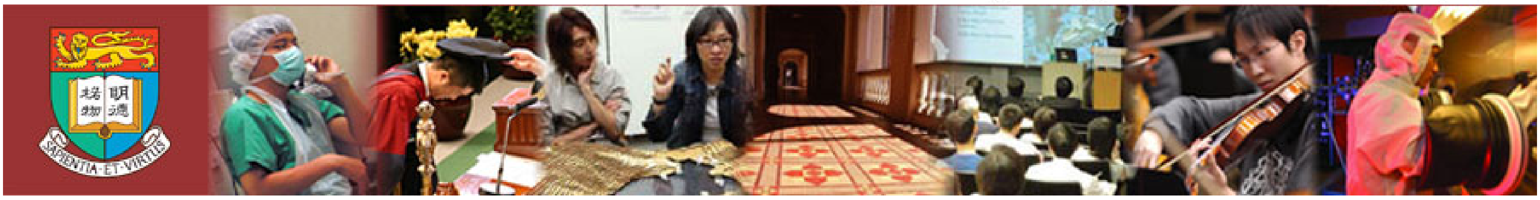

Title	Examining the neoglottal vibratory pattern of Cantonese tracheoesophageal speakers : a preliminary aerodynamic study using inverse-filtering
Author(s)	Chan, Ming-wa; 陳銘華
Citation	Chan, M. [陳銘華]. (2012). Examining the neoglottal vibratory pattern of Cantonese tracheoesophageal speakers : a preliminary aerodynamic study using inverse-filtering. (Thesis). University of Hong Kong, Pokfulam, Hong Kong SAR.
Issued Date	2012
URL	http://hdl.handle.net/10722/237891
Rights	The author retains all proprietary rights, (such as patent rights) and the right to use in future works.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Examining the neoglottal vibratory pattern of Cantonese tracheoesophageal speakers - A preliminary aerodynamic study using inverse-filtering

Chan Ming Wa

A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2012.

ABSTRACT

The present study examined the neoglottal vibratory pattern of Cantonese tracheoesophageal (TE) speakers by inverse-filtering the airflow signals obtained from eight superior TE speakers during phonation. The syllable /papapa/ was used for obtaining airflow signals, and the acoustic signals of the vowels /i, æ, a, ɔ, u/ were also obtained. Aerodynamic parameters obtained were compared between TE and laryngeal speakers. Results revealed that TE speakers exhibited comparable open quotient and airflow volume values but significantly smaller speed quotient values than laryngeal speakers. The marked difference in inverse-filtered airflow signals between TE and laryngeal speech of Cantonese is believed to be related to the use of different sounding mechanisms between the two speaking methods, and the unique vibratory nature of the neoglottis in TE speech.

INTRODUCTION

With the removal of the entire larynx during total laryngectomy, laryngeal cancer patients lose the ability to speak post-operatively. Therefore, seeking for an alternative speaking method after the surgery becomes a critical part of their rehabilitation. To date, a number of alaryngeal speaking methods are available, including pneumatic artificial laryngeal, electrolaryngeal, esophageal, and tracheoesophageal speech (Ng, Kwok, & Chow, 1997). Among them, tracheoesophageal (TE) speech can be considered as a panacea for the laryngeal cancer patients after total laryngectomy. Singer and Blom (1980) advocated that the intersection between the pharynx and the esophagus, known as the pharyngoesophageal (PE) segment, can be used as a substitute sound source for speech production. With the new sounding device, such voicing apparatus is called the neoglottis. Similar to laryngeal speech, TE speech is pulmonary driven such that the lungs still serve as a source for airflow and air pressure. It makes use of pulmonary air to set the sound source (PE segment) into periodic vibration. Yet, in TE phonation, vibration of the PE segment functions as a signal excitator and the source of voiced speech. The sound is subsequently modulated by the vocal tract during which neoglottal signal is acoustically filtered by the articulators to form different vowel and consonants (Moon & Weinberg, 1987).

