; MAXPHON, P=.03) and MINREST was smaller (P=.01) than in the subgroup without the appearance of a neoglottic bar at rest.
The assessment of the appearance of a neoglottic bar during phonation showed several relationships with the quantitative measures. In the subgroup with a neoglottic bar during phonation, SURPHON and SURREST, PROMPHON and PROMREST, and MAXPHON and MAXPHON/MAXREST were larger (SURPHON, PROMPHON, PROMREST, $P<.001$; SURREST, $P = .01$; MAXPHON, MAXREST, $P = .01$) and MINREST and MINPHON were smaller ($P < .001$) than in the subgroup without a neoglottic bar during phonation. These results indicate that the presence of a neoglottic bar during phonation was related to a shorter distance between the neoglottic bar and the anterior wall of the esophagus. Likewise, a larger subneoglottic distance during phonation was related to a larger SURPHON and a greater PROMPHON, as well as a relatively larger increase in MAXPHON.

The tonicity of the neoglottic bar during phonation, when divided into the subgroups hypotonicity, normotonicity, and hypertonicity, also showed a relationship with the quantitative measures (Table 5), with a clear distinction between the 3 levels of tonicity. For instance, whereas SURPHON, PROMPHON, MINREST, and MINPHON were distinctive between hypotonicity and normotonicy and between hypotonicity and hypertonicity, MAXREST and MAXPHON were distinctive between hypotonicity and hypertonicy and between normotonicy and hypertonicy.

The assessment of regurgitation of barium during phonation was also related to the quantitative measures. In

![Figure 1. Videofluoroscopy images of the situation at rest (A) and during phonation (B).](image-url)

![Figure 2. Schematic drawing of the videofluoroscopy images in Figure 1.](image-url)
the subgroup in which regurgitation of barium was observed during phonation, SURREST, SURPHON, PROMREST, PROMPHON, and MAXPHON were smaller (P = .02, P = .01, P = .02, P < .001, and P = .02, respectively) and MINREST and MINPHON were larger (MINREST, P = .003; MINPHON, P < .001) than in the subgroup in which no regurgitation of barium was observed during phonation.

These results indicate that regurgitation occurred when the neoglottic bar was small or not present as well as when the neoglottis was wide.

Regarding the cervical level of the neoglottis, stasis of barium above the neoglottic bar during phonation, and flattening of the neoglottic bar during swallowing, no relations with the quantitative measures were found.

**RELATION BETWEEN VOICE QUALITY AND VISUAL ASSESSMENT OF THE NEOGLOTTIS**

The visual assessment of the appearance of a neoglottic bar during phonation (Table 6) and the tonicity of the neoglottis (Table 7) showed significant relations with voice quality (P = .03 for appearance of a neoglottic bar; P = .02 for tonicity), such that a good voice was related to the appearance of a neoglottic bar during phonation.

Regarding tonicity, a good voice was related to a normotonic or hypertonic neoglottis. Among the good voices, a hypotonic neoglottis was never observed. The number of speakers in this analysis was 37, since 1 speaker with stricture also fell in the poor group; this speaker was not included in the statistical analysis.

The assessment of the appearance of the neoglottis at rest, regurgitation of barium during phonation, stasis of barium on the neoglottic bar during phonation, and flattening of the neoglottic bar during swallowing showed no relations with voice quality.

**RELATION BETWEEN VOICE QUALITY AND QUANTITATIVE MEASURES OF THE NEOGLOTTIS**

The index MAXPHON-MAXREST (P = .01), MINPHON (P < .001), and MINREST (P = .01) were different between the speaker groups (Table 8). On the basis of these 3 quantitative measures, a distinction can be made between a poor and a good voice, or between a reasonable and a good voice. The MINREST and MINPHON measures were smaller during phonation for a good voice than for a poor voice (MINREST, P < .001; MINPHON, P = .02), and better speakers showed a relatively larger increase in the MAXPHON-MAXREST index (P = .01). For the other quantitative measures, no relations with the overall judgment of voice quality were found.

**COMMENT**

Videofluoroscopy has proven to be an important tool for the assessment of the neoglottis for both esophageal and tracheoesophageal speech. Although videofluoroscopy is clinically valuable on the level of the individual...
patient, the descriptive nature of the evaluation and the lack of objective measures have hampered its widespread use as a research tool. The present study was started in order to develop videofluoroscopy into a more objective and reproducible evaluation instrument for the analysis of anatomical and morphologic characteristics of the neoglottis. Another objective of this study was the correlation of videofluoroscopy results with voice quality in tracheoesophageal speech, which has been lacking so far.

