An Acoustic Study of Vowels Produced by Alaryngeal Speakers in Taiwan

Jia-Shiou Liao

Purpose: This study investigated the acoustic properties of 6 Taiwan Southern Min vowels produced by 10 laryngeal speakers (LA), 10 speakers with a pneumatic artificial larynx (PA), and 8 esophageal speakers (ES).

Method: Each of the 6 monophthongs of Taiwan Southern Min (/i, a, ɔ, u, ə/) was represented by a Taiwan Southern Min character and appeared randomly on a list 3 times (6 Taiwan Southern Min characters × 3 repetitions = 18 tokens). Each Taiwan Southern Min character in this study has the same syllable structure, /V/, and all were read with tone 1 (high and level). Acoustic measurements of the 1st formant, 2nd formant, and 3rd formant were taken for each vowel. Then, vowel space areas (VSAs) enclosed by /i, a, ɔ, u/ were calculated for each group of speakers. The Euclidean distance between vowels in the pairs /i, a/, /i, u/, and /a, u/ was also calculated and compared across the groups.

Results: PA and ES have higher 1st or 2nd formant values than LA for each vowel. The distance is significantly shorter between vowels in the corner vowel pairs /i, a/ and /i, u/. PA and ES have a significantly smaller VSA compared with LA.

Conclusions: In accordance with previous studies, alaryngeal speakers have higher formant frequency values than LA because they have a shortened vocal tract as a result of their total laryngectomy. Furthermore, the resonance frequencies are inversely related to the length of the vocal tract (on the basis of the assumption of the source filter theory). PA and ES have a smaller VSA and shorter distances between corner vowels compared with LA, which may be related to speech intelligibility. This hypothesis needs further support from future study.

A

fter total laryngectomy, the anatomical changes in alaryngeal speakers’ vocal tract consist of the removal of the laryngeal structure, including vocal cords, and the absence of a connection between the trachea and the nose and mouth. Laryngectomees lose normal laryngeal function and phonation. In order to achieve intelligible speech to improve their quality of life, there are alternative methods for restoring voice. The two most popular types of alaryngeal phonation in Taiwan currently are pneumatic artificial laryngeal speech and esophageal speech.

In pneumatic artificial laryngeal speech, alaryngeal speakers with a pneumatic artificial larynx (PA) exhale air through the tracheostoma, vibrating a rubber diaphragm to create a fundamental sound. The fundamental sound is then directed through a pipe into their mouth. This is the source of air when they articulate. In esophageal speech, the sound source of esophageal speakers (ES) is air insufflated into the cervical esophagus. ES then immediately expel the air to vibrate the pharyngoesophageal (PE) segment, which consists of the musculature and mucosa of the lower cervical area. The PE segment divides the pharynx from the esophagus. Alaryngeal speakers regain phonation on the basis of their communicative needs, comorbidities, and personal preference for their chosen speech mode.

Because of the sound source, the means of vibrating the air, and the size and configuration of the vocal tract, the voice quality of alaryngeal speakers is different from that of laryngeal speakers (LA; Most, Tobin, & Mimran, 2000; Ng, Gilbert, & Lerman, 2001; Ng, Lerman, & Gilbert, 1998; Robbins, 1984; Xu, Chen, Lu, & Qiao, 2009). Acoustical and perceptual analysis of speech sounds are the methods for evaluating the voice quality of typical and atypical speakers. Nevertheless, the best way of describing vowels is in terms of their acoustic properties (Fant, 1970; Kent & Read, 2002; Ladefoged, 2003). For the acoustic analysis of vowels, the speech production of laryngeal and alaryngeal speakers has been examined for duration, intensity, fundamental frequency, jitter, shimmer, vowel space, and first formant (F1), second formant (F2), and third formant (F3) values (Cervera, Miralles, & González-Álvarez, 2001; H. Liu, Wan, Ng, Wang, & Lu, 2006; H. Liu, Wan,
that PA have higher F2 in Cantonese /u/ and Chu (2009) suggested that F1 values of Cantonese are more fronted or are made with higher tongue positions. Cervera et al. (2001) suggested that, compared with ES and LA (Ng & Chu, 2009). Compared with the studies related to esophageal and tracheoesophageal speech, the findings associated with vowel formants of PA are limited because of the speech modes preferred in English-speaking communities (Chalstreya, Bleacha, Cheunga, van Ravensteijn, Koopmans-van Beinum, & Hilgers, 1994; Ng & Chu, 2009; Ng et al., 1997; Xu et al., 2009).

### Formant Frequencies

A formant refers to vocal tract frequency. The lowest two vowel formant frequencies are the most relevant acoustic parameters for characterizing vowels (Fant, 1970; Kent & Read, 2002; Peterson & Barney, 1952). Patterns of change in formant frequencies have been used in basic and clinical research to reflect the activity of the vocal tract indirectly (Dromey, Nissen, Roy, & Merrill, 2008; Lee, Yu, Hsieh, & Lee, 2015; Liss & Weismer, 1992; H.-M. Liu, Tsao, & Kuhl, 2005). Peterson and Barney (1952) described the acoustic properties of English vowels uttered within the /hVd/ context by 33 men, 28 women, and 15 children. They found that the ranges of sounds within a vowel phoneme correlate closely with F1 and F2 (Fant, 1970; Kent & Read, 2002; Peterson & Barney, 1952; Pickett, 1999). The frequency of F1 is related to oral and pharyngeal constrictions. The more open the mouth or the more constricted the pharynx, the higher the F1 frequency. The frequency of F2 is related to tongue advancement (the length of the front cavity). F2 values are raised by a front tongue constriction and lowered by a back tongue constriction. Lip rounding lowers all formant frequencies because it lengthens the lip passage and brings F3 closer to F2 (Fant, 1970; Kent & Read, 2002; Peterson & Barney, 1952; Pickett, 1999). It is noteworthy that formant frequencies are a function of the size and shape of the vocal cavities. Source filter theory (Kent & Read, 2002), as applied to vowel production, assumes that the frequency of the vocal cord vibration and the resonant frequency of the vocal tract are independent. The resonant frequencies of the vocal tract are determined mainly by the length of the tract and its cross-sectional area (Fant, 1970; Kent & Read, 2002; Miller, 1989; Stevens, 2000).