Many researchers have attempted to make use of different acoustical descriptors to quantify TE voice quality and aimed to provide insights into the acoustical components of TE voice quality (Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995; Ng & Chu, 2009; Ng & Wong, 2009; Robbins, Fisher, Blom, & Singer, 1984). Acoustical parameters such as fundamental frequency (F0) and formant frequencies, vocal intensity, perturbation and noise, as well as temporal characteristics were used to acoustically describe TE voice characteristics. Studies showed consistent results; TE speech tended to exhibit lower F0 values, reduced vocal intensity, increased perturbation in F0 (jitter) and amplitude (shimmer), and elevated noise level. These comparative studies have also linked acoustical similarities and differences between laryngeal and alaryngeal speech with perceptual judgment of the respective voice quality. Results generally showed that TE speech is superior to other alaryngeal speech with respect to these acoustical measurements, which appeared comparable with those of laryngeal speech (Ng, 2011; Ng & Wong, 2009; Robbins et al., 1984). Meanwhile, TE voice quality was percieved as more acceptable among other alaryngeal voices (van As, Hilgers, Verdonck-de Leeuw, & Beinum, 1998). Many studies provided objective and quantifiable measurements to support the subjective perception of TE voice quality.

Researchers also focused on examining the vibratory characteristics of PE segment in TE speakers, vibration of which should closely correlate with TE sound quality. Previous studies aimed at examining the aerodynamic characteristics of TE speech in terms of airflow (Moon & Weinberg, 1987; Ng, 2011; Robbins et al., 1984; Searl, 2008; Searl, 2002; Searl & Evitts, 2004), and intra-oral pressure during TE phonation (Searl, 2007; Searl, 2002; Searl & Evitts, 2004). Results revealed that the intra-oral pressure associated with TE speech was greater than laryngeal speech. As suggested by some researchers, the elevated oral pressure among TE speakers could be result of several factors such as differences in phonatory mechanism and high source-driving pressure during phonation (Searl & Evitts, 2004).

Although previous comparative studies generally attributed the unique voice quality of TE speech to the use of neoglottis as a new voice source, the supra-neoglottal vocal tract still contributes to the speech sounds obtained at the mouth opening, as they, in theory, should have been modulated by vocal tract filter. Yet, there is a lack of information into the specifics of the vibratory principal of the neoglottis in TE speech. It is not known how the vocal tract of TE speakers affects the quality of TE voice. Meanwhile, as PE segment is not easily accessible, any real-time measurement of the neoglottis such as endolaryngoscopy can be considered invasive. The detailed vibratory pattern of PE segment during TE phonation remains unclear and is poorly understood. Without a detailed understanding of the neoglottal vibratory behavior during TE phonation, investigation of mechanism of TE speech is still considered incomprehensive and incomplete. This knowledge gap has hindered the complete understanding of how TE phonation works, and subsequently improvement in speech therapy regime for TE speakers for enhancing TE voice quality. Hence, the need to use an alternative means to examine the vibratory behavior of neoglottis is necessary.

Inverse-filtering

According to the source-filter theory, all speech signals obtained at the mouth opening have been modulated by vocal tract filter (Pickett, 1999). In order to examine the glottal or neoglottal signal from speech sounds, the effect of vocal tract resonance needs to be eliminated by inverse-filtering the airflow signals obtained from the mouth opening. The inverse filtering is based on the source-filter theory which assumes that the speech output (at the mouth opening) results from a sound source (glottal or neoglottal) which is modified by the vocal tract transfer function, forming an energy spectrum with a broadband energy peaks. With the removal of the effect of the acoustic filtering effect of vocal tract resonance, the original glottal or neoglottal waveform can be revealed. The resulting signal yields an estimation of the excitation signal (Löfqvist, 1991).

