An Acoustical and Perceptual Study of Vowels Produced by Alaryngeal Speakers of Cantonese

Manwa L. Ng Rhoda Chu

Speech Science Laboratory, Division of Speech and Hearing Sciences, University of Hong Kong, Hong Kong, SAR, China

Key Words

Acoustical analysis · Perceptual analysis · Cantonese · Alaryngeal speakers · Formant

Among the different alaryngeal speaker groups, naïve listeners tended to perceive the vowels produced by electro-laryngeal speakers more accurately.

Copyright © 2009 S. Karger AG, Basel

Abstract

Objective: The present study determines the acoustical and perceptual differences between alaryngeal and laryngeal speakers of Cantonese in the production of Cantonese vowels. **Subjects and Methods:** Formant frequencies (F1 and F2) associated with eight Cantonese vowels (/i, y, ε, a, ɐ, ɔ, œ, u/) produced by 40 alaryngeal (10 esophageal, 10 tracheoesophageal, 10 with pneumatic artificial larynx, 10 electro-laryngeal) and 10 laryngeal speakers were obtained. To assess how formants affect vowel perception, the vowels were identified by 20 naïve listeners, based on which confusion matrices were established. **Results:** Spectral analysis revealed that F1 and F2 values of Cantonese vowels produced by the alaryngeal speakers were significantly higher than those produced by laryngeal speakers, with the exception of F2 of /ε/, /i/ and /y/ produced by alaryngeal speakers using pneumatic artificial larynx, and /ε/ produced by electro-laryngeal speakers. **Conclusion:** The acoustic findings suggest a general shortening of the effective vocal tract length for resonance in alaryngeal speakers after laryngectomy. Percent correct identification of vowels from the perceptual experiment indicated similar error patterns in vowel identification between alaryngeal and laryngeal speaker groups.

Learning to use an alternative phonation method in order to regain verbal communication is a crucial part of postlaryngectomy speech rehabilitation. Currently, four types of alaryngeal phonation are being adopted by laryngectomees in Hong Kong. They are esophageal (SE), tracheoesophageal (TE), electrolaryngeal (EL) and pneumatic artificial larynx (PA) speech. In SE and TE speech, the pharyngoesophageal (PE) segment functions as a new sound source, the neoglottis. Upon phonation, the PE segment is set into vibration by passing air through it. Despite the use of a similar sound source, SE and TE speakers utilize different air reservoir mechanisms. For SE speakers, air is stored in the upper esophagus. Upon phonation, positive intraesophageal pressure pushes the air upwards and vibrates the PE segment. In TE phonation, pulmonary air expelled from the lungs vibrates the neoglottis [1]. Unlike SE and TE phonation, EL and PA speakers rely on an external device for sound generation. EL speakers make use of a transcervical electrolarynx which is placed on the anterolateral neck. In PA phonation, sound is made by setting the rubber reed located inside the pneumatic artificial larynx into vibration using egressive pulmonary air [2]. In

all types of alaryngeal phonation methods, sound generated by the neoglottis is shaped by the supralaryngeal articulators, which yields different speech sounds.

Due to the use of an alternative sound source, alaryngeal speakers' voice quality is no longer the same as that of laryngeal speakers. Alaryngeal voice quality has been of interest to many researchers, especially the voice quality of SE and TE speech. Studies have reported various aspects of voice associated with different types of alaryngeal phonation, including perceptual characteristics, intensity, rate of speech, voice pitch, and aerodynamics [3–11]. Alaryngeal speech of a tone language has also been investigated, particularly in Thai [12, 13], Mandarin [14–16], and Cantonese [17–19]. It is generally agreed that, using the PE segment as the neoglottis, SE and TE speakers show reduced voice pitch and pitch variation, reduced loudness and range of loudness, reduced rate of speech, elevated jitter, shimmer, and noise level. Robbins et al. [6] attributed the unique SE and TE voice quality to the diminished control of PE segment vibration.

Compared to SE and TE speech, the voice quality of EL and PA speech has only been studied to a limited extent. As EL and PA sound quality is largely determined by the devices being used, it is beyond the control of EL and PA speakers. Yet, both have been described as robotic, monotonic, and lacking in pitch and loudness variations [e.g., 8, 9, 19, 20]. Perhaps because of cultural differences, PA speech is not prevalent among English-speaking alaryngeal communities.

