Acoustic and Perceptual Evaluation of Voice and Speech Quality

A Study of Patients With Laryngeal Cancer Treated With Laryngectomy vs Irradiation

Caterina Finizia, MD, PhD; Hans Dotevall, MD; Elisabet Lundström; Jörgen Lindström, MD, PhD

Objective: To compare voice and speech function in patients who underwent laryngectomy with that of 2 control groups.

Design: A cross-sectional study comparing acoustic and temporal variables with perceptual evaluations in 3 subject groups.

Setting: University hospital in Göteborg, Sweden.

Subjects: Two groups of patients with laryngeal carcinoma were examined: 12 male patients who had laryngectomy and were using a tracheoesophageal prosthesis and 12 male patients treated with radical radiotherapy who had a preserved larynx. The third group consisted of 10 normal controls without laryngeal disease.

Main Outcome Measures: Acoustic variables were fundamental frequency, absolute fundamental frequency perturbation, speech rate, and maximum phonation time. Perceptual evaluation included 15 listeners' perceptual evaluation and the patients' self-assessment of speech intelligibility, voice quality, and speech acceptability.

Results: No significant acoustic or temporal differences were found between the laryngectomy and radical radiotherapy groups. There was a significant difference between the patient groups in perceptual evaluation. Both groups of patients differed from normal controls in acoustic and temporal measures, where the laryngectomy group generally deviated more from the normal controls than the patient group treated with radiotherapy. There was a weak, but significant, correlation between absolute fundamental frequency perturbation and perceived voice quality.

Conclusions: Perceptual evaluations could indicate significant differences between the patients who underwent laryngectomy and irradiated patients, where the acoustic analysis failed to reflect these differences. Both patient groups could be distinguished according to acoustic and temporal measures when compared with normal controls. The acoustic analyses were more sufficient in voices without severe dysfunction.


For patients with laryngeal carcinoma, the voice is affected by the disease and treatment. In terms of cure, the criterion standard for advanced laryngeal cancer (stages III and IV) has previously been laryngectomy (LE). In recent years, advances have been made in the development of organ-preserving therapy. However, the best primary treatment is still under debate, and different studies show contradictory results when comparing LE vs radical radiotherapy.1-3

The quality of voice in patients with advanced laryngeal cancer treated with radiotherapy is generally considered to be better than that after surgery. Retention of the larynx is the great advantage of radiotherapy in the treatment of laryngeal cancer, yet it remains unknown to what extent the vocal function is preserved or lost during the course of treatment. Voice-related problems after treatment indicate the need for voice rehabilitation not only for the patient who has lost the ability to communicate vocally after LE, but also for patients treated with radiotherapy.4,7

To our knowledge, there are no studies comparing patients who have had LE with patients treated with radiotherapy objectively and perceptually according to voice and speech function. The advances realized in different computerized voice analysis programs allow clinicians to more routinely analyze voice variables. Before application of the chosen variables to a large population, it is important that objective measures be validated against subjective assessments in small groups.4 Previous studies comparing perceptual and
PATIENTS AND METHODS

PATIENTS

Twelve men with tracheoesophageal speech as their primary mode of communication participated in the study. The primary treatment for all 12 patients was radical radiotherapy, but these 12 patients later underwent LE as salvage surgery. All patients in the western region of Sweden who underwent LE during a 6-year period (January 1, 1989, through December 31, 1994) were reviewed. During this period, 50 patients underwent LE, 40 of whom received a tracheoesophageal prosthesis (TEP). When our study started, all of the LE-treated patients who were alive (n = 26), who lived within a specified area (n = 20), and who had laryngeal cancer (n = 18) were contacted. Two patients declined to participate because of illness, and 2 patients no longer used their TEP. Two of the patients were women, and to eliminate the variability of acoustic variables between sexes, only male patients were studied. All of the LE-treated patients received at least 20 hours of voice training by a speech pathologist. None of the patients used esophageal or electrolaryngeal speech regularly.

CONTROL GROUPS

The cancer control group consisted of 12 men with laryngeal carcinoma, treated with radical radiotherapy, and with preserved larynxes. The patients were identified from the clinical records, and the inclusion criteria were, as far as possible, comparable age, tumor site, and the original TNM classification of the LE-treated patients (ie, before radiotherapy treatment). All patients contacted agreed to participate and were included in the study. The patients treated with radical radiotherapy did not receive any voice training. The last treatment had been completed at least 6 months for the LE-treated patients and 11 months for the radiotherapy-treated patients before the voice recordings were made (range, 6-72 and 11-72 months, respectively). None of the patients had active laryngeal disease.

