Aerodynamic Findings in Esophageal Voice

Sergio Motta, MD; Iacopo Galli, MD; Lino Di Rienzo, MD

Objective: To define the perceptive and aerodynamic characteristics of esophageal voice in relation to different rehabilitation modalities.

Design: Cross-sectional study comparing perceptive and aerodynamic variables in 3 subject groups.

Setting: Referral center.

Subjects: A total of 19 subjects who underwent total laryngectomy were divided into groups A and B. Group A consisted of 13 subjects (who required speech therapy)—8 good speakers (subset A1 who were >80% intelligible) and 5 mediocre speakers (subset A2 who were <70% intelligible). Group B consisted of 6 subjects with a tracheoesophageal prosthesis (who were >90% intelligible).

Main Outcome Measures: Perceptive variables included phonatory pauses and stomal noise. Aerodynamic variables included maximum phonation time, phonatory flow, phonatory volume, postphonatory volume, intensity, and articulatory pressure.

Results: Phonatory pauses and stomal noise statistically differentiated group A from group B and good speakers from mediocre speakers. Phonation time, phonatory volume, and phonatory flow were statistically higher in group B subjects compared with group A subjects. Postphonatory volume was significantly higher in group A. Intraoral pressure and postphonatory volume were statistically higher in subset A2 subjects compared with subset A1 subjects while maximum phonation time was significantly higher in subset A1 subjects compared with subset A2 subjects.

Conclusions: In subset A1 subjects a positive ratio between phonatory volume and phonatory flow was maintained with an adequate phonation time. In subset A2 subjects a reduced phonatory volume was associated with a more rapid dispersion of phonatory flow, lower duration of phonation, and frequent pauses; stomal noise and consonant hyperarticulation worsened the voice performance in this group. In group B subjects the positive ratio between phonatory volume and phonatory flow represented the prerequisite of speech without frequent pauses.


Total Laryngectomy represents an operation that drastically affects respiratory dynamics and phonation mechanisms, suppressing the normal verbal communication. The acquisition of a new voice by the laryngectomized subject is possible with the use of an electrolarynx, conventional speech therapy, and surgical prosthetic methods. Use of an electrolarynx is exclusively reserved for patients who did not profit from conventional speech therapy or in whom a tracheoesophageal prosthesis (TEP) cannot be applied. Speech therapy allows the acquisition of autonomously produced esophageal voice (EV) and, therefore, it is the most commonly used treatment in voice rehabilitation of laryngectomized patients. Surgical prosthetic methods, introduced in 1980 by Singer and Blom, tended to become rapidly widespread based on the excellent outcomes that they achieved. However, TEP is still an underused rehabilitative procedure; in fact, motivational and socioeconomic factors, as well as contraindications associated to pathologic manifestations of involved organs, limit its application.

Intelligibility of EV can vary according to several perceptive factors on the precise definition for which there is no general agreement. Furthermore, aerodynamic data in the study of EV physiology and, in particular, correlations between those data and the perceptive findings have not been defined as yet. The purpose of our study was to identify: the perceptive factors that affect EV intelligibility; EV aerodynamic characteristics; and the rehabilitative procedures that are more effective in achieving adequate EV.
SUBJECTS AND METHODS

SUBJECTS

The study subjects came consecutively to the Institute of Otorhinolaryngology, Ateneo “Federico II,” University of Naples, Naples, Italy, between January 1, 1990, and December 31, 1997. Twenty-two Italian-speaking male patients (mean age, 52 years) undergoing total laryngectomy were enrolled in this study; 16 received speech therapy for the acquisition of EV (group A) while 6 received a TEP (group B). Excluded were subjects who at the intelligibility test achieved word intelligibility percentages lower than 50% (group A, 1 subject); and subjects shown to be unable to undergo one or more examinations of the study protocol (group A, 2 subjects). Overall 19 subjects were examined of whom 13 were in group A and 6 were in group B. The application of TEP (Bivona-Colorado Low-Resistance, Bivona Medical Technologies, Gary, Ind) was performed according to the surgical procedure described by Motta.6

METHODS

At the end of rehabilitation (mean duration, 4 months for group A and 1 month for group B) each subject underwent the following: (1) an intelligibility test, (2) an acoustic-perceptive study, and (3) an aerodynamic study.