Comparing the voice quality associated with different types of alaryngeal speech, the possible change in vocal tract resonance characteristics after laryngectomy has received scant attention. According to the source-filter theory of vowel production [21], speech sounds are produced as a product of two components: sound source and vocal tract filter. Theoretically applicable to both laryngeal and alaryngeal phonation, the theory assumes the independence between the sound source and the supralaryngeal vocal tract filter. This assumption implies that, for alaryngeal speakers, the removal of the laryngeal source should not affect the filtering properties of the supralaryngeal vocal tract filter. By adopting a new sound source, alaryngeal speakers should be able to produce vowels in a way similar to laryngeal speakers. This suggestion is supported by a number of studies of alaryngeal vowel production. Schilling et al. consistently revealed little or no difference in the formant frequencies of German vowels produced by SE and NL speakers, and similar findings were also reported for Swedish alaryngeal speakers by Luchsinger, as cited in van As et al. [22].

However, contradictory findings were also documented. Kytta [23] reported higher formant values in all Finnish vowels produced by SE speakers compared with NL speakers, with the exception of F1 of /u, o, e/. When examining the vowel production among English-speaking laryngectomees, Sisty and Weinberg [24] revealed that formant frequencies were consistently higher in the English vowels produced in [hVd] context by SE and TE speakers when compared with NL speakers. Van As et al. [22] investigated the formant characteristics of Dutch vowels produced by TE speakers. They found that when TE and NL speakers produced nonsense syllables, they demonstrated comparable F1 values in vowels produced in nonsense syllables, but F2 values were higher for /a, i/ and lower for /u/ in TE speech than in NL speech. During a reading task the TE speakers consistently produced vowels with significantly higher F1 and F2 values compared to NL speakers. Cervera et al. [25] investigated the acoustic characteristics of Spanish vowels produced by SE and TE speakers. Significantly higher formant values were found in alaryngeal speech than in NL speech. Similar results were also reported in a recent study of Mandarin SE speech by Liu and Ng [20]. Vowels produced by SE speakers of Mandarin showed higher formant frequencies.

The discrepant findings regarding vowel formants associated with SE and TE speech may be related to the limitations associated with these studies, including a wide variety of languages being studied, and use of small sample sizes. In addition, past studies focused only on SE and TE speech in which the neoglottis is the PE segment. A systematic study of vowel formants produced by EL and PA speakers is still lacking. Possible changes to formant structures of vowels produced by EL and PA speakers are not known. In addition, data regarding the relationship between formant frequency patterns and perceptual quality of vowels produced by alaryngeal speakers are not presently available. Accordingly, the present study examined the vowel formants associated with different types of alaryngeal speech compared to laryngeal speech. F1 and F2 values were obtained from vowels produced by EL, PA, SE and TE speakers of Cantonese.

Methods

Participants

Forty alaryngeal and 10 laryngeal speakers participated in the study. The alaryngeal speakers included 10 EL, 10 PA, 10 SE, and 10 TE speakers who were carefully selected from the New Voice Club of Hong Kong by speech-language pathologists. The alaryn-

geal speakers were physically healthy and judged as superior speakers in their respective method of alaryngeal phonation by the speech-language pathologists. All participants were adult males who were native speakers of Cantonese. They were age-matched, and assessed with no known history of speech and hearing problems, except those associated with laryngectomy for the alaryngeal speakers.

The EL speakers were between 49 and 79 years old (mean age 65.3 years), with a duration of using EL speech between 3.25 and 22 years. The PA speakers were between 51 and 82 years of age (mean age 69.0 years), with a duration of using PA speech between 1 and 17 years. The SE speakers were 47–75 years of age (mean age 64.8 years), with a duration of using SE speech ranging from 3 months to 18 years. The TE speakers were 49–84 years old (mean age 71.2 years), with a duration of practicing TE speech ranging from 2 to 15 years, with a mean time of 5.75 years. The NL speakers were between 56 and 75 years old (mean age 61.8 years).

Procedures

Speech materials consisted of eight Cantonese vowels /i, y, ɛ, a, ɐ, ɔ, œ, u/. The participants were instructed to read aloud the vowels in a carrier phrase /ŋɔ̌ jiu tak _ pei nei t^hɛŋ/ ('I want to read _ to you'). With such a carrier phrase, the target vowel was preceded by a stop consonant /k/, and followed by another stop consonant /p/. This enabled the investigator to clearly and easily identify the target vowel in the waveform during analysis. Each vowel was produced three times. The order in which the vowels were produced was randomized. The speech samples were recorded by using a high-quality microphone. The acoustic signals were digitized at 20 kHz and 16 bits/sample, and recorded by using an MD recorder (Sony, MZN710). Before the recording took place, the participants were provided with an opportunity to practice in order to familiarize themselves with the recording format and the speech materials. To ensure the naturalness of productions, the speakers were instructed to produce the speech samples at a comfortable loudness and pitch level.