Formant measures have been applied in studies for alaryngeal speakers. PA, ES, and speakers with tracheoesophageal speech (TE) have higher F1 and F2 values than typical speakers in English (Sisty & Weinberg, 1972), Spanish (Cervera et al., 2001), Mandarin (H. Liu & Ng, 2009; H. Liu et al., 2005), and Cantonese (Ng et al., 1997). Ng and Chu (2009) suggested that F1 values of Cantonese vowels of PA are consistently higher than those of LA and that PA have higher F2 in Cantonese /a\, u, æ, œ, e/ than Spanish vowels produced by ES and TE have significantly higher formant values than those produced by LA (Cervera et al., 2001). Cervera et al. (2001) suggested that, compared with vowels of typical speakers, vowels produced by ES and TE are more fronted or are made with higher tongue positions. Dutch TE and LA have comparable F1 values in vowels from nonsense syllables. However, TE have higher F2 values in /a, ï/ and lower F2 values in /u/ compared with LA (van As, van Ravensteijn, Koopmans-van Beinum, Hilgers, & Pols, 1997). Possible explanations for these differences in formant frequencies are the reduced length of the vocal tract and the lowered position of the tongue resulting from the removal of the larynx (Christensen & Weinberg, 1976; Sisty & Weinberg, 1972). However, Damsté (as cited in van As et al., 1997) stated that vowels produced by ES and LA speakers differ little in their formant frequencies because there is little difference between the speakers’ buccopharyngeal cavities.

Fant (1970) and Stevens and House (1955) suggested that, when accounting for the variation in formant frequencies, the elevation and advancement of the tongue may not be sufficient to characterize the vowel quality. Factors such as the size of the tongue, the size of the vocal tract, the configuration of the oral and pharyngeal cavities, and the speech modes used should be taken into consideration (H.-M. Liu et al., 2005). As found in the study of Sisty and Weinberg (1972), the length of the altered vocal tract may be shortened and thus produce higher formant frequencies of vowels. Moreover, Kyutta (1964) stated that the back of the tongue is connected with the esophagus when the vocal cords are removed. This can have two possible explanations. First, the PE segment may shift from between the fourth and fifth cervical vertebrae to between the third and fourth (Bentzen, Guld, & Rasmussen, 1976; Damsté & Lerman, 1969; Diedrich & Youngstrom, 1966; Kyutta, 1964; van As, Op de Coul, van den Hoogen, Koopmans-van Beinum, & Hilgers, 2001). As an alternative, the raising of formant frequencies may occur because of the loss of the rearmost resonance cavity in the vocal tract (Bentzen et al., 1976; Damsté & Lerman, 1969; Diedrich & Youngstrom, 1966; Kyutta, 1964; van As et al., 2001). The sound source of PA is from the oral cavity inward, which is different from that of ES and LA. This suggests that the resonance patterns of PA are different from those of ES and LA. The speech production of PA is therefore different in nature from that of ES and LA (Ng & Chu, 2009). Compared with the studies related to esophageal and tracheoesophageal speech, the findings associated with vowel formants of PA are limited because of the speech modes preferred in English-speaking communities (Chalstreya, Bleacha, Cheunga, & van Hasselte, 1994; Ng & Chu, 2009; Ng et al., 1997; Xu et al., 2009).

### Vowel Space Area

Vowel space area (VSA) refers to the two-dimensional area enclosed by corner vowels such as English /i, æ, u, a/ in an F1/F2 plane (Fant, 1970; Peterson & Barney, 1952). Because F1 and F2 values correlate with the size and shape of the cavities created by tongue height (F1) and tongue advancement (F2), VSA is an acoustical reference for the displacements of the articulators.

The Euclidean distances (EDs) and Heron’s formula are applied to calculate VSA (Ball & Gibbon, 2013). Corner vowels have extreme formant frequency values in the F1/F2 plane, and they form the VSA. The length of each side is the distance between the pairs of corner vowels.

The equation for ED is as follows:

\[
ED_{xy} = \sqrt{(F_{1x} - F_{1y})^2 + (F_{2x} - F_{2y})^2}.
\]
Using frequency values of $F_1$ and $F_2$ from two different vowels, $x$ and $y$, this calculates the distance between vowel $x$ and vowel $y$ in hertz. For the VSA quadrilateral or trilateral, the ED values for the vowel pairs produced by a speaker are the length of each side.

English corner vowels /i, æ, a, u/ form a quadrilateral VSA. The English quadrilateral area can be divided into two separate triangular VSAs enclosed by /i, æ, a/ and /i, a, u/. The ED values for the vowel pairs /i, æ/, /i, a/, /æ, a/, /æ, u/, /i, u/, and /æ, u/ are used to calculate the VSAs enclosed by /i, æ, a/ and /i, a, u/. Heron’s formula below is for calculating the area of a triangle when the lengths of all three sides are known. Let $a$, $b$, and $c$ be the lengths of the sides of a triangle. The area is given by

$$A = \sqrt{s(s-a)(s-b)(s-c)},$$

where $s$ is half the perimeter, or $\frac{a + b + c}{2}$.

Estimation of VSA has been performed in various studies, such as those of speaker characteristics and speaking style (Flipsen & Lee, 2012; Krause & Braida, 2004; Tsao, Weismer, & Iqbal, 2006). VSA is greater for the slow talkers of both sexes of speakers, and VSA is larger in women than in men (Tsao et al., 2006). Sex differences in VSA are apparent in late adolescence and adulthood, and VSA declines notably with age (Flipsen & Lee, 2012; Peterson & Barney, 1952). Clear speech takes up a significantly larger speech area than conversational speech when both are uttered at a normal rate (Krause & Braida, 2004).

