Speech Performance of Adult Cantonese-Speaking Laryngectomees Using Different Types of Alaryngeal Phonation

Manwa L. Ng, *†Chui-Ling I. Kwok, and †Sau-Fong W. Chow

Illinois State University, Normal, Illinois, U.S.A.; and *Queen Elizabeth Hospital and †New Voice Club of Hong Kong, Hong Kong

Summary: The purpose of the present study was to compare the speech performance of four types of alaryngeal phonation—electrolaryngeal (EL), pneumatic artificial laryngeal (PA), tracheoesophageal (TE), and standard esophageal (SE) speech—by adult Cantonese-speaking laryngectomees. Subjective ratings of (1) voice quality, (2) articulation proficiency, (3) quietness of speech, (4) pitch variability, and (5) overall speech intelligibility were given by eight naive individuals who had no prior experience with any form of alaryngeal speech. Results indicated that SE and TE speech was perceived to be more hoarse than PA and EL speech. EL speech was associated with significantly less pitch variability, and PA speakers produced speech with the least amount of perceived noise. However, articulation proficiency and overall speech intelligibility were found to be comparable in all four types of alaryngeal speakers. Key Words: Laryngectomy—Electrolarynx—Pneumatic artificial larynx—Tracheoesophageal—Standard esophageal.

Laryngectomy is a procedure used to remove a pathological larynx, usually as a result of laryngeal cancer. After the larynx is removed, the upper portion of the trachea is attached to the anterior of the lower neck, where a permanent opening, the tracheostoma, is created. The tracheostoma serves mainly for respiration purposes (1). The procedure results in a total loss of the ability for normal laryngeal phonation due to the removal of the entire laryngeal structure including the vocal folds. Therefore, an important part of rehabilitation after laryngectomy is to seek alternative ways to regain phonation. Currently, there are four methods of regaining phonation after laryngectomy: (i) electrolaryngeal (EL) speech, (ii) pneumatic artificial laryngeal (PA) speech, (iii) tracheoesophageal (TE) speech, and (iv) standard esophageal (SE) speech.

EL speech

Phonation after laryngectomy can be achieved by using a transcervical electrolarynx, which is a handheld, battery-operated external phonatory device. Sound can be transmitted into the vocal tract by correctly coupling the vibratory membrane of the device to the anterolateral aspect of the neck. With appropriate lip, tongue, and jaw movements, different speech sounds can be produced (2). Currently, pitch-controllable and volume-controllable electrolarynges are available. These devices allow the modification of speech fundamental frequency and intensity level during phonation.

PA speech

Adapted from the Tokyo artificial larynx (3), the pneumatic artificial larynx is an external device whose stromal end is coupled to the tracheostoma.
Upon phonation, the pneumatic device user can expel air from the lungs, which sets the rubber reed inside the device into vibration, and sound is thus generated. A plastic tube, which is placed inside the oral cavity, directs the sound into the oval cavity for articulation.

**TE speech**

Introduced by Singer and Blom (4), TE puncture has become a common surgical method to help regain phonation after laryngectomy. During the procedure, a fistula is surgically created between the trachea and the esophagus. A unidirectional TE valve is then inserted into the fistula, allowing pulmonary air to enter into the esophagus through the valve’s tubing. During TE phonation, the tracheostoma must be blocked (usually digitally) so that air can only enter into the esophagus, instead of leaving the lungs via the stoma. The pharyngoesophageal (PE) segment, which is composed of muscle fibers of the inferior pharyngeal constrictor, the cricopharyngeus, and possibly the upper esophagus (5), will then be set into vibration. The PE segment functions as a neoglottis that, similar to the vocal folds in normal laryngeal phonation, serves as a new sound source after laryngectomy.

**SE speech**

In normal laryngeal phonation, air stored in the lungs is expelled through the larynx, which sets the vocal folds into vibration. However, in SE speech, the upper part of the esophagus serves as the air reservoir for phonation. During air intake, air is either injected or inhaled into the upper esophagus, which can retain up to 50 ml of air (6). Upon phonation, air is expelled and the PE segment is set into vibration.