The normal control group consisted of 10 normal speakers who were recruited as they came to the Sahlgrenska University Hospital, Göteborg, Sweden, accompanying patients visiting the hospital. The normal control group was matched for sex and age. These speakers had no history of cancer or vocal abnormality (Table 1).

SMOKING HABITS

One of the patients in the LE group claimed to be a non-smoker, 5 were smokers, and 6 were ex-smokers. All of the patients treated with radical radiotherapy had been smokers, and 8 were ex-smokers. In the normal control group, 3 subjects were nonsmokers, 3 were smokers, and 2 were ex-smokers.

TREATMENT

The 24 patients with laryngeal carcinoma had received radical radiotherapy as primary treatment. Patients with T1 disease received conventionally fractionated radiation therapy to a biologically equivalent dose of 62 to 68 Gy with only the larynx as the target volume. It was given as a 2-Gy/d fraction 5 days per week or as a 2.4-Gy/d fraction 4 times per week. Patients with T2 to T4 disease received either hyperfractioned accelerated radiation therapy, 2 fractions of 1.7 Gy/d with a total dose of 64.6 Gy in 4.5 to 5 weeks, or conventionally fractionated radiation therapy with a total dose of 62 to 67.2 Gy. In patients with T3 or T4 disease, the regional nodes at risk, ie, nodes suspected to harbor occult disease, were included in the target volume. The regional nodes were irradiated with a total dose of 40.8 to 50 Gy and with the same fractionation regimen as for the primary tumor.

One patient in the LE group and 2 patients treated with radiotherapy with T4 disease had initial chemotherapy, 2 cycles of cisplatin in combination with 5-fluorouracil. None of the LE-treated patients required removal or alteration of any oral structures during surgery.

PROCEDURE

All subjects came to Sahlgrenska University Hospital once, and voice samples were recorded. For the patients, clinical data and the present larynx status were examined, and for the normal controls, smoking habits were registered.

All subjects underwent recording under similar conditions, comfortably seated in a soundproof booth. They were asked to perform the following tasks: to read aloud a standard story of 89 words at a comfortable level of speech and to achieve the maximum duration of /a/ on 1 exhalation. Each subject was asked to sustain the vowel /a/ at a

acoustic variables for patients with different treatments for laryngeal carcinoma often showed disparities in the clinical materials and methods used, which prevents a fair comparison. Although there are difficulties in assessing objective voice variables in patients with laryngeal carcinoma, there is a need for voice quality quantification, and objective variables could be a part of the voice rehabilitation follow-up. In this study, we wanted to compare LE-treated patients with 2 control groups (patients treated with radiotherapy for laryngeal carcinoma and normal controls), assess different acoustic variables, and correlate them with the subjective assessment of speech intelligibility, voice quality, and speech acceptability. It was hypothesized that the 3 subject groups would show acoustic differences.

RESULTS

ACOUSTIC ANALYSIS

Fundamental Frequency

The patients treated with LE showed the lowest F0 values (mean, 105.2 Hz, and median, 98.5 Hz for running speech; and mean, 103.1 Hz, and median, 102.3 Hz for sustained /a/). All subject groups had wide F0 ranges within the groups, with patients treated with radiotherapy demonstrating the largest range (Figure 1). No statistical significance was found between the different groups.
There were no significant differences in the median F0 between running speech and sustained vowel /a/ independent of treatment group.

Absolute F0 Perturbation

For running speech, the highest values for mean (2.6%) and median (2.5%) were found in the LE group, followed by the radiotherapy group (mean, 2.2%; median, 1.5%), and finally normal controls (mean, 1.3%; median, 1.1%). There was great variability in perturbation values of sustained /a/ within the LE as well as the radiotherapy group. No statistical significance was found between the LE and radiotherapy patient groups. There was, however, a significant difference between the normal controls, who had lower values, and both the LE-treated patients and the radiotherapy-treated patients for running speech as well as for the sustained vowel /a/ (P<.001 for LE and P<.05 for radiotherapy) (Figure 1).

TEMPORAL ANALYSIS

Maximum Phonation Time

The mean and median values for maximum phonation time were 14.7 and 10.8 seconds in the LE group and 15.6 and 10.8 seconds in the radiotherapy group. The normal controls had the highest mean and median values (21.7 and 20.7 seconds, respectively). The LE-treated patients and the radiotherapy-treated patients

ACOUSTIC ANALYSIS

The recordings were sampled (rate, 16 kHz) on a 133-MHz personal computer and analyzed with the speech analysis software Soundswell to determine the following acoustic variables: (1) fundamental frequency (F0), (2) absolute F0 perturbation, (3) speech rate, and (4) maximum phonation time. These variables were chosen because we wanted to study a few, easily analyzed variables. The acoustic variables (F0 and absolute F0 perturbation) were calculated from the second sentence of the recorded text and from a 2.5- to 3-second-long portion, spliced from the middle part of the sustained vowel /a/.