Acoustic Analysis

Measurement of formant frequencies was accomplished using a combination of waveform and spectrographic displays (Praat, version 5.0.05). Formant values were estimated using linear prediction coding (LPC) algorithm. A 25-ms window was placed at the mathematical midpoint of each vowel and Praat automatically provided the corresponding F1 and F2 frequency values.

Perceptual Analysis

Twenty naïve listeners who had no prior experience with alaryngeal speech of any type were recruited for the listening task. The listeners were adult native Cantonese speakers with no known history of hearing problems. The recorded speech samples produced by laryngeal and alaryngeal speakers were mixed and presented to the listeners in a randomized order at a comfortable loudness level (approximately 65–70 dB) via high-quality headphones. Prior to the experiment, the listeners were provided with answer sheets on which the IPA symbols and Chinese characters representing the eight vowels were printed. The listeners were instructed to choose the vowel they think they have perceived upon listening to each stimulus. An interstimulus interval of approximately 3 s was provided. The listeners were given practice time to familiarize themselves with the stimuli and the format of the lis-

tening task. The results of perceptual and acoustic analysis were compared, and were used to reflect the relationship between the acoustic properties and the perception of Cantonese vowels.

Reliability

To evaluate for intrarater and interrater reliability, 20% of the speech samples were randomly selected from the entire data corpus from which F1 and F2 values were analyzed again by the investigator and a second investigator. The results were compared with the first F1 and F2 measurements. Pearson product-moment correlation coefficients (*r*) were used to reflect the intrarater and interrater reliability of F1 and F2 measurements. For intrarater reliability, the average absolute error and Pearson product-moment correlation coefficient of F1 measurements were 7.76 Hz and 0.998, respectively, whereas those of F2 were 25.05 Hz and 0.992, respectively. For interrater reliability, the average absolute error and Pearson product-moment correlation coefficient of F1 measurements are 17.29 Hz and 0.974, respectively, whereas those of F2 were 48.25 Hz and 0.954, respectively. Pearson product-moment correlation coefficients of the intrarater and interrater reliability of perceptual analysis were 0.801 and 0.835, respectively.

Results

Acoustic Analysis

The mean and standard deviation values of F1 and F2 of the Cantonese vowels produced by the five speaker groups are presented in table 1. In general, the formant frequency values associated with all alaryngeal speakers were higher than those with the laryngeal speakers, with the exception of F2 of /ɛ/, /i/ and /y/ for PA speakers, and F2 of /ɛ/ for EL speakers. For all eight Cantonese vowels, SE and TE speakers consistently demonstrated higher F1 and F2 values than NL speakers. This indicates that, with the use of a PE segment as the neoglottis, SE and TE speakers consistently produced Cantonese vowels with higher formant frequencies than NL speakers. The F1 values associated with EL and PA speech were consistently higher than that of NL speech. Except for the vowels /i/, /y/, and /ɛ/, EL and PA speakers exhibited greater F2 values than NL speakers.

The present study aimed at examining the difference in formants among different speaker groups. No attempt was made to compare the difference among different vowels. For each of the vowels, one-way Analyses of Variance (ANOVAs) were performed to assess for significant differences in F1 and F2 values among different speaker groups. Results revealed significant main effects among speaker groups for both F1 and F2 of all vowels. Subsequent Tukey post-hoc *t* tests were carried out to determine the differences among speaker groups for F1 and F2 values. The *t* tests for F1 frequency revealed clear differ-

Fig. 1. Vowel spaces of the corner vowels /i/, /a/, and /u/ produced by NL, EL and PA speakers (a) and NL, SE and TE speakers of Cantonese (b).