Speech after laryngectomy has been of interest to researchers. There have been a number of studies investigating different perceptual and speech rate characteristics of alaryngeal speech of English speakers (7), particularly those using SE and EL phonation. In general, SE speech was associated with a slower speech rate compared to normal laryngeal, TE, and EL speech (8,9). By studying 118 SE and 35 EL superior speakers reading a passage of 101 words, Shames et al. (8) reported a mean reading rate of 100.3 WPM for the 22 superior SE speakers studied. Filter and Hyman (12) also reported a mean reading rate of 100.1 WPM for SE speakers reading the My Grandfather prose. Similarly, Robbins et al. (9) recorded a mean rate of 99.1 WPM for SE speakers, 127.5 WPM for TE speakers, and 172.8 WPM for normal laryngeal speakers in reading the first paragraph of the Rainbow Passage.

With respect to perceptual ratings, Shames et al. (8) noted significantly better ratings on mean articulation score and word intelligibility for SE speakers than EL speakers. Similarly, Tikofsky (13) reported a lower speech intelligibility score in SE speakers than normal laryngeal speakers. In a study of five superior SE speakers and four superior EL speakers, Bennett and Weinberg (14) found that the speech of those superior SE speakers was more acceptable than that of EL speakers. In addition, a better rating on speech acceptability was found to be associated with a faster rate of speaking and a smaller amount of respiratory noise (15). Similar results were reported by Hoops and Noll (11).

However, contradictory findings were also reported. In a study of eight SE, eight EL, and eight normal laryngeal speakers, Hyman (16) found no significant difference in speech intelligibility between SE speakers and EL speakers. McCroskey and Mulligan (17) noticed greater intelligibility for EL than SE speakers when speech samples were judged by a group of naïve listeners, instead of trained listeners.

**PURPOSE OF THE STUDY**

Despite the large number of studies comparing speech performance between SE and EL speakers, discrepant results were reported. Perhaps due to the lack of prevalence in using PA phonation after laryngectomy, few studies included PA speakers as part of their subject population. The speech performance by PA speakers relative to SE and EL speakers is still not clear. Since the procedure of TE puncture is relatively new, few studies have reported the speech performance associated with TE phonation.

Even though various forms of alaryngeal phonation have been studied extensively with English speakers, there have been few studies investigating alaryngeal speakers of Cantonese. Ching et al. (18) reported the differences in perceptual accuracy of tones among normal, SE, TE, EL, and PA speakers of Cantonese. They concluded that alaryngeal
speakers of Cantonese, particularly speakers using the electrolarynx, failed to produce various Cantonese tones at a level of proficiency comparable to that of normal laryngeal speakers. However, no study has reported the global speech performance of alaryngeal speech of Cantonese in terms of voice quality, articulation, proficiency in tone production, and speech intelligibility.

The purpose of the present study was to describe and compare the speech performance of EL (Servox-type) speakers, PA (Taiwan tube) speakers, TE speakers, and SE speakers of Cantonese. The parameters investigated were the (a) voice quality based on the amount of hoarseness, (b) articulation proficiency, (c) pitch variability, (d) quietness of speech according to the amount of perceived noise, and (e) overall speech intelligibility.

METHODOLOGY

Speakers

Fifteen EL speakers (Servox type), 11 users of the pneumatic artificial larynx (Taiwan tube), 12 TE speakers, and 15 SE speakers of Cantonese participated in the present study. They were all males, with ages ranging from 48 to 80 years. The speakers had no known history of speech problems, except that associated with laryngectomy. They were all rated as superior speakers by three practicing speech pathologists in Hong Kong. The speakers were selected from the New Voice Club of Hong Kong, which is a nonprofit self-help organization for the laryngectomee community in Hong Kong. All speakers were literate, with at least a grade-school education. They were all able to read the speech materials used in the present study.

Listeners

Eight naïve adults (six females and two males) who were not familiar with alaryngeal speech of any type served as listeners. They were all native speakers of Cantonese with no known history of a hearing problem. All listeners were reported to possess at least a high-school education, and they demonstrated the ability to read and comprehend correctly the reading passage before the experiment.