Aerodynamic characteristic of laryngeal speech

Studies of the glottal characteristics in laryngeal speech have been widely reported (Eskenazi, Childers, & Hicks, 1990; Hertegård, Gauffin, & Karlsson, 1992; Holmberg et al., 1995; Lehto, Airas, Björkner, Sundberg, & Alku, 2007; Ng, Gilbert, & Lerman, 1997; Rothenberg, 1973; Timcke, von Leden, & Moore, 1958; Timcke, von Leden, & Moore, 1959). Results related glottal activities to aerodynamic and acoustic quantities and correlated the inversed glottal airflow with different voice qualities. Timcke and colleagues (1958) investigated glottal flow by expressing the relative duration of the different phases of a vibratory cycle in terms of open quotient and speed quotient. It was indicated that these two measurements are the convenient means for the basic description of vibratory cycle and quantification of the function of the voice source. Open quotient is the ratio between duration when the glottis is open and that of entire vibratory cycle. It is helpful in determining the valving efficiency of the vibrating vocal folds. Speed quotient is defined as the ratio between durations of the opening phase and the closing phase. Research revealed that the degree of tension of vocal folds and the airflow pressure contribute to the variation of speed quotient. It also provided an estimate of symmetry of glottal cycles (Timcke et al., 1958). The area under curve (AUC) is defined as the average area under the flow function for the interval of interest. Both open quotient and speed quotient reflect prominent temporal characteristics of the airflow function while AUC the airflow function provides an indication of average air expenditure. It is believed that the technique of inverse filtering could also be applied to estimate the vibratory pattern of the neoglottis in TE speakers.

The findings would help derive the neoglottal waveform with the use of airflow signal of TE speech. By providing a description of key aerodynamic characteristics of TE voice and predicting the proficiency of TE speech, the paucity of information regarding the vibratory principle of TE speech can be alleviated. Besides, the norms developed will help refining the treatment goals and strategies. Based on the previous discussion, it is hypothesized that:

1. The neoglottic vibration of Cantonese TE speech was variable in temporal and shape

characteristics (Qi & Weinberg, 1995)

2. The volume of the airflow in neoglottic vibration was comparable to that in glottis vibration (Moon & Weinberg, 1987).

METHOD

Subjects

Eight male TE speakers with ages ranging from 52 to 88 years (mean = 69.5 years) and eight male laryngeal (NL) speakers with ages ranging from 50 to 83 years (mean = 66.0 years) participated in the present study. All speakers were adult native speakers of Cantonese. Prior to the experiment, all speakers completed a case history questionnaire and confirmed that they had no reported history of any respiratory, speech, language and/or hearing problems, except that associated with laryngectomy for the alaryngeal speakers. All TE speakers were judged to be superior TE speakers by practicing speech therapists who were experienced in TE speech rehabilitation. Only superior TE speakers were selected as the study only focused on the aerodynamic characteristics associated with vibration of PE segment in proficient TE speakers of Cantonese. All alaryngeal speakers were selected from the New Voice Club of Hong Kong, which was a self-help organization for laryngectomees in Hong Kong.

Speech Materials

The speech materials included five monophthong vowels of Cantonese /i/, /æ/, /a/, /ɔ/ and /u/. During the experiment, the subjects were instructed to sustain the five vowels for approximately five seconds at a comfortable level of intensity and pitch. The formant frequencies were later obtained from these sustained vowels and used for further analysis.

The speakers were instructed to produce the syllable /ap^{h}ap^{h}a/ for three times. The vowel /a/ was used to extract volume velocity waveform at the neoglottis and the effect of the vocal tract filter under inverse filtering. The vowel /a/ was chosen because its formant frequencies fell within the range of the frequencies provided by the aerodynamic measurement device. This method of estimating the pressure below the vibratory sound source has been discussed previously by Smitheran and Hixon (1981) and validated by Löfqvist (1991).

Instrumentation and Measurements

Acoustic signals of the sustained vowels were recorded using a high-quality dynamic microphone (Shure, SM58), which was placed approximately six inches in front of the speaker's mouth during the experiment. All signals were recorded and stored in a computer for later analyses. The first and second formant frequency values of the vowel /a/ were used to acquire the inverse-filtered sub-neoglottal waveform. The airflow measurement was acquired by using an aerodynamic measurement system (MS110, Glottal Enterprise, NY).

Procedure

Before the recording, the aerodynamic system was carefully calibrated using the Glottal Enterprises airflow calibration units (PCU-1 & FCU-1, Glottal Enterprises, NY). Calibration was carried out by strictly following the procedure listed in the user's manual. All calibration signals were recorded for later signal analyses.