Table 1. Mean and standard deviation values of F1 and F2 frequency of Cantonese vowels produced by five groups of speakers

	Formant frequency, Hz															
	/i/		/y/		/ɛ/		/ɑ/		/ɐ/		/ɔ/		/œ/		/u/	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
NL	257 (23)	2,254 (143)	264 (26)	1,978 (237)	526 (53)	1,901 (179)	795 (54)	1,248 (89)	743 (55)	1,381 (118)	545 (54)	831 (117)	512 (55)	1,528 (87)	393 (55)	908 (485)
EL	769 (226)	2,251 (352)	771 (227)	2,145 (350)	705 (66)	1,812 (313)	999 (134)	1,409 (184)	967 (99)	1,514 (143)	726 (73)	931 (100)	691 (80)	1,686 (174)	658 (98)	1,455 (612)
PA	511 (218)	1,802 (238)	573 (308)	1,797 (186)	634 (134)	1,848 (180)	911 (112)	1,319 (99)	875 (153)	1,423 (118)	670 (62)	1,186 (196)	607 (147)	1,695 (219)	580 (97)	1,267 (209)
SE	383 (113)	2,357 (322)	339 (69)	2,211 (311)	592 (80)	2,135 (201)	898 (130)	1,364 (181)	850 (157)	1,564 (174)	620 (70)	929 (153)	565 (78)	1,735 (267)	457 (73)	1,067 (540)
TE	397 (316)	2,366 (471)	377 (209)	2,130 (355)	553 (81)	2,100 (212)	795 (104)	1,358 (160)	775 (124)	1,486 (196)	528 (101)	924 (235)	530 (87)	1,650 (182)	450 (142)	999 (336)

The corresponding standard deviation is shown in parentheses.

ences for the production of each vowel across the speaker groups ($p < 0.05$). The t tests for F2 frequency indicated that the PA group showed the greatest difference in vowel production compared to the other groups ($p < 0.05$).

Vowel spaces constructed based on the F1 and F2 values of the three corner Cantonese vowels /i/, /u/, and /a/ produced by the speaker groups are shown in figure 1. Examination of the figure reveals a striking similarity

among the vowel spaces of NL, SE and TE speech, whereas PA and EL speech demonstrated greater deviations from NL speech. Among the three quantal vowels /i/, /u/, and /a/ produced by NL, PA and EL speakers, the vowel productions for /a/ were found to be the closest to each other. The largest deviation was seen for the vowel /i/. The vowel /u/ produced by PA and EL speakers was found closer to /a/ than to /u/ produced by NL speakers.

Table 2. Confusion matrix of identification of vowels produced by laryngeal and alaryngeal speakers of Cantonese (percent correct identification)

Vowels perceived	Speakers	Vowels presented							
		/i/	/y/	/ɛ/	/a/	/ɐ/	/ɔ/	/œ/	/u/
/i/	NL	95.83	10.17	0.17	0.00	0.00	0.00	0.17	0.50
	EL	88.83	19.33	0.50	0.17	0.00	0.00	0.17	0.33
	PA	53.17	10.17	0.33	0.00	0.17	0.00	0.00	0.67
	SE	72.00	22.67	9.50	0.00	0.17	0.17	0.33	1.67
	TE	72.17	13.83	21.83	0.00	0.00	0.00	0.17	2.00
/y/	NL	3.67	88.33	0.50	0.00	0.00	0.00	0.67	0.00
	EL	10.50	78.50	0.00	0.00	0.00	0.33	0.17	0.00
	PA	43.17	79.00	0.67	0.00	0.00	0.00	1.33	1.00
	SE	8.33	60.50	3.17	0.00	0.00	0.00	6.33	10.83
	TE	6.00	62.33	0.17	0.00	0.00	0.33	0.50	19.33
/ɛ/	NL	0.17	0.00	88.00	0.00	0.33	0.00	13.00	0.00
	EL	0.17	0.50	81.00	0.17	0.00	0.17	15.17	0.33
	PA	1.83	1.17	65.83	0.17	0.00	0.67	21.50	0.67
	SE	0.17	0.33	77.83	16.83	2.33	2.33	20.83	0.83
	TE	0.17	0.00	64.50	12.50	0.67	1.00	11.83	0.17
/a/	NL	0.00	0.33	0.00	99.33	0.00	5.50	0.17	0.00
	EL	0.00	0.17	1.50	95.00	1.50	4.00	0.00	0.00
	PA	0.33	0.17	4.00	86.67	0.83	8.17	4.00	0.00
	SE	1.17	0.00	0.00	70.33	18.67	9.67	0.17	0.00
	TE	0.17	0.17	0.17	82.50	22.67	4.67	0.67	0.17
/ɐ/	NL	0.17	0.00	0.33	0.50	99.17	0.00	0.33	0.00
	EL	0.17	0.17	0.17	4.00	98.33	0.17	0.00	0.17
	PA	0.50	0.17	0.50	3.33	98.17	0.17	0.50	0.33
	SE	0.67	14.83	2.00	11.67	78.00	2.17	2.33	0.33
	TE	2.33	17.67	0.33	1.17	75.67	0.67	0.50	0.17
/ɔ/	NL	0.00	0.17	0.17	0.00	0.00	93.17	0.00	0.50
	EL	0.00	0.00	1.83	0.17	0.17	94.17	0.00	2.17
	PA	0.00	0.00	1.50	1.50	0.17	84.67	4.33	14.67
	SE	16.00	0.00	0.33	0.17	0.00	70.50	2.50	8.33
	TE	17.67	0.00	0.17	0.17	0.00	85.67	0.50	0.50
/œ/	NL	0.00	0.00	10.67	0.17	0.50	0.83	85.17	0.00
	EL	0.00	0.17	18.00	0.33	0.00	0.67	84.17	0.00
	PA	0.50	5.50	26.83	8.33	0.67	4.50	68.33	1.83
	SE	0.33	0.17	7.17	0.83	0.67	14.00	55.67	1.67
	TE	0.50	0.67	12.67	3.67	1.00	5.17	68.17	0.17
/u/	NL	0.17	1.00	0.17	0.00	0.00	0.33	0.50	99.00
	EL	0.33	1.17	0.33	0.17	0.00	0.33	0.33	97.00
	PA	0.50	3.83	0.33	0.00	0.00	1.83	0.00	80.67
	SE	1.33	1.50	0.00	0.17	0.00	1.17	11.83	76.33
	TE	1.00	2.00	0.17	0.00	0.00	2.50	17.67	77.50