Speech material

A short passage selected from a third-grade reading book served as the speech material. The passage consisted of 136 Chinese monosyllabic characters. Before the recordings, the speakers were given a brief practice session, so as to familiarize themselves with the reading material and the recording environment. During the practice and recording sessions, they were provided with cards on which Chinese characters of the reading passage were printed.

Recording procedure

All recordings took place in a quiet recording room located at the New Voice Club of Hong Kong. The speech samples were recorded on high-quality audio tapes (Sony UX-Pro 46) via a Sony TC-D5 Pro-II portable audio recorder with a Shure SM48 dynamic microphone. The microphone was placed ~10 in. in front of the mouth of the speaker. Mesh screen was used to cover the microphone to minimize the recording of extraneous noise. The speakers were instructed to read the short passage once at a comfortable loudness level. Since speech intensity contributes to better intelligibility, and different speaker groups are associated with different loudness levels, no attempt was made to adjust the loudness levels among different speaker groups during the recordings. The relative loudness difference among different speaker groups was retained.

Listening task

Before the listening session, the listeners were given a card on which the reading passage was printed. They were instructed to read and comprehend correctly the reading passage before the experiment.

The listening task took place in the sound-treated Language Laboratory at the City University of Hong Kong. The recorded speech samples were randomized and presented to the listeners via high-quality headphones. Throughout the listening session, loudness control was kept constant in an attempt to maintain the original loudness difference among different speaker groups. The listeners were provided with answer sheets and they were asked to rate the speech samples on 1-to-7 equal-interval scales (see Appendix A) based on the following speech parameters.

Speech parameters

Voice quality based on the amount of hoarseness

The rating of voice quality was based on the perceived amount of hoarseness in the speech samples. On the scale of 1 to 7, 1 corresponded to a voice with severe hoarse quality, whereas 7 referred to a normal voice.

Articulation proficiency

Articulation proficiency referred to the correctness of articulation perceived. It was rated accord-
ing to the amount of misarticulation or poor articulation in the reading passage. On the scale of 1 to 7, 7 referred to normal articulation, whereas 1 related to poor articulation.

**Quietness of speech**

Although all alaryngeal speakers demonstrated superior speech performance, a minimal amount of noise was still observed in their speech. The noise was associated with the neck radiation of the electrolarynx in the EL speakers or the respiratory and/or stoma noise associated with PA, TE, and SE speakers. Quietness of speech was rated according to the amount of noise perceived in the speech samples. On the rating scale, 1 referred to speech production with the most amount of noise, and 7 to noise-free production.

**Pitch variability**

Ratings on pitch variability was based on the presence of pitch variations. In Cantonese, there are six possible tones associated with each sound. They are high-level, high-rising, midlevel, low-falling, low-rising, and low-level tones (18). Even for the same syllable, such as /si/, sounds with different tones have different meanings (see Appendix B). A rating of 1 represented a total lack of pitch variability, while 7 reflected a voice with appropriate tone variations.

**Overall speech intelligibility**

Speech intelligibility referred to the extent to which the listeners were able to understand the short passage read by the alaryngeal speakers. Listeners were instructed to judge the speech intelligibility independent of previous knowledge of the reading passage. They were asked to judge the speech samples as if it were the first time they had heard the passage. A rating of 1 referred to poor speech intelligibility, while 7 referred to excellently intelligible speech.

**Reliability measures**

Five speech samples (~10% of all the speech samples) were randomly selected from the datum corpus and were presented to the eight listeners for judgment the second time. Pearson product-moment correlation was then calculated based on the results obtained from the first and second judgments of the five randomly selected samples.

**RESULTS**

Subjective ratings for the different parameters were analyzed statistically. The average score and the standard deviation of each speech parameter were calculated and the results are shown in Table 1. A considerable amount of individual variations can be observed. A significant correlation (r = 0.84) between the first and the second ratings on the five randomly selected samples was observed, indicating a good intrajudge reliability.

In analyzing the results, raw data collected in the listening task were tested using analysis of variance (ANOVA), which took into account the unequal number of speakers in a given speech mode.

**Voice quality**

Voice quality for each speaker was rated with respect to the amount of hoarseness present in the voice samples (see Table 1). The higher the ratings, the less the hoarseness and the better the voice quality perceived. Results of one-way ANOVA indicated a significant difference in voice quality among different types of alaryngeal speech (p < 0.01). A Scheffé post hoc test further indicated that significant differences existed between PA and TE and between PA and SE speakers at the 0.01 level of confidence.

