Relation between Videofluoroscopy of the Esophagus and the Quality of Esophageal Speech

Marina Lang Fouquet a, b Antônio José Gonçalves a Mara Behlau b, c

a Head and Neck Surgery Discipline, Santa Casa de Misericórdia de São Paulo Hospital, b Faculdade de Medicina, Universidade de São Paulo, and c Centro de Estudos da Voz – CEV, São Paulo, Brazil

Introduction

Total laryngectomy may greatly impact patient quality of life, primarily due to the loss of laryngeal voice and the associated communication deficits. Esophageal voice (EV), in Brazil, remains the primary rehabilitation technique to facilitate communication after total laryngectomy due to both social and economic constraints. However, the percentage of successful acquisition of EV varies greatly, ranging from 32 to 63% [1–5]. Several variables, including age and socioemotional status, as well as anatomical and functional factors have been linked to successful EV acquisition.

Several authors have described patient age as the primary factor associated with successful EV acquisition, suggesting that the complex processes associated with EV acquisition are highly dependent upon a patient’s capacity to learn new tasks, which diminishes with advancing age [4, 5]. Socioemotional factors including depression, loss of identity, lack of practice, inadequate instruction or lack of opportunities for communication [6] may also influence successful acquisition of EV. Finally, anatomical variables are highly correlated to EV acquisition including stenosis (occasionally referred to as difficulty swallowing solids), pseudovallecula [6, 7], cricopharyngeal bar (bulging of the posterior wall of the pharynx into the lumen of the hypopharynx), gastroesophageal reflux, hiatal hernia [6], dysphagia, small tracheostoma, perma-

Key Words
Evaluation · Fluoroscopy · Voice quality · Esophageal speech · Rehabilitation

Abstract

The goal of the current study was to compare the quality of esophageal speech and voice to videofluoroscopic features of the esophagus and pharyngoesophageal (PE) segment. The speech and voice characteristics of 30 laryngectomized patients were rated by 5 speech-language pathologists. Based on these ratings, patients were divided into 3 categories: fluent (n = 9), moderately fluent (n = 10) and nonfluent (n = 11). Videofluoroscopy of the PE region was then performed during both swallowing and voice production. An insufflation test and percutaneous pharyngeal plexus block were required in 9 patients to determine the etiology of poor esophageal voice production. The strongest videofluoroscopic indicators of nonfluent speakers were: (1) small or absent air reservoir and (2) lack of a vibrating PE segment. Fluent speakers presented with shorter PE segments (1.17 mm) compared to moderately fluent speakers (17.1–29.9 mm). Perceptually, fluent speakers presented with a predominantly rough vocal quality. In contrast, moderately fluent speakers presented with a tense quality. In addition, stoma blast noise was reduced in fluent speakers. Videofluoroscopic findings highly correlated with the quality of esophageal speech.
nent cannula [5], postradiotherapy edema and fibrosis, and/or tumor recurrence [8, 9]. Functionally, spasm, hypotonicity or hypertonicity of the pharyngoesophageal (PE) segment [8–13] have been associated with limited success in EV rehabilitation.

In contrast, successful acquisition of EV is associated with a patent, vibrating hypopharyngeal segment [14]. Parameters of vocal quality, including speech modulation and fluency, are considered essential for satisfactory esophageal speech [5]. It is, therefore, reasonable to assume a relationship between functional characteristics of the PE segment and the fluency and quality of esophageal voice and speech. Furthermore, both voice and swallowing share common anatomical structures, leading some to hypothesize a relationship between deficits in swallowing function and poor EV. Stenosis [12, 15], food stasis and PE spasm [7, 16] have all been reported as factors limiting EV, but no clear relationship between these deficits of deglutition and the acquisition of fluent EV has been described. Moreover, previous investigations have not varied bolus conditions during the evaluation of swallowing function as a means to further elucidate the potential relationship between dysphagia and EV. It is likely that varying the bolus consistencies may add substantial additional information regarding voice production.