Perceptual Analysis

The results obtained from the perceptual analysis are presented in a confusion matrix shown in table 2. The confusion matrix is constructed based on the identifica-

tion scores of vowels produced by each speaker group. It should be noted that, even for vowels produced by NL speakers, the percent correct identification was not perfect (93.5%). /ɛ/ and /œ/ produced by NL speakers were

more frequently confused with each other, and /y/ was easily confused with /i/. For the alaryngeal speaker groups, the highest accuracy was judged to be associated with EL (89.6%), followed by PA (77.1%), TE (73.6%) and SE (70.1%) speakers.

Discussion

The use of alternative phonation methods by laryngectomized patients unavoidably alters their sound quality. If the sound source and supralaryngeal filter are independent of each other, removal of the laryngeal source should not lead to any substantial change in the filtering properties of the supralaryngeal cavity, and the vowel formants should remain unchanged. However, contradictory findings were observed. Both F1 and F2 values were found to be significantly higher in alaryngeal speech than in laryngeal speech. Such findings are consistent with those reported by van As et al. [22], Kytta [23], Sisty and Weinberg [24], and Liu and Ng [26].

A formant pattern is determined by the configuration of the supralaryngeal filter. Formant frequencies change with the vocal tract configuration. The increased F1 and F2 values can be attributed to the reduced effective vocal tract length after laryngectomy. During laryngectomy, the larynx and the attached tissues between the tongue base and the trachea are removed. The surgical site is closed by suturing the anterior fibers of the inferior pharyngeal constrictor and the anterior fibers of the cricopharyngeus, which are attached to the larynx and the cricoid cartilage, respectively [1]. Assuming that the oral cavity remains unchanged after surgery, the location of the sound source and the way the sound source is coupled to the supralaryngeal vocal tract filter become the determining factors in the understanding of changes to formant values.

SE and TE Speech

During SE and TE phonation, the PE segment serves as the neoglottis. The elevated F1 and F2 values in SE and TE speech imply that the level of the PE segment in SE and TE speakers was higher than that of the vocal folds in NL speakers, resulting in a shortened vocal tract length for resonance during speech production. This is confirmed by previous imaging studies investigating the location and configuration of the PE segment in SE and TE speakers. Based on videofluoroscopic images, the PE segment in SE and TE speakers were found to be situated at roughly the level of the fourth to fifth cervical vertebrae

(C4 to C5), and may sometimes shift upward to C3/C4 [27–30]. The upward PE segment movement was attributed to the energy of pulmonary air during exhalation. Bentzen et al. [28] also observed that a high-seated PE segment during SE phonation was associated with better SE phonation. Damsté and Lerman [29] and Diedrich and Youngstrom [30] reported that the PE segment might vibrate at a level as high as the hypopharynx, at the level of C2–C3 upon phonation in some SE speakers.