**Articulation proficiency**

Average scores and standard deviation values of articulation proficiency for each type of speech are presented in Table 1. Results of an ANOVA showed that, regardless of the speech options, there was no significant difference in articulation proficiency among different types of alaryngeal speech.

**Quietness of speech**

Average ratings and standard deviation values of speech quietness for different types of alaryngeal speech are presented in Table 1. Results of an

**TABLE 1. Mean scores and standard deviations of ratings of different speech parameters for different types of alaryngeal speech**

<table>
<thead>
<tr>
<th>Speaker group</th>
<th>Voice quality (Mean)</th>
<th>Articulation proficiency (Mean)</th>
<th>Quietness of speech (Mean)</th>
<th>Pitch variation (Mean)</th>
<th>Overall speech intelligibility (Mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EL</td>
<td>3.38 (0.48)</td>
<td>4.10 (0.70)</td>
<td>3.58 (0.41)</td>
<td>2.40 (0.40)</td>
<td>4.15 (0.64)</td>
</tr>
<tr>
<td>PA</td>
<td>4.15 (0.56)</td>
<td>4.17 (0.65)</td>
<td>4.37 (0.55)</td>
<td>5.70 (0.66)</td>
<td>4.65 (0.72)</td>
</tr>
<tr>
<td>TE</td>
<td>2.56 (0.52)</td>
<td>3.91 (0.70)</td>
<td>3.94 (0.58)</td>
<td>4.48 (0.56)</td>
<td>4.41 (0.38)</td>
</tr>
<tr>
<td>SE</td>
<td>2.79 (0.66)</td>
<td>3.57 (0.71)</td>
<td>3.51 (0.63)</td>
<td>4.00 (0.65)</td>
<td>3.89 (0.69)</td>
</tr>
</tbody>
</table>

*The ratings were made on a scale of 1-7, where 7 referred to the best production and 1 referred to the worst production.*
ANOVA indicated a significant main effect for speaker type. Scheffé post hoc tests indicated significant differences between PA and SE speakers \((p < 0.01)\) with respect to the amount of noise perceived in the speech samples. Less noise appeared to be associated with PA than with SE speakers.

**Pitch variability**

Average scores and standard deviation values of pitch variability are presented in Table 1. The results of an ANOVA indicated a significant main effects for speaker type. Scheffé post hoc tests showed that significantly greater variation in pitch was associated with PA speakers \((p < 0.01)\) than other types of alaryngeal speech. PA speakers were able to produce speech with more pitch variations than other types of alaryngeal speakers.

**Speech intelligibility**

Average scores and standard deviation values of overall speech intelligibility are presented in Table 1. ANOVA revealed no significant difference in speech intelligibility among the four types of alaryngeal speakers.

**DISCUSSION**

Fifty-three superior alaryngeal speakers using four types of alaryngeal phonation participated in the present study. They were able to speak with minimal effort and maximum clarity, and their speech was readily intelligible as judged by the three speech pathologists. Alaryngeal speakers using a Servox-type electrolarynx produced speech sounds with a minimal amount of neck sound radiation, whereas speakers using the pneumatic artificial larynx had been able to maintain a near-leakless coupling between the tracheostoma and the device. SE and TE speakers were able to speak with a minimal amount of inspiratory and/or air intake noise.

In the present study, the voice quality of PA speech appeared to be significantly better than that of SE and TE speech. Both SE and TE speakers showed a lower score on voice quality, which implied that the voice quality of SE and TE phonations was perceived as worse than that of PA and possibly EL speech. This finding was consistent with that reported by Smith et al. (19), where greater jitter values were found in SE voice than in normal laryngeal voice. These greater jitter values might have been due to the use of the PE segment as a new sound source in both SE and TE phonation. Although TE and SE speakers were able to regain phonation after laryngectomy by using this new vibratory medium, due to the lack of voluntary control over the PE segment, greater amounts of aperiodicity were produced concomitantly. This greater amount of aperiodicity resulted in more noise in the speech signals. According to previous studies, a hoarse voice quality was found to be associated with greater amounts of noise (or aperiodicity) in the speech signals (20). In the case of PA and EL phonation, an external device was used as the new vibratory source. These external mechanical devices provided a regular sound source with a constant vibratory frequency. Therefore, less hoarse quality was perceived in these two forms of alaryngeal speech.