Moreover, VSA analysis has added to our understanding of the fields of language development and speech disorders. Children with dysarthria (Higgins & Hodge, 2001) or impaired hearing (Kent, Osberger, Netsell, & Hustedde, 1987), speakers with amyotrophic lateral sclerosis (Turner, Tjaden, & Weismer, 1995) or cerebral palsy (H. Liu et al., 2005; Weismer, Jeng, Laures, Kent, & Kent, 2001), and adults with partial glossectomy (Whitehill, Ciocca, Chan, & Samman, 2006) or total laryngectomy (Cervera et al., 2001; H. Liu & Ng, 2009) have relatively restricted VSA compared with the control group in the previously mentioned studies. The reduction of vowel formant frequencies or vowel centralization result in compressed VSA, which can be regarded as a feature of speech production deficits.

Alaryngeal speakers must learn to adjust their speech production to changed vocal tract dimensions after their total laryngectomy in order to reach acceptable speech accuracy. Analysis of acoustical VSA provides a measure of how speakers manage such adjustment and may help increase our knowledge of how anatomical and physical changes affect speech production by alaryngeal speakers.

Acoustic studies related to alaryngeal speakers have been done in several languages, including English (Sisty & Weinberg, 1972), Dutch (van As et al., 1997), Finnish (Kytta, 1964), Spanish (Cervera et al., 2001), Cantonese (Ng et al., 1997), and Mandarin (H. Liu & Ng, 2009). However, speaking with a pneumatic artificial larynx is not a commonly chosen option for alaryngeal speakers in English-speaking communities; therefore, there are fewer studies related to PA than studies associated with esophageal and tracheoesophageal speech (Boyd, Varvares, & Fitzmaurice, 1995; Chalstreya et al., 1994; Ng & Chu, 2009; Ng et al., 1997; Xu et al., 2009). Wang, Wang, Huang, and Tseng (2009) reported that 62.5% of 152 laryngectomees they surveyed in Taiwan used an artificial pneumatic introral laryngeal device to communicate, and 18.4% used esophageal speech. However, there is no study comparing Taiwanese PA and ES with respect to the acoustic parameters of $F_1$, $F_2$, and $F_3$ of the vowels they produce. Therefore, the aim of this study was to investigate the acoustical properties of Taiwan Southern Min vowels produced by PA, ES, and LA using $F_1$, $F_2$, and $F_3$ values and VSA.

**Method**

**Participants**

The participants in this study, like three quarters of the population of Taiwan, acquired Taiwan Southern Min as their first language (Huang, 1993). Southern Min is a member of the Hokkien family of Chinese languages. There were 10 PA, 10 PA, and eight ES participating, all men. All the alaryngeal participants in this study had undergone the total laryngectomy surgical procedure. The PA participants were aged between 49 and 76 years ($M = 57$ years) and had a PA speech experience of at least 2 years. Nine out of 10 PA were receiving radiation therapy. The ES participants were aged between 51 and 71 years ($M = 60$ years) and had a regular ES speech experience of at least 2 years. Six out of eight ES were receiving radiation therapy. The age-matched LA were aged between 41 and 73 years ($M = 57$ years). All the alaryngeal participants were from different cities of Taiwan and were members of the Association of Laryngectomees, a nonprofit organization. They attend a weekly 2-hr class in which they either review their communication skills or help newcomers. The weekly classes with speech therapists in hospitals are held in Taichung in central Taiwan and in Tainan in southern Taiwan. The class in Taipei is held in an education center. The alaryngeal speakers could choose which meeting place was convenient for them. None of the laryngeal participants self-reported that they had a hearing or speech disorder. The alaryngeal participants had no known history of hearing problems except those associated with their laryngectomy. In addition to this experimenter’s observation, the quality of each participant’s voice was judged from recordings by a speech pathologist who has more than 20 years of experience in teaching alaryngeal speakers how to produce phonation and assessing their speech performance in a sound-attenuated room in a laboratory. The assessment items adopted from Ng et al. (1997) included (a) voice quality, (b) articulation proficiency, (c) quietness of speech, (d) pitch variability, and (e) overall speech intelligibility. The 28 participants had at least acceptable overall speech intelligibility. The assessment results from the pathologist are listed in Appendix A.
**Data Collection—Taiwan Southern Min Vowels**

One of the Taiwan Southern Min syllable structures, /V/, was used for the speech materials in this study. The monophthongs of Taiwan Southern Min, /i, e, a, ë, u, æ/, do not differ in their length. /V/ syllables are meaningful in Taiwan Southern Min, so each vowel can be represented by a Taiwan Southern Min character. These characters are 'ì, è, í, æ, ə, ŭ' (meaning “he/she,” “mill,” “an initial particle,” “black,” “pollute,” and “edible chrysanthemum,” respectively)—words with /V/ syllable structure. Participants were asked to read a list of the Taiwan Southern Min monophthongs put in the /V/ context with tone 1 (high and level). For example, /u/ was put in /V/ to form /u/ and read with tone 1 (see Appendix B).

Eighteen Taiwan Southern Min characters were presented randomly and listed on a piece of paperboard for the participants to read. Each Taiwan Southern Min character appeared randomly on the list three times (6 Taiwan Southern Min characters × 3 repetitions = 18 tokens). Before recording, all participants were given a short warm-up session. All participants were instructed to read the test characters clearly and at a rate that they felt to be reasonably normal. In view of the size of the list, participants were allowed to self-correct in the event of slips of the tongue or misreading.

Ten laryngeal participants were recorded in a sound-attenuated room in the laboratory. For laryngeal participants, the microphone (P48, Audio-Technica, Tokyo, Japan) was in front of the mouth at a distance of about 15 to 20 cm. For alaryngeal participants, the microphone was in front of the mouth at a distance of about 15 to 20 cm. The microphone (P48, Audio-Technica, Tokyo, Japan) was in front of the alaryngeal participants who live in Taichung or who were placed in front of the alaryngeal participants in Taipei or Tainan. All participants were given about $40 as financial compensation.

## Data Measurement

The vowel onset was assigned at the point where the steady state of formant bars of the vowel began on the spectrogram. The vowel offset was assigned at the point where the formant bars ended. For the Taiwan Southern Min monophthongs, once the duration of the target vowel was determined, the F1, F2, and F3 of each vowel were measured at the midpoint in the duration of the vowel. These measurements were taken with linear predictive coding estimates derived from Praat (Boersma & Weenink, 2013). At the same time that formant values were being provided directly from the output of the formant tracking by Praat, wide-band spectrograms for each vowel were visually examined.