We, therefore, hypothesize that, even though perceptual evaluation of EV remains the primary tool for voice rehabilitation, videofluoroscopy may provide valuable insight into the PE segment location and morphology [15]. Furthermore, combining perceptual and instrumental evaluations may establish improved rehabilitation strategies and outcomes. To address these issues, we utilized a combined evaluative approach by which we analyzed esophageal speech and swallowing with various bolus consistencies videofluoroscopically and compared these data with auditory perceptual analyses.

Materials and Methods

Thirty patients who underwent laryngectomy (26 male, 4 female; 30–81 years; mean age 60.3 years) participated in the current study. Of the 30 patients, 19 underwent total laryngectomy alone, 7 underwent total laryngectomy with partial pharyngectomy, 2 underwent total laryngectomy with thyroidectomy, 1 underwent total laryngectomy with resection of the soft palate, and 1 underwent total laryngectomy with total glossectomy. These individuals did not present with additional, overt health problems and had no previous history of myotomy or botulinum toxin injection to the PE segment. The current study was approved by the Ethics Committee.

All subjects had previously undergone at least 12 EV rehabilitation sessions. Gender differences were not considered in the current study due to the relatively low number of female subjects. Furthermore, previous literature suggests that sex is not a factor in EV acquisition [2, 9, 17–20].

Perceptual Analysis

Each subject was recorded for perceptual analysis and also submitted to videofluoroscopy. For perceptual analysis, the parameters selected were: (1) voice quality and (2) speech fluency. Voice quality was classified as either rough, tense, or whisper. EV was also classified according to the type of voicing (constant or intermittent) and the presence of related noise (i.e. stoma blast). Fluency of speech was evaluated according to the Speech Fluency Scale [21]. This scale is divided into 7 degrees, in which A refers to no sound production, B to partial control, single sounds under fair control, C to simple words produced, D to combination of 2–3 words, E to use of some sentences, F to consistent use of sentences, and G to fluent speech with no hesitation. Utilizing this classification scheme, patients were divided into three groups: 9 fluent speakers (F, G), 10 moderately fluent speakers (C–E), and 11 nonfluent speakers (A, B). Seven patients were classified as whisperers (A).

Speech samples were collected in a quiet room using a mono-condenser microphone placed 45° lateral from the speakers’ mouth connected to a professional voice recorder (Marantz, PMD 213). All subjects were asked to speak at a comfortable pitch and loudness for at least 1 min. Five speech-language pathologists with at least 3 years of experience with head and neck cancer rehabilitation evaluated each sample.

Videofluoroscopic Analysis

Videofluoroscopic evaluation was performed during both swallowing and voice production, in the lateral and anteroposterior planes. A coin was placed inferior to the left mastoid of each subject for measurement reference. Swallowing was analyzed using three different consistencies: liquid (barium solution), puree (baby food with barium) and solid (wafer biscuit dipped into barium solution). For voice production analyses, subjects were instructed to produce a sustained ‘a’ or produce ‘pa’ 10 times if prolonged phonation was not possible. In addition, subjects were instructed to count from 1 to 10. Subjects with no voice (grade A) or pharyngeal voice production (grade B) were additionally submitted to an esophageal insufflation test [22]. If no voice was achieved utilizing the insufflation test, a percutaneous lidocaine block of the pharyngeal plexus was performed [10]. After 30 min, the insufflation test was then repeated.