The actual recording took place in a sound-isolated room in order to minimize recording of extraneous noise. The participants were first explained the purpose of the study and were given sufficient practice time to ensure they were familiar with the recording environment and the speech materials. All the speech samples were produced at a comfortable level of pitch and loudness. To obtain the formant frequencies, they were asked to produce five sustained vowels /i/,

/æ/, /a/, /ɔ/ and /u/ for approximately five seconds. The acoustic signals were recorded by using a dynamic microphone (SM58, Shure) and stored in a computer. To obtain the airflow signals, an undivided anesthesia face mask was held against the speakers' face, covering the mouth and the nose, to avoid any possible air leakage blockage during data collection. The airflow signals were recorded and digitized, and stored in a computer for further analyses.

Data Analysis

Analysis of airflow signals was carried out by using WaveViewPro program which is the accompanying software for the Aerodynamic Measurement System (MS 110, Glottal Enterprises, NY). Based on the airflow waveform, average airflow rate value was measured.

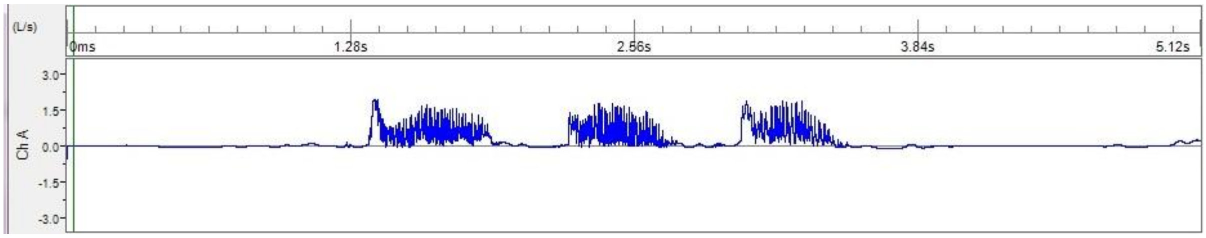

Figure 1: Illustration of airflow rate measurement from an airflow signal by using the WaveViewPro program.

To facilitate inverse filtering of the airflow signals, the first two formant frequencies (F1 & F2) were obtained acoustically from the sustained vowel /a/ using the software Praat. The calculated F1 and F2 values were used as inputs to the inverse filtering unit in order to obtain the best neoglottal waveform. With F1 and F2 as inputs, airflow signals were inverse-filtered to derive the neoglottal waveform. The best inverse-filtered neoglottal waveform was obtained after adjustment of the damping controls on the inverse filter device was completed. The optimal result was interpreted as a neoglottal waveform, which shows a maximally flat horizontal closed phase with minimal remaining formant ripple. For each speaker, 50 cycles of the waveform were analyzed. With the derived neoglottal waveforms, the opening phase, closing phase, and the closed phase were identified for each cycle. The duration of these phases was calculated from the waveform, using which the open quotient, speed quotient, and AUC values were calculated.

Statistical Analyses

To assess the difference in neoglottal waveform of TE speakers and glottal waveforms of laryngeal speakers, Mann Whitney U tests of the parameters derived from the inverse-filtered airflow signals between two groups (TE vs. laryngeal speakers) were used.

Reliability Measure

As human judgment was involved in the data analysis, reliability was assessed. A total of 20% of the entire signals were randomly selected for re-analysis by the investigator and by a second investigator who was also experienced in carrying out acoustic analysis. Pearson Product-Moment Correlation coefficient was used to depict the inter- and intra-examiner reliability of airflow measurements. Correlation coefficient of airflow for both inter- and intra-examiner reliability were 0.91 and 0.99, respectively ($p < .01$), indicating high agreement among the raters and multiple repetitions of rating respectively.

RESULTS

Demographic information of the TE speakers

Demographic information of TE speakers including their age, gender, presence of radiation therapy, myotomy, the duration of using TE phonation and type of TE puncture voice prostheses is shown in Table 1. In the present study, all TE speakers were Provox-valve users, and have been using TE speech as their major mode of phonation for periods ranging from one to six years.