In this study, the first three formants were used to outline the acoustic properties of a vowel; however, using the recommended Praat values, five formants were in fact extracted below the ceiling value of 5000 to 5500 Hz. When spurious formant values came up, one of the following methods was used to get formant values instead. First, if the formant track on the spectrogram (which generally is smooth for a good quality sound) makes an abrupt jump at the point at which a formant value is measured, a visual estimation of the formant value from the horizontal center of the formant band would be made from the spectrogram. Second, when two formants are very close together, formant track errors can occur easily because the two formants are regarded as a single formant. The solution in this situation is to increase the number of formants in the entry from five to six or seven at Praat setting. The original red formant track on the spectrogram would become two separate formant tracks, and then formant values would be taken directly from the output of the formant tracking. About 17% of the F1 to F3 values from LA were produced by one of the two above-mentioned methods. About 33% of the F1 to F3 values from PA were produced by one of the two above-mentioned methods. About 38% of the F1 to F3 values from ES were produced by one of the two above-mentioned methods.

In the literature, the presence of noise in pathological voices may create difficulty in measuring formant values (Sisty & Weinberg, 1972; van As et al., 2001). This problem occurred more often in ES than in PA in this study.

### Reliability

To evaluate for intrarater and interrater reliability, 10% of the F1, F2, and F3 formant values from the entire vowel data corpus were randomly selected and remeasured by this investigator and a second investigator. The results were compared with the first F1, F2, and F3 measurements. Pearson product–moment correlation coefficients (r) were used to reflect the intrarater and interrater reliability of F1, F2, and F3 measurements. For intrarater reliability, the average absolute error and Pearson product–moment correlation coefficient of F1 measurements were 24.88 Hz and .967, respectively; those of F2 were 31.00 Hz and .993, respectively; and those of F3 were 50.67 Hz and .980, respectively. For interrater reliability, the average absolute error and Pearson product–moment correlation coefficient of F1 measurements were 25.47 Hz and .967, respectively; those of F2 were 36.52 Hz and .989, respectively; and those of F3 were 74.02 Hz and .966, respectively.
Results

In this study, F1, F2, and F3 values were used to define the acoustic characteristics of the spoken Taiwan Southern Min monophthongs. In this study, each Taiwan Southern Min monophthong was read three times by each participant. Each participant’s F1 values of the same vowel from three readings were averaged for statistical analysis; the same procedure was applied for F2 and F3 values.

Repeated measures analysis of variance (ANOVA) was applied to the data sets of Taiwan Southern Min monophthongs to determine whether F1, F2, and F3 values did vary as expected. Mauchly’s test of sphericity was used to test the repeated measures assumption of sphericity. If the sphericity assumption was violated, then Greenhouse–Geisser adjusted values were reported instead. Mean values for F1, F2, and F3 were further compared in pairwise comparisons using Bonferroni’s adjustment for multiple comparisons.

First Formant

A two-way mixed-design ANOVA was conducted with group (LA, PA, and ES) as the between-subjects factor and vowel (/i, e, a, ɔ, u, ə/) as the within-subject factor. Table 1 reports the means and standard deviations of F1 for each Taiwan Southern Min vowel from different speech groups (see Figures 1, 2, and 3). A significant difference between vowels exists in F1 values, $F(2, 55) = 182.482$, $p < .001$, $\eta^2_p = .880$, and a significant difference between groups exists in F1 values, $F(2, 25) = 25.085$, $p < .001$, $\eta^2_p = .667$. The Vowel × Group interaction is significant in F1 values, $F(4, 55) = 2.801$, $p = .030$, $\eta^2_p = .183$. In order to get a deeper insight into the Vowel × Group interaction, one-way repeated measures ANOVAs were performed to test simple main effects.

Three simple main effects tests were used to see whether the F1 values of each of the six Taiwan Southern Min vowels produced by participants from the same speech group differed significantly from each other, and six simple main effects tests were applied to see whether the F1 values from the three speech groups differed for the same Taiwan Southern Min vowel.

In F1 values, a significant difference between vowels exists in the speech of LA, $F(5, 45) = 155.443$, $p < .001$, $\eta^2_p = .945$. Mean values for F1 (/i/: $M = 278$, $SD = 44$; /u/: $M = 313$, $SD = 45$; /e/: $M = 426$, $SD = 55$; /a/: $M = 448$, $SD = 53$; /ə/: $M = 538$, $SD = 54$; /ɔ/: $M = 770$, $SD = 70$) were further compared in post hoc pairwise comparisons using Bonferroni corrections for multiple comparisons. Significant differences ($p < .05$) were found between Taiwan Southern Min vowels except for those in the vowel pairs /i, u/ and /ə, ɔ/.

Table 1. Mean and standard deviation (SD) of the first, second, and third (F1, F2, and F3) formant frequencies (Hz) of Taiwan Southern Min vowels produced by 10 laryngeal speakers (LA), 10 speakers with a pneumatic artificial larynx (PA), and eight esophageal speakers (ES).