Visual assessment of the anatomical and morphologic characteristics of the neoglottis appears to be facilitated by judging the different phases of movement of the neoglottis (at rest, during phonation, and when swallowing) separately. With the method used by others,22,24 in which this distinction was not applied, consensus judgments appeared to be less easy and efficient. The method presented herein leaves only one subjective parameter—tonicity of the neoglottis during phonation. The remaining parameters are more or less objective because of their use of clear dichotomies and/or numbers.

Only 3 of 30 patients showed a double neoglottic bar during phonation, which contrasts with the findings of others, who have described this phenomenon in 5 of 16 and 3 of 4 patients.25,26 The neoglottis was located at the level of C4-5 in the majority of our patients, which is more cranial than the C5-6 level reported for the majority of patients in studies of esophageal speech.15,18,29 In our study, this level tended to rise by approximately half a vertebra during phonation, which contrasts with the findings of others.23,26,28 It is likely that this upward shift in tracheoesophageal voice was caused by the greater aerodynamic effect in the pulmonarily driven tracheoesophageal speech.

Our method of obtaining quantitative measures was easy and straightforward, using digital images and special image evaluation software with a reference marker to allow calculation of exact distances and surface areas. The quantitative measures showed a large variability in the speaker group that was reflected in the relatively high SDs. Differences were found between rest and phonation for the maximal subneoglottic distance and the prominence of the neoglottis in the standard group, indicating a dynamic change of the neoglottis from rest to phonation.

Perceptual evaluation is still the gold standard for the evaluation of voice quality.11 In the method applied, trained judges distinguished 3 voice quality subgroups: 11 speakers with poor voices, 14 with reasonable voices, and 15 with good voice quality. Voice quality is not easily defined. In this study, a voice that was considered close to normal was called good, and a voice that was very deviant from normal was called poor. A voice that was not very close to normal but also not very deviant, for instance, slightly bubbly or rough, was judged as reasonable. The high intrarater reliability of the perceptual evaluation shows that our subgroupings were consistent and reliable and therefore useful for comparison with the other outcome measures in this study.

### Table 4. Quantitative Measures of the Neoglottis for the Standard and Reconstruction Groups

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Standard</th>
<th>Reconstruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>MINPHON, mm</td>
<td>10.6 (4.7)</td>
<td>25.0 (13.9)</td>
</tr>
<tr>
<td>MAXPHON, mm</td>
<td>5.2 (3.8)</td>
<td>22.3 (11.9)</td>
</tr>
<tr>
<td>SURPHON, mm²</td>
<td>88 (67)</td>
<td>209 (116)</td>
</tr>
<tr>
<td>PROMPHON, mm</td>
<td>8.0 (5.6)</td>
<td>11.0 (6.7)</td>
</tr>
<tr>
<td>MAXPHON-MAXREST</td>
<td>1.8 (1.9)</td>
<td>15.0 (9.9)</td>
</tr>
<tr>
<td>MINPHON, mm</td>
<td>2.05 (1.39)</td>
<td>0.61 (0.30)</td>
</tr>
<tr>
<td>MAXPHON, mm</td>
<td>9.1 (4.6)</td>
<td>12.0 (7.0)</td>
</tr>
<tr>
<td>SURPHON, mm²</td>
<td>38 (27)</td>
<td>289 (170)</td>
</tr>
<tr>
<td>PROMPHON, mm</td>
<td>4.6 (5.37)</td>
<td>6.7 (3.7)</td>
</tr>
<tr>
<td>MAXPHON-MAXREST</td>
<td>0.61 (0.30)</td>
<td>0.61 (0.30)</td>
</tr>
</tbody>
</table>

*Values are mean (SD). MINPHON indicates minimal distance at rest; MAXPHON, maximal distance during phonation; SURPHON, surface area of the neoglottis during phonation; PROMPHON, prominence of the neoglottis during phonation; and MAXPHON, prominence of the neoglottis at rest.