In the intelligibility test a panel of 6 members participated, drawn for each subject from a pool of 60 students enrolled in the university course Logopedics. Each subject, placed 3 m from the observers, was invited to produce 20 phonetically balanced words, of which 10 were bisyllabic and 10 were trisyllabic. During the test the observers turned their backs to the subject to rule out a possible interference of lateral reading on the results of the test and to prevent the identification of subjects with a TEP. Based on the percentage of word intelligibility, group A subjects were subdivided into the following 2 subsets: subset A1, composed of 8 subjects defined as good speakers, for whom intelligibility was shown to be higher than 80% (mean intelligibility, 97%); subset A2, composed of 5 subjects, considered mediocre speakers for whom intelligibility was between 50% and 70% (mean intelligibility, 64%). For group B subjects the percentage of word intelligibility was shown to be higher than 90% (mean intelligibility, 99%). The statistical analysis performed using the Mann-Whitney test showed the following: intelligibility values were statistically not different in the comparison between group A and group B subjects (P=.15); statistically different values in the comparison between subset A1 and subset A2 subjects (P<.002); and statistically not different values between subset A1 and group B subjects (P=.85); and statistically different values between subset A2 and group B subjects (P<.001).

For the acoustic-perceptive study, performed with the same modalities used for the intelligibility test, each subject was required to produce a series of progressive numbers from 1 to 10; the panel had to report the number of pauses perceived and the evaluation of stomal noise, score based on a 3-grade scale (grade 1, no stomal noise; grade 2, discontinuous stomal noise; and grade 3, continuous stomal noise).

The aerodynamic study was carried out by a computerized system with physical flow calibration and automatic calibration of air pressure and sound intensity, equipped with a facial mask connected with 3 transducers (Aerophone II; F. J. Electronics, Copenhagen, Denmark). This examination allowed the following observations: (1) During vowel /e/ emission the maximum phonation time was expressed in seconds and tenths of a second; mean phonatory flow, expressed in milliliters per second, indicating the ratio between phonatory volume and maximum phonation time; phonatory volume, reported in milliliters, equal to the volume of air expelled to produce the voice; residual postphonatory volume, expressed in milliliters, which represents the volume of air expelled not being sonorized or expelled with sound intensity levels not recordable by the instrumentation (<30-dB sound pressure level); and sound intensity expressed in decibles of sound pressure level. (2) During the production of the phoneme pi-pi repeated 3 times, using a small silicon tube placed into the oral cavity for approximately 5 cm, the mean articulatory pressure, expressed the level of intraoral pressure in cm H2O during the articulation of the phoneme. (3) The aerodynamic study was carried out by one of us (S.M.) who was unaware of the results of the intelligibility test but not of the interpretation of the results. The statistical comparison was performed using the Mann-Whitney test (SPSS 8.0; SPSS, Chicago, Ill). P<.05 was accepted as the minimal level of statistical significance.

dynamic characteristics, in relation to the different rehabilitation methods used; and the possible correlations between the findings of the perceptive study of EV and those of the aerodynamic study. Furthermore, our study intends to contribute to the knowledge of the pathophysiological mechanisms underlying the EV to improve the rehabilitation techniques used at present.

RESULTS

The results of this study are summarized in the Table. In the comparison between group A and group B, in subjects undergoing TEP rehabilitation (group B) there was a lower number of pauses (P<.01); a more reduced stomal noise (P<.001); higher values of maximum phonation time (P<.001), phonatory flow (P<.02), and phonatory volume (P<.02); and a lower residual postphonatory volume (P<.02). Statistically similar values were observed for sound intensity (P=.89) and intraoral articulatory pressure (P=.46). In the comparison between subset A1 and subset A2, in subjects with good conventional EV (subset A1), there was a lower number of pauses (P<.0001), a more reduced stomal noise (P<.01), a longer maximum phonation time (P<.01), lower values of articulatory pressure (P<.03), and residual postphonatory volume (P<.01). Phonatory flow was higher in subset A2 subjects but without statistical significance (P=.12); phonatory volume was shown to be higher in subset A2 subjects but without statistical significance (P=.17); sound intensity was higher, but without statistical significance, in subset A2 subjects (P=.52). In the comparison between subset A1 and group B sub-
The sound emission produced by subjects with EV acquired by conventional speech rehabilitation or the application of a TEP was previously studied by perceptive and spectrographic analysis to assess the intelligibility of EV and its acoustic characteristics.11-14 However, studies on air phenomena underlying the EV and conditioning its efficacy are lacking. Max et al10 suggested that the EV factors they studied (frequency, intensity, and maximum duration of phonation) were correlated with different modalities and ill-defined aerodynamic mechanisms. Ill-defined aerodynamic mechanisms would vary according to the modality of EV production (whether conventional or secondary to TEP). However, to our knowledge, there are no reports in the literature to support this hypothesis even if there are interesting observations by Nieboer and Schutte15; these authors studied the relation between the intratracheal and esophageal pressure in patients with a TEP to define the resistance to the air flow by the prosthesis, finding a mean (SD) phonatory flow higher in patients with a TEP (140 [40] mL/s) compared with conventional subjects with EV (70 [40] mL/s). The EV pathophysiology has been analyzed in-depth based on perceptive and aerodynamic studies.