Three speech-language pathologists, with at least 2 years of experience with videofluoroscopic evaluation, analyzed the images by consensus. None of these speech-language pathologists participated in the perceptual voice analysis. The speech-language pathologists characterized the swallowing according to the following parameters: (1) the presence and severity of residue, (2) the presence of pseuodvallecula, and (3) the presence of a cricopharyngeal bar and its position relative to the cervical vertebrae. The PE region was analyzed during phonation according to the following parameters: (1) the presence of an air reservoir and its position relative to the cervical vertebrae, (2) the presence of the PE segment during esophageal phonation, and (3) qualitative characteristics of the PE segment during phonation (frictional or
vibrating) including the participation or not of pharyngeal bar during phonation. All video images were then digitalized for evaluation. The size and anteroposterior diameter (in millimeters, using the digital scale available in the software) of both the PE segment and the air reservoir were analyzed using Adobe-Photoshop program (version 5.5). All photos were analyzed in the lateral plane, utilizing the coin as a reference. The anteroposterior measure of the PE segment was taken from the most cranial portion of the segment up to the prevertebral line. One speech-language pathologist, with 7 years of experience with head and neck cancer rehabilitation and videofluoroscopic evaluation (first author), performed all measurements, blinded to esophageal speech quality. An engineer, with no knowledge of the experimental design, reviewed all measurements. When the PE segment or air reservoir were not visible (e.g. shoulder obstructing view or total esophageal constriction) or when the esophagus was filled with air, the image was classified as ‘not available’.

Samples of videofluoroscopic images are shown in figures 1–3. Figure 1 shows the PE segment of a good esophageal speaker during phonation. Figure 2 shows a videofluoroscopic image of a moderate esophageal speaker during phonation, and figure 3 shows a videofluoroscopic image of a poor esophageal speaker during the insufflation test.

For statistical analyses, the Kruskal-Wallis test was applied to detect possible differences between the three groups of speakers. When a difference was achieved, the Mann-Whitney test was used to pair the groups. A level of 5% (0.050) of significance was employed.

Results

The results of the perceptual analysis are presented in table 1. In table 2 the visual, descriptive parameters and measures obtained by videofluoroscopy are presented according to esophageal speech fluency. All information was obtained in the lateral plane, as the anteroposterior views did not offer satisfactory visualization. The results of the 9 nonfluent esophageal speakers requiring an insufflation test and block of the pharyngeal plexus are presented in table 3. Although 11 nonfluent speakers, classified as either ‘A’ (n = 7) or ‘B’ (n = 4), were identified, 2 were capable of producing esophageal speech for isolated sounds and could be evaluated without insufflation. Finally, table 4 shows videofluoroscopic views of a prominent cricopharyngeal bar identified during phonation.

Discussion

Successful acquisition of EV following total laryngectomy appears to be related to several factors. In the current study, we sought to utilize both videofluoroscopy and perceptual voice analyses in an attempt to identify physiological variables correlated to successful acquisition. The importance of videofluoroscopy for the evaluation of the structures involved in esophageal phonation has been previously described in the literature[14, 23, 24]. These data provide valuable information regarding the patency of the PE segment. The current study sought to expand upon these previous findings by visualizing the lateral plane, since anteroposterior views tend to be limited as previously described by Bentzen et al. [25] and McIvor et al. [13]. Our data suggest that videofluoroscopic examination complements perceptual voice analyses as it provides additional insight into the anatomical and physiological correlates of EV production. Particularly in patients who have been largely unsuccessful in EV acquisition, videofluoroscopy may provide valuable information and guide therapeutic management.
As a complement to videofluoroscopy, we obtained perceptual voice analysis data including EV proficiency. We analyzed the factors that could have influenced voice proficiency, such as age, extent of surgery, duration following surgery, radiotherapy, chemotherapy and presence of pseudoepiglottis. With regard to vocal quality, perceptual analysis consisted of the evaluation of vocal quality, presence of disrupted speech, clunk, stoma blast, buccal speech, mouth whispering and imprecise articulation. All fluent esophageal speakers had a predominantly hoarse vocal quality, while the moderately fluent esophageal speakers presented with both hoarseness as well as a tense vocal quality. Not surprisingly, the most common vocal quality associated with esophageal speech is the hoarse vocal quality due to irregularity of PE segment vibration [26, 27]. Casper and Colton [8] described the esophageal vocal quality as hoarse and rough. Even considering that the fluent esophageal speakers’ vocal quality deviated from laryngeal voice, the hoarse vocal quality is likely more socially acceptable than the harsh or tense vocal quality [28].