### Table 5. Relations Between Quantitative Measures and Tonicity

<table>
<thead>
<tr>
<th>Quantitative Measures</th>
<th>Overall</th>
<th>Hypotonic (n = 14)</th>
<th>Normotonic (n = 15)</th>
<th>Hypertonic (n = 8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MINPHON, mm</td>
<td>.001</td>
<td>5.1 (3.9)</td>
<td>8.1 (4.7)</td>
<td>10.6 (4.7)</td>
</tr>
<tr>
<td>MAXPHON, mm</td>
<td>.001</td>
<td>9.8 (3.4)</td>
<td>9.5 (4.4)</td>
<td>10.6 (4.7)</td>
</tr>
<tr>
<td>SURPHON, mm²</td>
<td>.001</td>
<td>86 (74)</td>
<td>90 (63)</td>
<td>106 (72)</td>
</tr>
<tr>
<td>PROMPHON, mm</td>
<td>.001</td>
<td>10.2 (7.3)</td>
<td>12.9 (5.8)</td>
<td>11.0 (6.7)</td>
</tr>
</tbody>
</table>

*Values are mean (SD). For expansions of abbreviations, see the first footnote to Table 4.

*For these parametric measures, analysis of variance and the post hoc Tukey test were used.

†For these nonparametric measures, the Kruskal-Wallis test and Mann-Whitney test with Bonferroni correction were used.

‡For these parametric measures, analysis of variance and the post hoc Tukey test were used.
Some correlations were observed between clinical parameters and quantitative measures. The differences between the standard and reconstruction groups were not surprising because of the larger extent of surgery in the latter group. Our results suggest that the present reconstruction techniques result in less favorable conditions of the neoglottis. Furthermore, our quantitative measures indicate that neck dissection and age have an influence.

The MINPHON measure was smaller in the group without neck dissection and the MINREST measure was smaller in patients under 70 years of age. These quantitative measures are correlated with good voice quality. The finding that neck dissection seems to influence the anatomy of the neoglottis has not been reported earlier for tracheoesophageal voice and needs to be studied in larger series before any conclusions can be drawn, especially since results for esophageal voice are conflicting. Smith et al16 found that patients with a laryngectomy only had better voice results than patients with an additional neck dissection, whereas Richardson21 did not find any such influence.

Investigating the relationship between the results of the visual assessment and the quantitative measures was one of the main objectives of this study. Replacement of the more subjective visual assessments of the neoglottic characteristics by more objective and precise quantitative measures could allow the evaluation of videofluoroscopy recordings in a more consistent and standardized manner, and the results within and between studies could become easier to compare in the future. It is noteworthy that relations between quantitative measures and visual assessment were found for all parameters except for flattening of the neoglottic bar during swallowing and stasis of barium above the neoglottic bar during phonation. This suggests that the majority of the visual assessments might be replaceable by these quantitative measures.

Another important part of this study was to investigate the possible relations between the neoglottis and voice quality. To the best of our knowledge, this particular study has not yet been performed for tracheoesophageal speech. Results showed that voice quality was related to the appearance of a neoglottic bar during phonation and to the tonicity of the pharyngoesophageal segment. This is most obvious in the good group, members of which always had a visible neoglottic bar during phonation; none were hypotonic, but some were hypertonic. All 3 types of tonicity were seen in the poor and reasonable voice groups, which suggests that tonicity is not as clear an indicator of voice quality as might be expected. However, these results make it clear that hypotonicity is an unfavorable condition of the neoglottis regarding voice quality and should be avoided. On the other hand, hypertonicity of the neoglottis is clinically much less of a problem, since it still can be associated with a good voice and, in relevant cases, can be fixed relatively easily by surgical intervention (myotomy)31,32 or medical intervention (botulinum toxin type A injection [Botox; Allergan Inc, Irvine, Calif]). Hypotonicity can be fixed only by exerting digital pressure on the external neck, thereby enabling approximation of the esophageal tissues during phonation. Other forms of surgical intervention or phonosurgery of the neoglottis are not yet available for this problem. Since it is common nowadays to perform a neurectomy of the pharyngeal plexus and/or a myotomy of the pharyngeal muscles during total laryngectomy,31,32 it should be stressed that care should be taken to avoid overcorrection, which could result in hypotonicity.

Results of the investigation of the relations between voice quality and quantitative measures were interesting. The most important factor appeared to be MINPHON—the closer the neoglottis, the better the voice quality. This shows the relevance of closure for sound production, something that is already well known for normal laryngeal voices. Surprisingly, this measure was not used in earlier studies regarding esophageal speech. However, some quantitative measures were performed in some studies. They

Table 6. Relations Between Appearance of a Neoglottis During Phonation and Voice Quality*

<table>
<thead>
<tr>
<th>Neoglottis During Phonation</th>
<th>Poor</th>
<th>Reasonable</th>
<th>Good</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>1 or 2</td>
<td>7</td>
<td>9</td>
<td>13</td>
<td>29</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>14</td>
<td>13</td>
<td>38</td>
</tr>
</tbody>
</table>

*P=.03. Values are numbers of patients.