<table>
<thead>
<tr>
<th>Vowel</th>
<th>/i/</th>
<th>/e/</th>
<th>/a/</th>
<th>/u/</th>
<th>/ɔ/</th>
<th>/ə/</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 LA</td>
<td>278 (44)</td>
<td>426 (55)</td>
<td>770 (70)</td>
<td>313 (45)</td>
<td>538 (54)</td>
<td>448 (53)</td>
</tr>
<tr>
<td>PA</td>
<td>463 (94)</td>
<td>538 (58)</td>
<td>1044 (187)</td>
<td>476 (53)</td>
<td>678 (124)</td>
<td>585 (78)</td>
</tr>
<tr>
<td>ES</td>
<td>351 (57)</td>
<td>525 (78)</td>
<td>857 (102)</td>
<td>425 (83)</td>
<td>587 (95)</td>
<td>598 (82)</td>
</tr>
<tr>
<td>F2 LA</td>
<td>2212 (207)</td>
<td>2011 (116)</td>
<td>1247 (67)</td>
<td>693 (56)</td>
<td>791 (86)</td>
<td>1111 (173)</td>
</tr>
<tr>
<td>PA</td>
<td>2022 (159)</td>
<td>1930 (146)</td>
<td>1480 (165)</td>
<td>1131 (176)</td>
<td>1281 (178)</td>
<td>1362 (116)</td>
</tr>
<tr>
<td>ES</td>
<td>1825 (310)</td>
<td>1665 (305)</td>
<td>1330 (85)</td>
<td>825 (152)</td>
<td>990 (126)</td>
<td>1186 (146)</td>
</tr>
<tr>
<td>F3 LA</td>
<td>3105 (174)</td>
<td>2465 (244)</td>
<td>2378 (171)</td>
<td>2252 (289)</td>
<td>2678 (204)</td>
<td>2635 (196)</td>
</tr>
<tr>
<td>PA</td>
<td>3012 (260)</td>
<td>2501 (201)</td>
<td>2395 (281)</td>
<td>2107 (255)</td>
<td>2244 (238)</td>
<td>2404 (258)</td>
</tr>
<tr>
<td>ES</td>
<td>2955 (330)</td>
<td>2469 (210)</td>
<td>2436 (330)</td>
<td>2039 (255)</td>
<td>2396 (452)</td>
<td>2292 (421)</td>
</tr>
</tbody>
</table>
A significant difference between vowels exists in the speech of PA, \( F(2, 17) = 44.425, p < .001, \eta_p^2 = .832 \).

Mean values for F1 (/i/: \( M = 463, SD = 94 \); /a/: \( M = 476, SD = 53 \); /e/: \( M = 538, SD = 58 \); /a/: \( M = 585, SD = 78 \); /a/: \( M = 678, SD = 124 \); /a/: \( M = 1044, SD = 187 \)) were further compared in post hoc pairwise comparisons using Bonferroni corrections for multiple comparisons. Significant differences \( (p < .05) \) were found between Taiwan Southern Min vowels except for the vowels in the pair /i, u/ and the triplet /e, o, o/.

A significant difference between vowels exists in the speech of ES, \( F(5, 35) = 83.925, p < .001, \eta_p^2 = .923 \).

Mean values for F1 (/i/: \( M = 351, SD = 57 \); /a/: \( M = 425, SD = 83 \); /a/: \( M = 525, SD = 78 \); /a/: \( M = 598, SD = 82 \); /a/: \( M = 587, SD = 95 \); /a/: \( M = 857, SD = 102 \)) were further compared in post hoc pairwise comparisons using Bonferroni corrections for multiple comparisons. Significant differences \( (p < .05) \) were found between Taiwan Southern Min vowels except for the vowels in the pair /i, u/ and the triplet /e, o, o/.

Significant differences between the groups of speakers exist in all the vowels—/i/: \( F(2, 25) = 18.037, p < .050, \eta_p^2 = .591 \); /a/: \( F(2, 25) = 9.120, p < .001, \eta_p^2 = .422 \); /a/: \( F(2, 25) = 11.318, p < .001, \eta_p^2 = .475 \); /a/: \( F(2, 25) = 18.938, p < .001, \eta_p^2 = .602 \); /a/: \( F(2, 25) = 5.484, p < .011, \eta_p^2 = .305 \); /a/: \( F(2, 25) = 13.003, p < .001, \eta_p^2 = .510 \). Mean values for F1 were further compared in pairwise comparisons using Bonferroni’s adjustment for multiple comparisons. LA and PA differ significantly \( (p < .05) \) in all the vowels: /i/ (LA: \( M = 278, SD = 44 \); PA: \( M = 463, SD = 94 \)), /a/ (LA: \( M = 313, SD = 45 \); PA: \( M = 476, SD = 53 \)), /a/ (LA: \( M = 426, SD = 56 \); PA: \( M = 538, SD = 58 \)), /a/ (LA: \( M = 448, SD = 53 \); PA: \( M = 585, SD = 78 \)), /a/ (LA: \( M = 538, SD = 54 \); PA: \( M = 678, SD = 124 \)), and /a/ (LA: \( M = 770, SD = 70 \); PA: \( M = 1044, SD = 187 \)). LA and ES differ significantly \( (p < .05) \) in /a/ (LA: \( M = 313, SD = 45 \); ES: \( M = 425, SD = 83 \)), /a/ (LA: \( M = 426, SD = 55 \); ES: \( M = 525, SD = 78 \)), and /a/ (LA: \( M = 448, SD = 53 \); ES: \( M = 598, SD = 82 \)). PA and ES differ significantly \( (p < .05) \) in /i/ (PA: \( M = 463, SD = 94 \); ES: \( M = 351, SD = 57 \)) and /a/ (PA: \( M = 1044, SD = 187 \); ES: \( M = 857, SD = 102 \)).