Voice production was also rated as either constant or intermittent. This evaluation allowed for the assessment of how the patient effectively produced syllables during speech. A statistically significant difference was observed between fluent and nonfluent esophageal speakers, an expected outcome as nonfluent esophageal speakers had sporadic esophageal productions. Nevertheless, it is interesting to highlight the sound production values in Table 2; all fluent esophageal speakers presented with constant sound production, while 7 moderately fluent speakers had constant sonority and 3 occasional sonority. The literature is sparse regarding this parameter, however, there are some authors that mention the importance of training sound production in all syllables in EV acquisition [8, 29].

According to Doyle [29], two vocal behaviors requiring intervention are the clunk that occurs with air injection into the esophagus and the stoma blast that is the result of turbulent pulmonary inspiration and expiration during esophageal speech production. In the current study, there was no significant relationship between esophageal voice fluency and clunk, since the latter is present both in the fluent and moderately fluent speakers. However, there was a statistically significant relationship between stoma blast and EV fluency. Higher frequency and severity of stoma blast were observed in moderately fluent and nonfluent esophageal speakers when

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Groups of speakers</th>
<th>Groups of speakers</th>
<th>Groups of speakers</th>
<th>Groups of speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fluent</td>
<td>moderately</td>
<td>nonfluent</td>
<td>total</td>
</tr>
<tr>
<td></td>
<td>n %</td>
<td>fluent</td>
<td>n %</td>
<td>nonfluent</td>
</tr>
<tr>
<td>Vocal quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predominantly hoarse</td>
<td>9 30.0</td>
<td>7 23.3</td>
<td>2 6.7</td>
<td>18 60.0</td>
</tr>
<tr>
<td>Predominantly tense</td>
<td>0 0</td>
<td>3 10.0</td>
<td>2 6.7</td>
<td>5 16.7</td>
</tr>
<tr>
<td>Mouth whispering</td>
<td>0 0</td>
<td>0 0</td>
<td>7 23.3</td>
<td>7 23.3</td>
</tr>
<tr>
<td>Sound production</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>0 0</td>
<td>0 0</td>
<td>7 23.3</td>
<td>7 23.3</td>
</tr>
<tr>
<td>Present</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermittent</td>
<td>0 0</td>
<td>3 10.0</td>
<td>4 13.3</td>
<td>7 23.3</td>
</tr>
<tr>
<td>Constant</td>
<td>9 30.0</td>
<td>7 23.3</td>
<td>0 0</td>
<td>16 53.3</td>
</tr>
<tr>
<td>Stoma blasts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>4 13.3</td>
<td>0 0</td>
<td>7 23.3</td>
<td>36.6</td>
</tr>
<tr>
<td>Present</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slight</td>
<td>4 13.3</td>
<td>3 10.0</td>
<td>1 3.4</td>
<td>8 26.7</td>
</tr>
<tr>
<td>Moderate</td>
<td>1 3.4</td>
<td>4 13.3</td>
<td>1 3.4</td>
<td>6 20.1</td>
</tr>
<tr>
<td>Severe</td>
<td>0 0</td>
<td>3 10.0</td>
<td>2 6.6</td>
<td>5 16.6</td>
</tr>
</tbody>
</table>

Statistical significance: vocal quality: fluent × moderately fluent, p < 0.001; sound production: fluent × nonfluent, p = 0.042; stoma blasts: fluent × moderately fluent, p = 0.004, and fluent × nonfluent, p = 0.019.
### Table 2. Visual parameters and measures obtained by videofluoroscopy, according to the groups of esophageal speech fluency