**INTELLEGIBILITY OF EV AND PERCEPTIVE STUDY**

Various factors drawn from the acoustic findings of some listeners have been considered in the literature to evaluate the quality of EV.11,12,16,17 First is the percentage of words correctly articulated by patients and, therefore, recognized by listeners. Literature data documentation that EV can achieve satisfactory intelligibility independent of the rehabilitation modality.3,11 Our observations show that the conventional EV can also afford satisfactory results with percentages of correctly identified words comparable to those observed in subjects with a TEP.

Second, speech fluency can be studied based on the number of words articulated per minute or the num-

---

**Perceptive and Aerodynamic Findings in 19 Study Subjects**

<table>
<thead>
<tr>
<th>Group*</th>
<th>Case No.</th>
<th>Words, %</th>
<th>Pause, No.</th>
<th>Stomal Noise, Grade†</th>
<th>Sound Intensity, dB SPL‡</th>
<th>Phonatory Volume, mL</th>
<th>Phonatory Flow, mL/s</th>
<th>Phonation Time, s</th>
<th>Residual Volume, mL</th>
<th>Articulatory Pressure, cm H2O</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1</td>
<td>100</td>
<td>0.0</td>
<td>1.2</td>
<td>72</td>
<td>48</td>
<td>40</td>
<td>1.2</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>100</td>
<td>0.2</td>
<td>2.3</td>
<td>76</td>
<td>22</td>
<td>21</td>
<td>1.0</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>100</td>
<td>0.0</td>
<td>1.0</td>
<td>69</td>
<td>70</td>
<td>47</td>
<td>1.5</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>97</td>
<td>0.5</td>
<td>1.2</td>
<td>79</td>
<td>99</td>
<td>58</td>
<td>1.7</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>98</td>
<td>0.0</td>
<td>1.2</td>
<td>69</td>
<td>60</td>
<td>44</td>
<td>1.3</td>
<td>7</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>100</td>
<td>0.0</td>
<td>1.0</td>
<td>74</td>
<td>583</td>
<td>153</td>
<td>3.8</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>83</td>
<td>0.7</td>
<td>1.3</td>
<td>76</td>
<td>125</td>
<td>88</td>
<td>1.4</td>
<td>5</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>100</td>
<td>0.5</td>
<td>2.3</td>
<td>72</td>
<td>39</td>
<td>17</td>
<td>1.3</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>97 ± 6</td>
<td>0.2 ± 0.4</td>
<td>1.5 ± 0.6</td>
<td>73 ± 3</td>
<td>131 ± 186</td>
<td>58 ± 44</td>
<td>1.6 ± 0.9</td>
<td>3.4 ± 3.8</td>
<td>40 ± 5</td>
<td></td>
</tr>
</tbody>
</table>

**COMMENT**

---

©2001 American Medical Association. All rights reserved.
ber of syllables emitted per each air emission\textsuperscript{3,9,18} or, as in our study, by the number of pauses perceived during a predefined vocal test. We preferred the latter method because it directly shows the relationship between the EV quality and fluency alterations. Kischk and Gross\textsuperscript{9} have studied comparatively the articulation velocity in laryngectomized patients rehabilitated by conventional speech therapy, a TEP application, or with an electrolarynx but did not find statistically significant differences between the study groups.

Our findings confirm that the absence or the reduced number of pauses leads to a good understanding of EV; this is documented by the outcomes in laryngectomized subjects with conventional EV who were good speakers (subset A\textsubscript{1}) and in those with a TEP (group B).