According to Bentzen et al. [28], better SE voice quality was associated with a high-seated PE segment. Since superior SE speakers were recruited in the present study, it is likely that the PE segment of our SE and TE speakers was high-seated, perhaps at the level of C4–C5 when at rest, and C3–C5 during SE and TE phonation. Comparing the level of vocal folds in laryngeal speakers at C5 [31], the effective length of the vocal tract for resonance in our SE and TE speakers was believed to be shorter than that in NL speakers. The shortened vocal tract length may explain the significantly higher F1 and F2 values observed in SE and TE vowels.

PA Speech

During PA phonation, pulmonary air sets the rubber reed of the PA into vibration. Directly determining the voice quality, the pattern of rubber reed vibration was controlled by the tension and configuration of the reed. For instance, the tenser the reed, the faster was the vibration, and the higher was the perceived pitch. Sound generated by the vibrating reed was transmitted to the oral cavity via the mouth tube. However, instead of an outward propagation of the sound wave as in NL, SE, and TE speakers, the PA sound wave propagated from the oral cavity inward. Such a posterior direction of sound propagation implied a different pattern of resonance, and may have generated reflections within the vocal tract. Although detailed interaction between sound waves within the vocal tract is not yet known, it is believed that the overlapping and interference of sound waves inside the vocal tract results in a complex resonance pattern. In addition, due to the distance needed for the reflected wave to travel back and forth, a considerable time lag existed between the original and reflected waves. This corresponded to the reverberant quality of the PA sound [19]. Because of the complex resonance pattern, PA speech was not comparable with NL, SE, and TE speech where sounds were generated at one end of the resonant tube and propagate outward.

As the new sound source was not located at the end of the resonant tube as in SE, TE and NL speakers, the oral

cavity anterior and posterior to the mouth tube of the pneumatic device could be considered as the back and front resonant cavities, respectively. Both resonating cavities contribute comparably to the oral resonance. As a result, the length of the resonant tube is considerably shorter than the entire vocal tract length. This may explain why formant values in PA speech were much higher, even higher than in SE and TE speech.

EL Speech

During EL speech, sound was generated by the mechanical vibration of the diaphragm of the electrolarynx. Sound was transmitted into the oral cavity mainly from the neck and floor of the mouth (sound energy coming from the electrolarynx was transmitted to the neighboring structures in a diffused manner). The sound wave from the floor of the mouth propagated superiorly towards the palate. Some energy was reflected and propagated inferiorly towards the floor of the mouth and further towards the PE segment in the inferior part of the pharynx when it reached the palate. The reflected sound wave then interfered with the original wave and resulted in a complex resonance. The direction of sound wave propagation in EL speech was different from that in NL, SE, TE and PA speech, which might explain why EL was incomparable with the other three types of alaryngeal speech. Combined with the lack of acoustic energy in the low frequency region associated with EL speech [e.g., 20], the resonance of EL speech was considerably more complicated.

Applying Fant's [21] tube model to the understanding of vocal tract resonance in EL speech, as the new sound source was no longer located at the end of the vocal tract resonance tube as in SE, TE and NL speech, the effective vocal tract length for resonance was different. By placing the electrolarynx on the anterolateral neck, and sound mainly coming from the superior neck and the floor of the mouth, the effective resonant tube became smaller and shorter when compared with NL, SE, and TE speakers. This might explain why the EL formants were even higher than those of SE and TE speakers which were associated with a shorter effective vocal tract.

Listeners' Identification of Vowels

Percent correct identification of the vowels produced by laryngeal and alaryngeal speakers revealed interesting findings. It is noticed that the vowels produced by NL speakers were not perfectly correctly identified by listeners. The lack of perfect identification even for NL speech implies that, in order to correctly identify Cantonese

vowels, other perceptual cues may be needed besides formant frequencies. / ϵ / and / $\text{\textcircled{e}}$ / appeared to be more frequently confused with /i/ and /y/. The confusion between / ϵ , $\text{\textcircled{e}}$ / and /i, y/ may be due to the acoustical similarity between the vowel pairs: they both have similar F1 values. Vowels with similar formant values were located closer to each other in the vowel space, and discrimination between them therefore became more difficult and less successful, particularly when contextual cues were not available. Specifically, listeners tended to perceive / $\text{\textcircled{e}}$ / as / ϵ / more likely than / ϵ / as / $\text{\textcircled{e}}$ /, and /y/ as /i/ more likely than /i/ as /y/. This pattern of confusion appears to be related to the different frequencies of occurrence of the vowels in Cantonese. According to Stokes and Wong [32], the vowels /i/ and / ϵ / have higher frequency of occurrence than the vowels /y/ and / $\text{\textcircled{e}}$. It is believed that listeners tended to identify the vowels as the ones with higher frequency of occurrence (as they are heard more frequently), especially when the listeners were not sure about which vowels they had perceived during the listening experiment.