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Groups of speakers</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fluent</td>
<td>moderately fluent</td>
<td>nonfluent</td>
<td>total</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
</tbody>
</table>

#### Puree stasis

- **Absent**
  - n: 4, %: 13.3
  - n: 2, %: 6.7
  - n: 7, %: 23.3
  - n: 13, %: 43.3

- **Present**
  - Slight
    - n: 4, %: 13.3
    - n: 3, %: 10.0
    - n: 3, %: 10.0
    - n: 10, %: 33.3
  - Moderate
    - n: 1, %: 3.3
    - n: 3, %: 10.0
    - n: 1, %: 3.3
    - n: 5, %: 16.7
  - Severe
    - n: 0, %: 0
    - n: 2, %: 6.7
    - n: 0, %: 0
    - n: 2, %: 6.7

#### Solid food stasis

- **Absent**
  - n: 4, %: 13.3
  - n: 2, %: 6.7
  - n: 7, %: 23.3
  - n: 13, %: 43.3

- **Present**
  - Slight
    - n: 4, %: 13.3
    - n: 3, %: 10.0
    - n: 3, %: 10.0
    - n: 10, %: 33.3
  - Moderate
    - n: 1, %: 3.3
    - n: 3, %: 10.0
    - n: 1, %: 3.3
    - n: 5, %: 16.7
  - Severe
    - n: 0, %: 0
    - n: 2, %: 6.7
    - n: 0, %: 0
    - n: 2, %: 6.7

#### Air reservoir

- **Absent**
  - n: 0, %: 0
  - n: 0, %: 0
  - n: 8, %: 26.7
  - n: 8, %: 26.7

- **Present**
  - n: 9, %: 30.0
  - n: 10, %: 33.3
  - n: 3, %: 10.0
  - n: 22, %: 73.3

#### Air reservoir AP diameter

- 1–9.9 mm
  - n: 0, %: 0
  - n: 1, %: 3.3
  - n: 1, %: 3.3
  - n: 2, %: 6.6
- 10–19.9 mm
  - n: 5, %: 16.7
  - n: 5, %: 16.7
  - n: 2, %: 6.6
  - n: 12, %: 40.0
- 20–30 mm
  - n: 4, %: 13.3
  - n: 2, %: 6.6
  - n: 0, %: 0
  - n: 6, %: 20.0

- **Not available**
  - n: 0, %: 0
  - n: 2, %: 6.6
  - n: 8, %: 26.7
  - n: 10, %: 33.3

#### Air reservoir length

- 1–29.9 mm
  - n: 1, %: 3.3
  - n: 1, %: 3.3
  - n: 1, %: 3.3
  - n: 3, %: 10.0
- 30–49.9 mm
  - n: 3, %: 10.0
  - n: 2, %: 6.6
  - n: 0, %: 0
  - n: 5, %: 16.7
- 50–100 mm
  - n: 3, %: 10.0
  - n: 1, %: 3.3
  - n: 1, %: 3.3
  - n: 5, %: 16.7

- **Not available**
  - n: 2, %: 6.6
  - n: 6, %: 20.0
  - n: 9, %: 30.0
  - n: 17, %: 56.6

#### PE segment

- **Not visible**
  - n: 0, %: 0
  - n: 1, %: 3.3
  - n: 9, %: 30.0
  - n: 10, %: 33.3

- **Visible**
  - Frictional
    - n: 3, %: 10.0
    - n: 0, %: 0
    - n: 1, %: 3.3
    - n: 4, %: 13.3
  - Vibration
    - n: 6, %: 20.0
    - n: 9, %: 30.0
    - n: 1, %: 3.3
    - n: 16, %: 53.3

#### PE segment length

- 1–17 mm
  - n: 7, %: 23.3
  - n: 2, %: 6.6
  - n: 1, %: 3.3
  - n: 10, %: 33.3
- 17.1–29.9 mm
  - n: 0, %: 0
  - n: 5, %: 16.7
  - n: 0, %: 0
  - n: 5, %: 16.7
- 30 mm–
  - n: 1, %: 3.3
  - n: 2, %: 6.6
  - n: 1, %: 3.3
  - n: 4, %: 13.3