Results also indicated that there was essentially no difference in articulation proficiency when the different speaker groups were compared. This suggests similar articulatory gestures in all forms of alaryngeal phonation (EL, PA, TE, and SE) and was related mainly to the premorbid articulation ability of the laryngectomees. This notion supports the assumption underlying the source–filter theory of speech production (21), which assumes that the articulatory system (the filter) and the laryngeal system (the source) are mutually independent. The alaryngeal speakers were able to speak (i.e., both to phonate and to articulate) with a normal larynx before the surgery. Even though the laryngeal structure was excised, the articulatory system was essentially intact, leaving the laryngectomees the ability to articulate normally.

In addition, significantly less noise was perceived in PA than SE speakers of Cantonese. The noise perceived could be caused by the air turbulence present at the tracheostoma and/or the sound of air intake found exclusively in SE phonation. Despite the fact that alaryngeal speakers demonstrated superior speech performance, the possible “klunking” noise during SE speech production might also contribute to the greater amount of perceived noise. In the case of PA speech, the superior speakers who participated in the present study were all able to maintain a fairly complete seal by coupling the spongy end of the pneumatic artificial larynx to the stomal opening. Hence, minimal noise was produced.

Due to the limitation of the device, speakers using the electrolarynx demonstrated a very limited pitch variation (mean score = 2.40). The pitch of EL speech, which corresponds to the rate of vibra-
tion of the mechanical membrane of the Servox-type electrolarynx, is essentially predetermined by the device. Despite the fact that the device was pitch and volume controllable, all EL speakers failed to manipulate the device to achieve different pitches throughout the course of recording. The problem relates to the fact that Cantonese is a tone language (see Appendix B), in which each syllable/word possesses its own pitch contour. Users of the electrolarynx were simply not fast enough to vary the pitch to contrast words with different tones. This notion has been discussed by Gandour et al. (22) in their study of Thai alaryngeal speakers and by Ching et al. (18) in their study of Cantonese alaryngeal speakers.

All forms of alaryngeal phonation demonstrated similar levels of speech intelligibility. This showed that acceptable and intelligible speech could be achieved by different forms of alaryngeal phonation. Each type of alaryngeal phonation was associated with its own perceptual characteristics (shown by the different scores for different parameters). With sufficient training, alaryngeal speakers can learn socially acceptable speech, and all existing methods of alaryngeal phonation can become a viable tool for verbal communication.

Acknowledgment: This study was supported by the New Voice Club of Hong Kong.

REFERENCES


APPENDIX A

Answer Sheet Listeners: __________
Subject Code: ______ Date: __________

1. Voice Quality (according to the amount of hoarseness)
   
   1 2 3 4 5 6 7
   severe acceptable clear
   hoarseness

2. Articulation Proficiency

   1 2 3 4 5 6 7
   poor acceptable good

3. Quietness of Speech

   1 2 3 4 5 6 7
   very noisy acceptable quiet

4. Pitch Variability

   1 2 3 4 5 6 7
   monotonic acceptable appropriate

5. Overall Speech Intelligibility

   1 2 3 4 5 6 7
   poor acceptable good

APPENDIX B

Examples of different Cantonese tones with syllable /si/

<table>
<thead>
<tr>
<th>Tone</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>High level</td>
<td>&quot;Poem&quot;</td>
</tr>
<tr>
<td>High rising</td>
<td>&quot;History&quot;</td>
</tr>
<tr>
<td>Mid level</td>
<td>&quot;To try&quot;</td>
</tr>
<tr>
<td>Low falling</td>
<td>&quot;To serve&quot;</td>
</tr>
<tr>
<td>Low rising</td>
<td>&quot;Market&quot;</td>
</tr>
<tr>
<td>Low level</td>
<td>&quot;Time&quot;</td>
</tr>
</tbody>
</table>