Although the overall identification scores were lower for alaryngeal vowels, the patterns of confusion exhibited by the four types of alaryngeal speech were found to be similar to that of NL speech. The two vowel pairs (/ ϵ / and / $\text{\textcircled{e}}$ /, and /i/ and /y/) produced by EL, PA, SE, and TE speakers appeared to be more confusing than other vowels. There was also a higher tendency for listeners to perceive / $\text{\textcircled{e}}$ / as / ϵ / than / ϵ / as / $\text{\textcircled{e}}$ / in all four types of alaryngeal speech, and to perceive /y/ as /i/ than /i/ as /y/ in SE, TE and EL speech, with the exception of the confusion pattern in the vowel pair /i/ and /y/ in PA speech. In PA speech, there was a higher tendency for listeners to perceive /i/ as /y/ than /y/ as /i/. This can be explained by the limitation of the PA. The placement of the mouth tube in the oral cavity via the corner of the mouth during PA phonation inevitably rounded speakers' lips even when producing the lip-spread vowel /i/. The speakers were not able to form appropriate lip-spreading in the production of /i/, and ended up producing a sound similar to the lip-rounded vowel /y/ instead. This was analogous to a person producing /i/ with a cigar in the mouth.

Despite the great deviation in formants of EL vowels, listeners achieved the highest percent correct identification of the vowels produced by EL speakers when compared with other types of alaryngeal speech. Yet, the vowel spaces of NL and EL speech were found to be quite dissimilar. The inconsistency between formants and identification accuracy of EL vowels perhaps suggests that, in addition to F1 and F2 values, other cues exist that contribute to the accurate perception of alaryngeal vowels.

Future studies should examine other possible perceptual cues such as duration and intensity in order to better understand the perception of vowels produced by EL speakers.

Conclusion

Spectral analysis revealed discrepant F1 and F2 values associated with Cantonese vowels between alaryngeal and laryngeal speech. Generally speaking, alaryngeal speech exhibited higher F1 and F2 values of Cantonese vowels than NL speakers. SE and TE speakers consistently showed higher F1 and F2 values in vowel production than NL speakers, suggesting a shorter effective length of

the resonating vocal tract in SE and TE speakers when using the PE segment as the neoglottis. Due to the use of an external sound source in EL and PA speech, the location of the new sound source and the coupling between the source and the filter determine the new formant frequencies.

The vowel spaces of NL, SE, and TE speech were more similar, whereas considerably greater deviation was seen in the PA and EL vowel spaces. Despite the great deviation in the vowel space between EL and NL speech, listeners achieved the highest percent correct identification of vowels produced by EL speech. This suggests that additional perceptual cues are contributing to the correct identification of vowels especially in the case of EL speech.