- **Not available**
  - n: 1, %: 3.3
  - n: 2, %: 6.6
  - n: 9, %: 30.0
  - n: 11, %: 36.6

Statistical significance: puree stasis: moderately fluent × nonfluent, p = 0.021; solid food stasis: moderately fluent × nonfluent, p = 0.021; air reservoir: fluent × nonfluent, p = 0.001; moderately fluent × nonfluent, p = 0.001; air reservoir AP diameter: fluent × nonfluent, p = 0.003; moderately fluent × nonfluent, p = 0.011; air reservoir length: fluent × nonfluent, p = 0.001; PE segment: fluent × nonfluent, p = 0.036; moderately fluent × nonfluent, p < 0.001; PE segment length: fluent × nonfluent, moderately fluent × nonfluent and fluent × moderately fluent, p < 0.001.

AP = Anteroposterior.
compared to fluent esophageal speakers (table 1). The acceptability of esophageal speech depends on decreased stoma blast noise [30], since it may distract and annoy the listener as well as mask the EV, which may also be characterized by reduced loudness [8, 29, 31].

To complement these perceptual analyses, videofluoroscopic examination was performed with various consistencies of contrast in order to assess stasis relative to consistency as well as to determine the potential for a relationship between observed stasis and esophageal speech acquisition. Perry et al. [12] and Diedrich [15] described a correlation between diminished barium transit through the PE segment in cases with stenosis, suggesting that this finding is contributory to poor EV. Gatenby [7] and Müller-Miny [16] also reported decreased liquid barium transit associated with PE spasm. However, no previous investigations utilized puree and solid consistencies in their evaluations. In the current study, moderately fluent esophageal speakers presented with stasis with purees and solids more frequently and severely than nonfluent esophageal speakers, indicating that stasis may not be directly related to poor esophageal speech acquisition. It is important to point out, however, that in our cohort there were no cases of stenosis and none of the subjects had severe complaints of dysphagia. In all cases, residue could be cleared after multiple swallows.

Table 3. Insufflation test and block of the pharyngeal plexus from the nonfluent speaker group

<table>
<thead>
<tr>
<th>Speech</th>
<th>Insufflation test</th>
<th>Lidocaine injection test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>positive</td>
<td>negative</td>
</tr>
<tr>
<td>Mouth whispering</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Pharyngeal speech</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 4. Presence of cricopharyngeal bar according to the groups of speakers

<table>
<thead>
<tr>
<th>Cricopharyngeal bar</th>
<th>Groups of speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fluent</td>
</tr>
<tr>
<td></td>
<td>n</td>
</tr>
<tr>
<td>Absent</td>
<td>2</td>
</tr>
<tr>
<td>Present</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistical significance: 0.487.

In the current study, fluent and moderately fluent speakers differed from the nonfluent group with regard to the presence of an air reservoir (table 1), concurring with previous findings by Robe et al. [23], Van Weissenbruch et al. [9] and Dantas et al. [32]. In addition, the anteroposterior diameter and the length of the air reservoir did not distinguish between fluent and moderately fluent esophageal speakers. However, air reservoir diameter differed between these groups and nonfluent speakers, and air reservoir length differed between fluent and nonfluent speakers (table 2). Interestingly, these data contradict previous reports [9, 33].

With regard to PE segment length, fluent esophageal speakers had shorter PE segments (1–17 mm) than moderately fluent speakers (17.1–29.9 mm, table 2). Measurements for nonfluent speakers were not available. These data concur with previous reports [34]. We hypothesize that less effort is required to eject air from a shorter PE segment in order to produce vibration, thus facilitating improved voice production.