To our knowledge, there are no reports in the literature showing which alterations in the study factor could occur in laryngectomized subjects with mediocre or poor EV. In this regard our observations clearly showed that in these cases (subset A\textsubscript{2}), the number of pauses was markedly higher (Table) and, therefore, the reduced intelligibility was also consequent to an inadequate speech fluency.

Third, the presence of parasite stomal noise (this phenomenon has a different genesis). In patients with conventional EV the stomal noise, usually higher in mediocre speakers, is caused by an incorrect coordination of phonorespiratory mechanisms, indicating a partial rehabilitation failure. In subjects with a TEP it is attributable to an imperfect adhesion of the prosthesis to the stoma during the vocal emission or, when a tracheostomal valve is used, to its unsatisfactory functioning—these are casual, mechanical, readily identifiable and, therefore, eliminating factors.

**AERODYNAMIC STUDIES AND PATHOPHYSIOLOGY OF EV**

Our study showed that the aerodynamic procedures play a major role in the definition of the factors involved in the genesis of EV as well as in the identification of some important functional patterns able to interfere with the vocal performance of patients and also to hypothesize possible “compensation” phenomena which may deteriorate the speaking proficiency. We found similar values of sound intensity in the different study groups, in agreement with the observations by Kischk and Gross.\textsuperscript{9} This finding shows that there are no differences in the performance of the pseudoglottis as the source of vibration with reference to the esophageal or pulmonary air supply, as confirmed by the study by Isman and O’Brien\textsuperscript{18} that ruled out significant differences in the structure of the pseudoglottis examined by videofluoroscopy in subjects able to speak with both types of voice. The aerodynamic tests have documented that fluency depends on the volume of phonatory air, its flow during the vocal emission and on the ratio between these 2 factors.

In subjects with EV produced by conventional rehabilitation, the air supply was drastically conditioned by the reduced capacity of esophageal storage. In subjects with a satisfactory EV, it was observed that the limited esophageal air volume corresponds to a reduced flow and, therefore, a positive ratio between the 2 factors and, in turn, a good maximum duration of phonation was maintained. In subjects with mediocre intelligibility, there was a reduced phonatory volume and a more rapid (relatively high values of phonatory flow), less effective dispersion of phonatory air (presence of a proportionally high residual postphonation volume) involving a negative phonatory volume–flow ratio and, consequently, a lower duration of phonation with frequent interruptions. The increased articulatory pressure, usually higher in mediocre speakers, should be correlated with the need for these subjects to increase the intraoral pressure during the production of explosive consonants, not sustained by an adequate aerodynamic flow. This consonant hyperarticulation while it tended to compensate for the reduced effectiveness of the esophageal air, further impaired the quality of the voice.

In subjects with TEP the relevant amount of air provided by the pulmonary bellows, its regular flow that followed the normal pneumophonic coordination, and the positive relationship between phonatory volume and phonatory flow represented the prerequisite of a satisfactory EV produced without the interference of frequent pauses.

**CONCLUSIONS**

In this study we have examined laryngectomized subjects undergoing voice rehabilitation with conventional speech therapy and the use of a TEP, able to produce more or less satisfactory speech but, in any case, showing an adequate function of pseudoglottic sphincter. In subjects with conventional rehabilitation who are good speakers, despite the reduced phonatory volume, there was a positive volume–flow ratio because the subjects seemed to be able to maintain a reduced flow and, therefore, air dispersion could be better controlled.

In conventional rehabilitation subjects who are mediocre speakers, the air volume used for phonation was lower while its flow was proportionally high with a consequent negative volume–flow ratio, worsened by the presence of a relatively high residual postphonation volume. As a consequence of the reduced maximum phonation time, the number of pauses increased and the intelligibility was severely impaired in this group. Furthermore, there could be uncorrected articulatory compensation (with increased intraoral pressure) that, in turn, contributed to the worsening of EV intelligibility.

In subjects with a TEP, all excellent speakers, both phonatory air volume and phonatory airflow were relevant; the positive ratio between phonatory volume and phonatory flow represented the prerequisite of a satisfactory EV produced without the interference of frequent pauses.

Accepted for publication February 6, 2001.

Corresponding author and reprints: Sergio Motta, MD, Clinica Otorinolaringoiatrica, Universita Cattolica del Sacro Cuore, Largo A. Gemelli, 00168 Rome, Italy (e-mail: sermotta@Yahoo.it).
REFERENCES

5. Seinsch W. The voice prosthesis: an underutilized procedure in rehabilitation [in German]? HNO. 1993;41:595-596.