References

- Diedrich WM: Anatomy and physiology of esophageal speech; in Salmon SJ (ed): Alaryngeal Speech Rehabilitation, ed 2. Austin, Pro-Ed, 1999.
- Shirley JS: Artificial larynxes: types and modifications; in Keith RL, Darley FL (eds): Laryngectomy Rehabilitation, ed 3. Austin, Pro-Ed, 1994.
- Bellandese MH, Lerman JW, Gilbert HR: An acoustic analysis of excellent female esophageal, tracheoesophageal, and laryngeal speakers. *J Speech Lang Hear Res* 2001;44:1315–1320.
- Bennett S, Weinberg B: Acceptability ratings of normal, esophageal, and artificial laryngeal speech. *J Speech Hear Res* 1973;16:608–615.
- Weinberg B, Horii Y, Smith BE: Long time spectral and intensity characteristics of esophageal speech. *J Acoust Soc Am* 1980;67:1781–1784.
- Robbins J, Fisher HB, Blom EC, Mark IS: A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production. *J Speech Hear Disord* 1984;49:202–210.
- Schutte HK, Nieboer GJ: Aerodynamics of esophageal voice production with and without a Groningen voice prosthesis. *Folia Phoniatr Logop* 1992;54:8–18.
- Weiss MS, Basili AC: Electrolaryngeal speech produced by laryngectomized subjects – perceptual characteristics. *J Speech Hear Res* 1985;28:294–300.
- Isshiki N, Tanabe M: Acoustic and aerodynamic study of a superior electrolarynx speaker. *Folia Phoniatr Logop* 1972;24:65–76.
- Motta S, Galli I, Di Rienzo L: Aerodynamic findings in esophageal voice. *Arch Otolaryngol Head Neck Surg* 2001;127:700–704.
- Snidecor JC, Isshiki N: Air volume and air flow relationships of six esophageal speakers. *J Speech Hear Disord* 1965;30:205–216.
- Gandour J, Weinberg B, Petty SH, Dardarananda R: Rhythm in Thai esophageal speech. *J Speech Hear Res* 1986;29:563–568.
- Gandour J, Weinberg B, Petty SH, Dardarananda R: Tone in Thai alaryngeal speech. *J Speech Hear Res* 1988;53:23–29.
- Liu H, Ng M, Wan M, Wang S, Zhang Y: The effect of tonal changes on voice onset time in Mandarin esophageal speech. *J Voice* 2008;22:210–218.
- Liu H, Ng M, Wan M, Wang S, Zhang Y: Effects of place of articulation aspiration on voice onset time in Mandarin esophageal speech. *Folia Phoniatr Logop* 2007;59:147–154.
- Liu H, Wan M, Wang S, Wang X: Acoustic characteristics of Mandarin esophageal speech. *J Acoust Soc Am* 2005;118:1016–1025.
- Ching TYC, Williams R, van Hasselt A: Communication of lexical tones in Cantonese alaryngeal speech. *J Speech Hear Res* 1994;37:557–571.
- Ng M, Gilbert HR, Lerman JW: Fundamental frequency, intensity, and vowel duration characteristics related to perception of Cantonese alaryngeal speech. *Folia Phoniatr Logop* 2001;53:36–47.
- Ng M, Kwok CLI, Chow SFW: Speech performance of adult Cantonese-speaking laryngectomees using different types of alaryngeal phonation. *J Voice* 1997;11:338–344.
- Liu H, Ng M: Electrolarynx in voice rehabilitation. *Auris Nasus Larynx* 2007;34:327–332.
- Fant G: Acoustic Theory of Speech Production. The Hague, Mouton, 1970.
- van As CJ, van Ravesteijn AMA, Koopmans-van Beinum FJ, Hilgers FJM, Pols LCW: Formant frequencies of Dutch vowels in tracheoesophageal speech. *Proc Inst Phonet Sci* 1997;21:143–153.
- Kytta J: Finnish oesophageal speech after laryngectomy: sound spectrographic and cineradiographic studies. *Acta Otolaryngol* 1964;195:1–94.
- Sisty NL, Weinberg B: Formant frequency characteristics of esophageal speech. *J Speech Hear Res* 1972;15:423–438.
- Cervera T, Miralles JL, Gonzles-lvarez J: Acoustical analysis of Spanish vowels produced by laryngectomized subjects. *J Speech Lang Hear Res* 2001;44:988–996.
- Liu H, Ng M: Formant characteristics of vowels produced by Mandarin esophageal speakers. *J Voice* 2007, E-pub ahead of print.
- van As CJ, Op de Coul BMR, van den Hoogen FJA, Koopmans-van Beinum FJ, Hilgers FJM: Quantitative videofluoroscopy: a new evaluation tool for tracheoesophageal voice production. *Arch Otolaryngol Head Neck Surg* 2001;127:161–169.
- Bentzen N, Guld A, Rasmussen H: X-ray video-tape studies of laryngectomized patients. *J Laryngol Otol* 1976;90:655–666.
- Damst PH, Lerman JW: Configuration of neoglottis: an X-ray study. *Folia Phoniatr (Basel)* 1969;21:347–358.
- Diedrich WM, Youngstrom KA: Alaryngeal Speech. Springfield, Thomas, 1966.
- Mller TB, Reif E: Pocket Atlas of Cross-Sectional Anatomy: Computer Tomography and Magnetic Resonance Imaging, vol 1: Head, Neck, Spine and Joints. Stuttgart, Thieme, 1994.
- Stokes SF, Wong IM: Vowel and diphthong development in Cantonese-speaking children. *Clin Linguist Phon* 2002;16:597–617.