Table 7. Relations Between Tonicity of the Neoglottis During Phonation and Voice Quality*

<table>
<thead>
<tr>
<th>Tonicity</th>
<th>Poor</th>
<th>Reasonable</th>
<th>Good</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypotonic</td>
<td>6</td>
<td>8</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Normotonic</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>Hypertonic</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>14</td>
<td>13</td>
<td>37</td>
</tr>
</tbody>
</table>

*P=.02. Values are numbers of patients.

Table 8. Relations Between Quantitative Measures and Voice Quality*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Overall</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MINPHON, † mm</td>
<td>.009</td>
<td>3.6 (3.6)</td>
<td>2.5 (1.8)</td>
<td>0.5 (1.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P &gt; .99</td>
<td>P = .02</td>
<td></td>
</tr>
<tr>
<td>MINREST, † mm</td>
<td>.001</td>
<td>3.8 (4.0)</td>
<td>2.5 (2.6)</td>
<td>0.3 (0.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P &gt; .99</td>
<td>P = .02</td>
<td></td>
</tr>
<tr>
<td>MAXPHON-MAXREST</td>
<td>.005</td>
<td>1.4 (0.6)</td>
<td>1.7 (1.5)</td>
<td>2.9 (1.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P &gt; .99</td>
<td>P = .02</td>
<td></td>
</tr>
</tbody>
</table>

*Values are mean (SD). For expansions of abbreviations, see the first footnote to Table 4. †For these nonparametric measures, the Kruskal-Wallis test and Mann-Whitney test with Bonferroni correction were used.
were always performed during phonation, and consisted of the length of the neoglottis,29 the prominence of the neoglottis,19 the dilatability of the esophagus below the neoglottis,29 and the width of the hypopharynx.17 The only measure comparable to one of those used in our study was the “dilatability” of the esophagus, expressed by the MAXPHON-MAXREST index, which appeared to influence voice quality.29 These authors observed a relation between the acquisition of esophageal speech and the width of the esophagus, indicating that a wider subneoglottic distance was related to a better voice. They presumed that a greater amount of air would be obtained within the lumen, providing a sufficient amount of air for voice production. In the present study, we also found that the MAXPHON-MAXREST index differed between the voice quality groups. The good speakers showed a relatively larger increase in MAXPHON compared with MAXREST. Presumably, in tracheoesophageal speech, the increase in maximal subneoglottic distance is related to the tension of the neoglottis together with the flexibility of the tissues of the neck. In this respect, it should be noted that tension of the neoglottis that is too high (ie, extreme hypertonicity or spasm) led to a relatively large increase in the maximal subneoglottic distance, with very poor or even absent voice. Since there was a clear correlation between the visual assessment and the quantitative measures, it seems possible to substitute for the classic visual parameters the quantitative measures MINREST and MPHON, as well as the MAXPHON-MAXREST index. The only exception was the situation in which there was hypertonicity or spasm of the neoglottis in conjunction with fibrosis of the surrounding tissues in the neck. In such a case, hypertonicity did not lead to extreme dilatation of the subneoglottic region.

The finding that the quantitative measures showed no difference between the bad and the reasonable group can be explained best by the assumption that the neoglottic characteristics leading to a judgment of a poor or reasonable voice quality are diverse. Problems with saliva interference and/or regurgitation may decrease voice quality substantially, irrespective of the tonicity of the neoglottis. In conclusion, this study showed that it is possible to analyze videofluoroscopy images in a more quantitative manner and that some objective measures can be applied to replace the descriptive parameters formerly used to assess the neoglottis. The clear correlation of the videofluoroscopy results with good voice quality and the objective nature of the proposed assessment protocol may increase the usefulness of this imaging technique for the rehabilitation of individuals with laryngectomies.

Accepted for publication July 13, 2000.

Maurits and Anna de Kock Stichting provided financial support for the equipment needed for the speech recordings and the listening experiment.

We thank Louis Pols, PhD, for his critical review of the manuscript; the patients for their participation in this study; Guus Hart, MSc, for his help with the statistics; and M. B. van Herk, PhD, for providing and adapting the Drawer software for the quantitative measures. We also thank Benita Scholten, BS, Rianne Polak, BS, Marike Koster, BS, and Brigitte Boon-Kamma, BS, speech pathologists, for their participation in the data collection.

Corresponding author and reprints: Frans J. M. Hilgers, MD, PhD, Department of Otolaryngology-Head and Neck Surgery, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands (e-mail: fhilg@nki.nl).

REFERENCES