Previous data suggest that increased PE segment tonus after total laryngectomy plays a critical role in the failure of esophageal speech acquisition [10, 24, 35]. In the current study, an insufflation test [22] was performed during videofluoroscopy in 2 subjects with pharyngeal voice and 7 with mouth whispering (table 3). In these cases, the PE segment walls remained occluded with no vibration, as well as minimal or no air reservoir formation. Parapharyngeal nerve blockage with lidocaine injection was then performed [10], and 8 out of 9 patients were able to produce esophageal speech. We therefore conclude that poor esophageal speakers likely present with PE spasm (table 3). No subjects presented with stenosis. These results concur with previous data by Singer and Blom [10], suggesting that 16 out of 129 patients presented with PE spasm. In addition, Gatenby et al. [7] reported that of 17 patients with esophageal speech inadequate for social
The presence of a cricopharyngeal bar is easily detected via the lateral view during videofluoroscopy. It is defined as a mucosal mass at the posterior wall of the pharynx [34], such as a muscular prominence [25, 36, 37], as a submucosal mass [38] or neoglottic bar [39]. We opted to use the term cricopharyngeal bar, as that is most commonly utilized in the Brazilian literature [40]. Several researchers have observed a cricopharyngeal bar vibrating with the PE segment, potentially enhancing successful speech acquisition [17, 20, 25, 34, 36–41]. However, these data remain controversial as many other sources suggest that this bowing is highly correlated with poor esophageal speech acquisition [6, 31].

The presence of a cricopharyngeal bar was observed in 63.3% of our subjects (table 4). Although these data did not achieve statistical significance, likely due to a relatively small sample size, cricopharyngeal bars were more common in fluent (77%) and moderately fluent (80%) esophageal speakers, suggesting a potential relationship between a cricopharyngeal bar and esophageal speech acquisition. We believe that if the air injection into the esophagus is favorable, the presence of a flexible mass facilitates the establishment of sound.

Although many other factors may contribute to EV acquisition, including age [4, 5, 18, 19, 25, 31, 34], extent of surgery [2, 18, 37, 42], postoperative duration [33, 34], presence of pseudoglottis [6, 31] and radiotherapy [20, 42, 43], these factors were not significant in the present study. In addition, we explored a potential relationship between chemotherapy and esophageal speech acquisition, which was also not significant. Our data reinforce the importance of achieving constant sound production and stoma blast reduction as one of the first goals of voice therapy to ensure optimal esophageal vocal quality. However, there is an obvious anatomic factor: PE segment length. Since this segment is composed of muscular fibers of the pharyngeal inferior constrictor muscle, cricopharyngeal muscle and the superior portion of the esophagus, the length of the PE segment may be intrinsic to the patient. Another hypothesis is that PE segment length may be affected by the size of the surgical resection and/or surgical closure technique. It is also hypothesized that prolonged use of esophageal speech may introduce functional adjustments to the structures involved in phonation, as mentioned by Vrtička and Svoboda [36]. This adjustment is not only important for fluency of speech but also for intelligibility since it plays a role as an articulatory organ, helping to achieve adequate voice onset time in plosives, specially for aspirated stops, as in Mandarin [44].

Conclusions

(1) The perceptual vocal quality of fluent esophageal speakers is characterized as predominantly rough, while moderately fluent speakers are rated as rough and strained. Nonfluent speakers, in contrast, produce pharyngeal, buccal speech and mouth whispering. Nonfluent and moderately fluent speakers differ from fluent speakers with regard to stoma noise.

(2) Fluent and moderately fluent speakers have a vibrating PE segment appreciable on videofluoroscopy. The PE segment is shorter in fluent speakers. Visualization of the PE segment is limited in nonfluent esophageal speakers. In addition, nonfluent speakers typically present with PE spasm as well as a small or no air reservoir.

(3) Perceptual and videofluoroscopic analyses are complementary, adding important information about EV and esophageal speech production and are critical to guiding the rehabilitative process.

References