**Second Formant**

A two-way mixed-design ANOVA was conducted with group (LA, PA, and ES) as the between-subjects factor and vowel (/i, e, a, o, u/) as the within-subject factor. Table 1 reports the means and standard deviations of F2 for each Taiwan Southern Min vowel from the different speech groups. A significant difference between vowels exists in the F2 values, \( F(3, 74) = 241.585, p < .001, \eta_p^2 = .906 \), and a significant difference between groups exists in the F2 values, \( F(2, 25) = 20.690, p < .001, \eta_p^2 = .623 \). The Vowel × Group interaction is significant in the F2 values, \( F(6, 74) = 9.69, p < .001, \eta_p^2 = .437 \). In order to get a deeper insight into the Vowel × Group interaction, one-way repeated measures ANOVAs were performed to test simple main effects.
In F2 values, a significant difference between vowels exists in the speech of LA, F(3, 23) = 238.419, p < .001, \( \eta_p^2 = .964 \). Mean values for F2 (/i: M = 693, SD = 56; /a:/ M = 761, SD = 36; /u:/ M = 1111, SD = 173; /a: M = 1247, SD = 67) were further compared in post hoc pairwise comparisons using Bonferroni corrections for multiple comparisons. Significant differences (p < .05) were found between Taiwan Southern Min vowels except for the vowel pairs /i, e/, /u, \( \bar{a} \)/, and /a, /a/. A significant difference between vowels exists in the speech of PA, F(3, 25) = 67.491, p < .001, \( \eta_p^2 = .882 \). Mean values for F2 (/u: M = 1131, SD = 176; /a:/ M = 1281, SD = 178; /l: M = 1362, SD = 116; /a: M = 1480, SD = 165; /e: M = 1930, SD = 146; /i: M = 2022, SD = 159) were further compared in post hoc pairwise comparisons using Bonferroni corrections for multiple comparisons. Significant differences (p < .05) were found between the speech groups. A significant difference between vowels exists in ES, F(2, 14) = 29.860, p < .001, \( \eta_p^2 = .810 \). Mean values for F2 (/u: M = 825, SD = 152; /a:/ M = 990, SD = 126; /l: M = 1186, SD = 146; /a: M = 1330, SD = 85; /e: M = 1665, SD = 305; /i: M = 1825, SD = 310) were further compared in post hoc pairwise comparisons using Bonferroni corrections for multiple comparisons. Significant differences (p < .05) were found between Taiwan Southern Min vowels except for the vowels in the pairs /i, e/, /u, \( \bar{a} \)/, and /a, /a/. A significant difference between vowels exists in ES, F(2, 14) = 41.024, p < .001, \( \eta_p^2 = .621 \), and a significant group difference exists in F3 values, F(2, 25) = 2.846, p < .050, \( \eta_p^2 = .185 \). The Vowel × Group interaction is not significant in F3 values, F(7, 86) = 9.69, p > .050, \( \eta_p^2 = .126 \). One-way repeated measures ANOVAs were not performed to test simple main effects because the Vowel × Group interaction is not significant in F3.

**Third Formant**

A two-way mixed-design ANOVA was conducted with group (LA, PA, and ES) as the between-subjects factor and vowel (/i, e, a, \( \bar{a} \), u, /a/) as the within-subject factor. Table 1 reports the means and standard deviations of F3 for each Taiwan Southern Min vowel from the different speech groups. A significant vowel difference exists in F3 values, F(3, 86) = 41.024, p < .001, \( \eta_p^2 = .621 \), and a significant group difference exists in F3 values, F(2, 25) = 2.846, p < .050, \( \eta_p^2 = .185 \). The Vowel × Group interaction is not significant in F3 values, F(7, 86) = 9.69, p > .050, \( \eta_p^2 = .126 \). One-way repeated measures ANOVAs were not performed to test simple main effects because the Vowel × Group interaction is not significant in F3.

**Vowel Space: ED**

The three corner vowels of Taiwan Southern Min, /i, a, /a/, were used to define the vowel trilateral in the vowel space and illustrate maximal acoustic distinctiveness. ED was used to measure numerically the separation between corner vowels in Taiwan Southern Min and to see whether the distance between any two designated corner vowels varied significantly across the three speaker groups. The ED values for the vowel pairs /i, u/, /i, a/, and /a, u/ produced by LA, ES, and PA are shown in Table 2.

One-way ANOVA was performed. A significant difference between the groups is found in the vowel pairs /i, u/, F(2, 25) = 18.057, p < .050, and /i, a/, F(2, 25) = 9.724, p < .050. Mean values of ED for the vowel pairs /i, a/, and /a, u/ were further compared in post hoc pairwise comparisons using Fisher’s least significant difference for multiple comparisons. Again, for the vowel pairs /i, a/, and /a, u/, significant differences (p < .05) are found between LA and PA and between LA and ES.

**Vowel Space: Area**

The lengths of the three sides of the VSAs—representing the ED values for the vowel pairs /i, a/, /i, /a/, and /a, u/ produced by laryngeal speakers (LA), speakers with a pneumatic artificial larynx (PA), and esophageal speakers (ES) and their /i, a, /a/ vowel space areas.

Table 2. Mean and standard deviation (SD) of Euclidean distance values (in Hz) for the vowel pairs /i, u/, /i, /a/, and /a, u/ produced by laryngeal speakers (LA), speakers with a pneumatic artificial larynx (PA), and esophageal speakers (ES) and their /i, a, /a/ vowel space areas.

<table>
<thead>
<tr>
<th>Participant group</th>
<th>Vowel pair</th>
<th>Vowel space area</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA</td>
<td>1520 (170)</td>
<td>362934 (72823)</td>
</tr>
<tr>
<td>PA</td>
<td>1480 (165)</td>
<td>271684 (124414)</td>
</tr>
<tr>
<td>ES</td>
<td>1300 (152)</td>
<td>245815 (91961)</td>
</tr>
</tbody>
</table>
from LA, PA, and ES. The statistical result suggests that LA, PA, and ES differ significantly in their VSA, $F(2, 25) = 3.704, p = .039$. Post hoc comparisons using Fisher’s least significant difference test indicated that LA ($M = 362934, SD = 72823$) and PA ($M = 271684, SD = 124414$), and LA ($M = 362934, SD = 72823$) and ES ($M = 245815, SD = 91961$), differ significantly in VSA. However, PA ($M = 271684, SD = 124414$) and ES ($M = 245815, SD = 91961$) do not differ significantly in VSA.

Discussion

Formants

This study provides acoustic information on the vowel production of PA and ES. Our results are in accordance with previous studies showing that alaryngeal speakers tend to have higher F1 or F2 values compared with LA (see Table 1). Studies have suggested that the higher formant frequencies of alaryngeal speakers might result from the fact that they have shorter vocal tracts after their total laryngectomy (Christensen & Weinberg, 1976; Sisty & Weinberg, 1972). Source filter theory and evidence from physiological studies have been proposed to provide the explanation for these results (Damsté & Lerman, 1969; Isman & O’Brien, 1992; Ng & Chu, 2009). Source filter theory assumes that the length of the vocal tract and its cross-sectional area are the determinants of the resonance frequencies. Moreover, formant frequencies are inversely related to the length of the vocal tract. Evidence from the physiological studies suggests that the back of the tongue is connected with the esophagus after the laryngeal structures have been removed. This may lead to the shift of the PE segment from between the fourth and fifth cervical vertebrae to between the third and fourth (Bentzen et al., 1976; Damsté & Lerman, 1969; Diedrich & Youngstrom, 1966; Kytta, 1964; van As et al., 2001). Furthermore, ES use the PE segment to vibrate the sound source. The upward displacement of the PE segment from between the fourth and fifth cervical vertebrae to the level of the second to third cervical vertebrae is possible when ES articulate (Diedrich & Youngstrom, 1966; Ng & Chu, 2009). This explains the shortening of the vocal tract during phonation.