The F0 extraction of the Soundswell program is based on the double-peak-picking algorithm described by Dolan-sky in 1955 and modified by Ternström. The input signals were first low-pass filtered at 250 Hz and high-pass filtered at 50 Hz. The filtered signal was fed to a peak-tracking circuit, which identified the dominant positive and negative peaks and computed the period time. For each pitch period, the period time was selected that was closest to its own running median value during the 7 most recent periods. This period time was inverted and expressed as the raw F0 contour. For perturbation measures, the raw F0 contour was low-pass filtered at 20 Hz, with a third-order Butterworth filter, which gives a smoothed version of the raw F0 contour (the F0 trend). The perturbation output signal is the relative percentage difference between the F0 trend and the raw F0 contour. In this study we analyzed the mean of the rectified perturbation signal, the absolute perturbation.

As previously described, the aperiodicity in especially the alaryngeal voices makes them difficult to test acoustically, and, in our study, 1 patient from the LE group was excluded from the acoustic evaluation because the presence of aperiodic events throughout his signal prevented extraction of an F0. We validated the Soundswell F0 extraction by comparison with 3 other methods: linear spectrum analysis, manual extraction of the F0 in the voice signal, and the perceptual evaluation of 2 professional listeners. Pearson correlation coefficient was 0.89, 0.91, and 0.81 (vowel /a/) and 0.76, 0.94, and 0.95 (story) for linear spectrum analysis, manual extraction of F0, and perceptual evaluation, respectively. The comparison showed a significant correlation (P<.001) between the F0 Soundswell extraction and the 3 methods.

PERCEPTUAL ANALYSIS

The listeners' panel judged the 2 patient groups treated for laryngeal carcinoma. The listeners' panel in this investigation included 10 inexperienced listeners (unfamiliar with tracheoesophageal or irradiated laryngeal speech) and 5 experienced listeners (practicing speech-language pathologists). The listeners rated the voices on 3 perceptual items: speech intelligibility, voice quality, and speech acceptability. The reliability of listeners' evaluation considered a sufficiently high intrajudge (77%-95%) and interjudge (69%-85%) reliability.

The 2 patient groups and the normal controls judged their own voices according to the perceptual items: speech intelligibility, voice quality, and speech acceptability.

The subjects and listeners were instructed to mark a vertical line on a 100-mm visual analog scale at the point of the scale that most closely resembled their answer. The scale was anchored at each end by opposite extremes of the attribute. The higher the score, the better the voice.

The study was approved by the local ethical committee on November 9, 1994.

STATISTICAL METHODS

For descriptive purposes, we used mean and median values and box-plot graphs, displaying the 10th, 25th, 75th, and 90th percentiles of the variables. For correlation analysis, Pitman nonparametric permutation test was used. In addition, Pearson correlation coefficient was calculated for descriptive purposes. For comparison between 2 groups, the Fisher nonparametric permutation test was used. Significance was determined at the P<.05 level.
showed more variability within the groups compared with the normal controls (Figure 2).

### Speech Rate

Great individual variability was found for speech rate within the LE group, the radiotherapy group, and the normal controls, and only small nonsignificant differences were found between the different groups (Figure 2). The total number of breathing pauses was about the same for all groups (LE: mean, 7.9; median, 8.5; radiotherapy: mean, 8.2; median, 9.0; and normal controls: mean, 7.6; median, 7.5) but the LE-treated patients required a longer inhalation pause time (mean, 9.7 seconds; median, 8.7 seconds) than the radiotherapy-treated patients (mean, 8.2 seconds; median, 7.5 seconds) and normal controls (mean, 7.4 seconds; median, 7.2 seconds).

### PERCEPTUAL EVALUATION

The listeners’ perceptual evaluations and the subjects’ self-assessment are summarized in Figure 3. Some of these data have been described previously.16 The LE-treated patients’ self-assessment of voice quality and speech acceptability deviated most from the listeners’ evaluation. In the LE-treated patients’ self-assessment, the result for speech intelligibility was significantly lower \( (P<.05) \) when compared with both the radiotherapy-treated patients and normal controls. The result was also significantly lower when the LE-treated patients were compared with normal controls in voice quality \( (P<.01) \) and speech acceptability \( (P<.05) \).

In listeners’ perceptual evaluations, the listeners judged the LE-treated patients significantly lower than the radiotherapy-treated patients in speech intelligibility \( (P<.01) \), voice quality, and speech acceptability \( (P<.001) \).
CORRELATIONS

The correlation matrix for all variables is shown in Table 2. In the comparison between perceptual evaluation and acoustic measures, there was a weak, but significant, correlation for the absolute F0 perturbation regarding both patients’ and listeners’ evaluation. The correlation between absolute F0 perturbation and the item “speech intelligibility” was lower than that between “voice quality” and “speech acceptability.” The intercorrelation between the listeners’ perceptual evaluation of “voice quality” and F0 perturbation for sustained vowel /a/ is shown in Figure 4. The correlation coefficient was 0.59, and 80% of the patients who were judged by the listeners to have insufficient voice quality (<65%) were captured by the perturbation /a/ test.

For listeners, the maximum phonation time and speech rate also correlated significantly with the perceptual evaluation, and the correlation between speech rate and speech intelligibility was 0.60 (P < .01).

For F0, the correlation coefficient of running speech vs sustained vowel /a/ was 0.59 (P < .01). The strongest correlations for acoustic variables were shown for F0 absolute perturbation /a/ vs story 0.79 (P < .001).

The most important finding in this study was that no significant differences were found between LE-treated and irradiated patients in acoustic and temporal measures, whereas significant differences were found in the perceptual evaluation. The LE-treated and radiotherapy-treated patients could be distinguished when compared with normal controls. The acoustic variable absolute F0 perturbation was found to be the factor showing the best correlation with the other variables, and there was a weak but significant correlation between absolute F0 perturbation and perceived voice quality.

Table 2. Correlation Matrix and Statistical Significance for Different Variables*

<table>
<thead>
<tr>
<th></th>
<th>F0(s)</th>
<th>F0(a)</th>
<th>PF0(s)</th>
<th>PF0(a)</th>
<th>/a/</th>
<th>w/m</th>
<th>SI(p)</th>
<th>VQ(p)</th>
<th>SA(p)</th>
<th>SI(l)</th>
<th>VQ(l)</th>
<th>SA(l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0(s)</td>
<td>0.59*</td>
<td>0.22</td>
<td>0.20</td>
<td>-0.26</td>
<td>-0.19</td>
<td>-0.05</td>
<td>-0.17</td>
<td>-0.12</td>
<td>0.15</td>
<td>-0.01</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>F0(a)</td>
<td>0.10</td>
<td>0.17</td>
<td>0.17</td>
<td>-0.22</td>
<td>-0.00</td>
<td>-0.04</td>
<td>-0.17</td>
<td>-0.06</td>
<td>0.34</td>
<td>0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF0(s)</td>
<td>0.79‡</td>
<td>-0.34§</td>
<td>-0.08</td>
<td>-0.24</td>
<td>-0.45§</td>
<td>-0.36§</td>
<td>-0.28</td>
<td>-0.55†</td>
<td>-0.48§</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF0(a)</td>
<td>-0.32§</td>
<td>0.49§</td>
<td>0.28</td>
<td>0.32</td>
<td>0.26</td>
<td>0.48§</td>
<td>0.42§</td>
<td>0.44§</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a/</td>
<td>0.49§</td>
<td>0.33</td>
<td>0.20</td>
<td>0.24</td>
<td>0.60‡</td>
<td>0.42§</td>
<td>0.51§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/m</td>
<td>0.77‡</td>
<td>0.88‡</td>
<td>0.58‡</td>
<td>0.58‡</td>
<td>0.58‡</td>
<td>0.58‡</td>
<td>0.58‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI(p)</td>
<td></td>
<td>0.87‡</td>
<td>0.48‡</td>
<td>0.50‡</td>
<td>0.51‡</td>
<td>0.51‡</td>
<td>0.51‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VQ(p)</td>
<td></td>
<td></td>
<td>0.82‡</td>
<td>0.88‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SA(p)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI(l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VQ(l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SA(l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* F0(s) indicates mean fundamental frequency for running speech; F0(a), mean fundamental frequency for sustained vowel /a/; PF0(s), absolute F0 perturbation for running speech; PF0(a), absolute F0 perturbation for sustained vowel /a/; /a/, maximum phonation time in seconds; w/m, speech rate in words per minute; SI(p), VQ(p), and SA(p), patients’ assessment for speech intelligibility, voice quality, and speech acceptability, respectively; and SI(l), VQ(l), and SA(l), listeners’ assessment for speech intelligibility, voice quality, and speech acceptability, respectively. The perceptual evaluation of the listeners does not include normal controls. The correlations with F0 and PF0 include 11 patients with laryngectomy, 12 with radiotherapy, and 10 normal controls. The other variables include all subjects.