In general for LA, the frequency of F1 is related to oral and pharyngeal constrictions. The more open the mouth or the more constricted the pharynx, the higher the F1 frequency. F2 values are raised by a front tongue constriction and lowered by a back tongue constriction (Fant, 1970; Kent & Read, 2002; Peterson & Barney, 1952). However, this association between articulation and acoustics in the speech of LA might be different for alaryngeal speakers. Alaryngeal speakers have changes in their anatomy and physiology resulting from their total laryngectomy. The size and extent of resection into the pharynx, and the reconstructed shape and length of the PE segment, vary depending on the surgical alterations (Casper & Colton, 1993). Furthermore, the extent of the upward displacement of the PE segment during phonation varies among ES depending on the length of the PE segment, the speech-testing materials, and the point in the speech when the measurement was taken (Isman & O’Brien, 1992). For PA, 1.0 to 1.5 in. of the tube of the pneumatic artificial larynx is placed into the corner of the mouth and directed up toward the roof of the mouth. The tube of the pneumatic artificial larynx of PA may confine the movement of the tongue or jaws. It is therefore uncertain to what degree F1 and F2 correlate with the extent of constriction of the pharynx, front tongue, or back tongue in alaryngeal speakers.

There is a many-to-one association between articulation and acoustics. In other words, the same set of F1 to F3 values can be obtained from vocal tracts of various shapes (Atal, Chang, Mathews, & Tukey, 1978; Gay, Lindblom, & Lubker, 1981). Participants with a fixed mandible are able to produce vowels with formant frequency patterns showing the ranges of variation of those produced by typical speakers (Lindblom, Lubker, & Gay, 1979). When steady-state vowels are produced both normally and with a bite block, the formant patterns of the bite-block vowels are similar to those of the naturally spoken vowels (Gay et al., 1981). Compensatory articulation of fixed-mandible vowels or bite-block vowels is possible to facilitate a listener’s comprehension (Gay et al., 1981; Lindblom et al., 1979). On the basis of the above-mentioned findings, two possible interpretations can be derived from the formant frequencies of the vowels of PA and ES.

First, PA and ES have the normal articulatory patterns of LA. The formant patterns of each Taiwan Southern Min vowel produced by PA and ES correlate with tongue height or tongue advancement, just like those of LA. However, in addition to the anatomical and physical changes that PA and ES have undergone, their motor control of the PE segment is not comparable with the motor control capability of LA (Moon & Weinberg, 1987). In alaryngeal speech, therefore, how formant patterns correlate with tongue height or tongue advancement is uncertain.

As an alternative, PA and ES may have abnormal articulation patterns, with formant frequencies resulting from compensatory articulation intended to facilitate their listeners’ comprehension. In that case, the correlation between alaryngeal speakers’ formant frequencies and the degree of tongue height and tongue advancement is not as significant as it is with LA. One of these two interpretations from our results may be confirmed by future studies of the myoelastic properties of the esophageal sphincter and the effect of the tube of the pneumatic artificial larynx in the vocal tract during phonation (Moon & Weinberg, 1987; Sisty & Weinberg, 1972; van As et al., 2001).

Following this line of argument, speech production is listener oriented. If the speech production of PA and ES is as intelligible as that of LA, we would expect PA and ES to have formant patterns in each Taiwan Southern Min vowel similar to those of LA. In their F1 values, those vowels differ from each other except for those in the pairs /i, u/ and /a, o/ in the speech of LA. Taking the frequency pattern in F1 of vowels from LA as a base, we find similar results in PA and ES. Taiwan Southern Min vowels from PA and ES also differ from each other in their F1 values.
except in the pairs /i, u/ and /ə, ə/. However, in the speech of PA and ES, the midvowel /e/ does not differ from the other two midvowels /ə, ə/ in their F1 values. PA and ES underwent total laryngectomy; they have similar anatomical changes in their vocal tract. Formant frequencies are acoustic cues for speech perception. PA and ES do not provide the acoustic cue to distinguish /e/ from /ə, ə/, which implies that their limited articulatory capabilities prevent them from making the finer adjustment in their vocal tract that would enable them to make this distinction.

In their of F2 values, LA and PA have similar formant patterns in Taiwan Southern Min vowels. Their Taiwan Southern Min vowels differ significantly from each other except for those in the pairs /i, e/, /u, ə/, and /ə, ə/. However, PA additionally do not distinguish the vowels in the pairs /ə, ə/ and /i, e/, /u, ə/, /ə/ is a midback vowel. The difficult articulatory adjustment for PA is distinguishing the midback /ə/ from its neighbors, low vowel /a/ and middle vowel /a/. ES, in addition to their undifferentiated vowels in the pairs /i, e/ and /u, ə/, do not distinguish vowels in the vowel pairs /i, e/, /a, e/, and /a, e/.

The formant patterns (the distances between F1 and F2) for each vowel produced by LA, PA, and ES are systematically related (see Figures 4, 5, and 6). The results also indicate that the relative positions of the vowels are generally maintained (see Figures 7 and 8). On this point, researchers (Miller, 1989; Pickett, 1999) proposed that vowel identity depends on the intervals between formants rather than absolute formant values, which is how the speaker can be understood. This may explain why the overall speech intelligibility of PA and ES in reading the vowels in the testing materials was still judged to be at least acceptable (as shown in Appendix A) even though PA and ES differ from LA in separating /e/ from /ə, ə/ in terms of F1 values and PA and ES differ from LA in separating vowels in the pairs /i, e/, /ə, ə/, /i, e/, /u, ə/, and /ə, ə/ in terms of F2 values.