†P < .01.
‡P < .001.
§P < .05.
The assumption that the results from 3 groups (LE, radiotherapy, and normal controls) would show acoustic differences was confirmed in that the results generally showed a trend for normal controls to have the best results, followed by the radiotherapy group and finally the LE group. The differences between the LE and radiotherapy groups were, however, nonsignificant. These findings are in agreement with previous observations that acoustic perturbation measures can differentiate patients with laryngeal cancer from normal controls. In our study, however, the acoustic and temporal measures failed to confirm the significant perceptual differences in voice and speech quality between LE-treated and radiotherapy-treated patients. This might be explained by the small number of patients and variables analyzed, and the possibility of an insensitive method.

Another consideration is the difference in voice therapy, where the patients treated with radiotherapy might have received better voice quality if they had been given the possibility of voice therapy after treatment.

The advantage of perceptual ratings and the disadvantage of acoustic measures in severe voice abnormality were described by Rabinov et al. They compared the reliability of perceptual ratings of roughness with acoustic measures of F0 perturbation produced by several voice analysis systems. The study of Rabinov et al showed that, overall, listeners agreed as well as or better than objective algorithms. Listeners disagreed in predictable ways, whereas automatic algorithms differed in a seemingly random fashion. Finally, the reliability of the listeners increased with the severity of voice abnormality, whereas the objective methods were more sufficient in normal voices. The authors concluded that acoustic measures may have advantages over perceptual measures for discriminating among essentially normal voices. For clinical purposes, however, perceptual measures are probably superior, at least for perturbation measures, to current acoustic analysis systems. Despite this fact, we believe acoustic measures to have a value in research settings, for validating perceptual evaluation, and in longitudinal rehabilitation and treatment evaluation. To support and complement perceptual evaluation of voice quality and communication after treatment for laryngeal carcinoma, objective variables could be used.

The temporal results for tracheoesophageal fistula gives the possibility of lung-powered speech, the LE-treated patients still require a larger lung volume to power TEP speech. This is because of higher resistance in the pharyngoesophageal segment, which requires a higher pressure and/or higher airflow. To what degree aerodynamic variables might influence TEP speech in voice and speech quality is not well known, but a higher pressure is needed for TEP speech. Woodson et al reported irradiated patients to have higher laryngeal resistance, resulting in complaints about voice fatigue and impaired loudness of the voice. When rating their voice quality, the irradiated patients considered not only the sound of the voice but also the effort required to speak. Aerodynamic measures are not used as commonly as acoustic measures, because data acquisition is more complex, and fewer commercial systems are available. However, measurements of airflow and air pressure could probably provide useful clinical information and will be considered as an option in future investigations of this population.
The incidence of laryngeal cancer in Sweden (world standardized rate) is 2.0 cases for males and 0.2 cases for females annually per 100,000 inhabitants. These patients may have their most vital and unique human functions affected, which could result in considerable communication problems. Restoration of effective communication possibilities after LE, primarily with a TEP, should be a major priority in the rehabilitation of the LE-treated patients, but the need for voice rehabilitation in patients treated with radiotherapy should not be forgotten.

CONCLUSIONS

The significant differences found between LE- and radiotherapy-treated patients in the perceptual evaluation could not be confirmed in the acoustic and temporal measures used in the study. Both LE- and radiotherapy-treated patients could be distinguished when compared with normal controls. Objective variables could support and complement perceptual evaluation of voice and speech quality after treatment for laryngeal carcinoma.

Accepted for publication September 8, 1998.

The study was made possible by grants from the Assar Gabrielsson Foundation, the King Gustav V Jubilee Clinic Cancer Research Foundation, the Hjalmar Svensson Foundation, The Sahlgrenska Hospital Foundations, the Göteborg Medical Society, and the Medical Faculty, Göteborg University, Göteborg, and the Rosa and Emanuel Nachmanssons Foundation, the Aximans Foundation, the Olympus Optical AB Foundation, and the Swedish Medical Society, Stockholm, Sweden.

We thank Lennart Nord for his assistance with the acoustic analyses, Henrik Ahlbom for statistical advice, and Stellan Hertegård for expert advice.

Reprints: Caterina Finizia, MD, PhD, Department of Otorhinolaryngology, Head and Neck Surgery, Sahlgrenska University Hospital, S-413 45 Göteborg, Sweden (e-mail: Caterina.Finizia@orlss.gu.se).

REFERENCES