**VSA**

Acoustic VSA is a quantitative measure that requires plotting the corner vowels in an F1/F2 plane; VSA therefore reflects the extent of vocal tract movements in the production of corner vowels. However, this procedure has not been applied to PA and ES in Taiwan. The VSA within the trilateral formed by the three Taiwan Southern Min corner vowels, /i, a, u/, is used as an indicator to determine the extent of tongue height (F1) and tongue advancement (F2) for LA. For the alaryngeal speakers, articulatory working space in the vocal tract is changed after the total laryngectomy surgery. The VSA found in PA and ES is reduced because those speakers have a significantly shorter distance between /i/ and /a/ and between /i/ and /u/ compared with LA (see Table 2 and Figure 8). The results are in accordance with the previous studies related to partial glossectomy (Whitehill et al., 2006) and total laryngectomy (H. Liu & Ng, 2009) that show reduced VSA. However, there is no difference in the distance in the nonfront vowel pair, /a, u/, among LA, PA, and ES. The VSA of ES is located almost entirely inside the vowel space of LA (see Figure 8). The VSA of PA shifts away from that of LA toward higher frequencies. The VSA of PA and ES is significantly smaller than that of LA.

VSA has been used as an index of the accuracy of vowel articulation, which signifies gross motor control of the tongue and jaw coordination. Studies have reported a positive correlation between speech intelligibility and the size of VSA formed by the corner vowels on the F1/F2 coordinates for typical and atypical speakers. Bradlow, Torretta, and Pisoni (1996) found that the intelligibility of key words in Harvard sentences (Fisher, Doddington, & Goudie-Marshall, 1986) produced by 10 men and 10 women...
was correlated with the vowel space dispersion and the F1 range. Differences in F2 between /i/ and /u/ in British English correlated significantly with word intelligibility (Hazan & Markham, 2004). Typical talkers who have larger VSAs are judged to be more intelligible than talkers who have smaller VSAs (Bradlow et al., 1996). Clear speech uses a significantly larger speech area than conversational speech when the two kinds of speech are uttered at a typical speaking rate (Krause & Braida, 2004). Speakers with amyotrophic lateral sclerosis or cerebral palsy and children with dysarthria have a restricted VSA, and there is a significant correlation between measures of intelligibility and vowel space (Higgins & Hodge, 2001; H.-M. Liu et al., 2005; Weismer et al., 2001). VSA has potential clinical applications for providing descriptive insight into the population of laryngectomees. This study did not investigate the relationship between VSA and speech intelligibility. Future studies may clarify the relationship between alaryngeal speakers’ VSA and their speech intelligibility for clinical application.

Conclusions

This study examined the acoustic properties of six Taiwan Southern Min vowels produced by 10 LA, 10 PA, and eight ES. The lowest three formants were measured for six vowels. In accordance with previous studies, alaryngeal speakers tend to have higher formant frequencies in F1, F2, or both compared with LA in each vowel. Source filter theory assumes that the resonance frequencies are inversely related to the length of the vocal tract. Alaryngeal speakers have a shortened vocal tract because of their total laryngectomy. Therefore, higher formant frequencies have been associated with the shortened vocal tract of PA and ES after total laryngectomy. The VSA of PA and ES is significantly smaller than that of LA, which implies that the speech of PA and ES is less readily intelligible than that of LA. This speculation needs further support from future study.

Acknowledgments

This research was supported by Research and Development Division at Chung Shan Medical University Grant F1010024. We thank Nan-Mei Wang for her valuable input on the initial plan of this study. We are grateful to the Association of Laryngectomy in Taiwan and to Hsiu-jin Lai, a speech pathologist, for helping to recruit participants.

References


Appendix A
Ratings of Different Speech Parameters by Participant Group

<table>
<thead>
<tr>
<th>Speech parameter</th>
<th>LA 1 2 3 4 5 6 7 8 9 10</th>
<th>PA 1 2 3 4 5 6 7 8 9 10</th>
<th>ES 1 2 3 4 5 6 7 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice quality</td>
<td>7 7 7 7 7 7 6 7</td>
<td>7 6 7 5 6 5 4 4 6 5</td>
<td>6 4 5 6 6 5 4 4</td>
</tr>
<tr>
<td>Articulation proficiency</td>
<td>7 7 7 7 7 7 6 7</td>
<td>7 5 7 6 5 5 4 6 5</td>
<td>6 4 5 7 6 5 4 4</td>
</tr>
<tr>
<td>Quietness of speech</td>
<td>7 7 7 7 7 7 7 7</td>
<td>7 5 6 4 6 5 3 3 5 4</td>
<td>4 2 3 7 5 4 3 4</td>
</tr>
<tr>
<td>Pitch variability</td>
<td>7 7 7 7 7 7 7 6 7</td>
<td>7 6 6 5 6 5 6 4 6 4</td>
<td>6 4 4 7 5 4 4 3</td>
</tr>
<tr>
<td>Overall speech intelligibility</td>
<td>7 7 7 7 7 7 7 7 6 7</td>
<td>7 5 6 5 5 4 5 6 4</td>
<td>6 4 4 7 5 4 4 4</td>
</tr>
</tbody>
</table>

Note. The ratings scale is from 1 to 7 (7 = best speech production, 1 = worst speech production). The speech parameters are adopted from Ng et al. (1997). LA = laryngeal speakers; PA = speakers with a pneumatic artificial larynx; ES = esophageal speakers. LA and PA participants are each numbered one to ten. ES participants are numbered one to eight.

Appendix B
Examples of Different Taiwan Southern Min Tones With Syllable /tcian/

<table>
<thead>
<tr>
<th>Tone pitch</th>
<th>Southern Min character</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>High and level</td>
<td>真</td>
<td>Real</td>
</tr>
<tr>
<td>High to low</td>
<td>診</td>
<td>To examine a patient</td>
</tr>
<tr>
<td>Low</td>
<td>遺職</td>
<td>To move forward</td>
</tr>
<tr>
<td>Middle short</td>
<td>職</td>
<td>A position or duty</td>
</tr>
<tr>
<td>Low rising</td>
<td>烏</td>
<td>The Chin dynasty</td>
</tr>
<tr>
<td>Middle flat</td>
<td>煙</td>
<td>Exhausted or utmost</td>
</tr>
<tr>
<td>High short</td>
<td>一</td>
<td>One</td>
</tr>
</tbody>
</table>