Table 1. Demographic information of tracheoesophageal (TE) speakers.

	Speakers	Age (years)	Gender	Radiation therapy	Myotomy	Duration of using TE speech (years)
	TE1	52	Male	Yes	No	1
	TE2	59	Male	Yes	No	5
	TE3	63	Male	Yes	Not known	5
	TE4	70	Male	No	No	5
	TE5	72	Male	Yes	Not known	5
	TE6	83	Male	Yes	No	6
	TE7	69	Male	Yes	Not known	6
	TE8	88	Male	Yes	Yes	6
Mean		69.5				4.88
Range		52-88				1-6

Shape and temporal characteristics of airflow signals

Representative inverse-filtered airflow signal of a TE speaker and a NL speaker are provided in Figures 2.

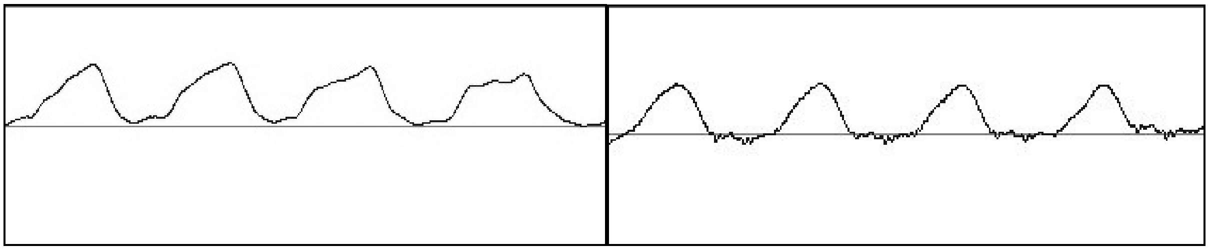

Figure 2. Typical inverse-filtered airflow signals associated with a TE speaker (left) and a NL speaker (right).

Inverse-filtered airflow signals

Open quotient (OQ)

Average open quotient (OQ) values associated with TE and laryngeal phonation are depicted in Table 2. Mean OQ measures for TE speakers and laryngeal speakers ranged from 0.23 to7.25 and 0.62 to 0.83, respectively. TE speakers generally had comparable OQ with laryngeal speakers. A Mann Whitney U test was performed to determine if OQ values were significantly different between TE and laryngeal speakers. Results revealed that there was no significant difference between two groups ($p \geq .05$).

Speed quotient (SQ)

Average speed quotient (SQ) values obtained from TE and laryngeal speakers are shown in Table 3. Mean SQ measures for TE and laryngeal speakers ranged from 0.51 to 2.85 and 1.25 to 2.21, respectively. In general, TE speakers were associated with smaller speed quotient with laryngeal speakers. Results of the Mann Whitney U test showed significant difference between the TE and laryngeal speaker groups ($p \leq .05$). The speed quotient values were significant smaller for the TE speaker than for the laryngeal speakers.

Table 2. Mean, standard deviation and range values of open quotient (OQ) associated with tracheoesophageal and laryngeal speech.

TE speech	Mean	Range	Laryngeal speech	Mean	Range
TE1	0.89 (0.04)	0.78-0.96	NL1	0.64 (0.10)	0.53-0.83
TE2	0.80 (0.07)	0.67-0.92	NL2	0.77 (0.04)	0.69-0.82
TE3	0.80 (0.07)	0.61-0.91	NL3	0.75 (0.04)	0.64-0.79
TE4	0.71 (0.11)	0.48-0.93	NL4	0.71 (0.06)	0.65-0.80
TE5	0.75 (0.09)	0.52-0.90	NL5	0.71 (0.04)	0.62-0.77
TE6	0.63 (0.05)	0.52-0.77	NL6	0.72 (0.05)	0.65-0.79
TE7	0.85 (0.04)	0.75-0.93	NL7	0.76 (0.02)	0.69-0.78
TE8	0.71 (0.06)	0.51-0.92	NL8	0.71 (0.04)	0.67-0.76
Mean	0.77 (0.08)			0.72 (0.04)	

Table 3. Mean, standard deviation, and range values of speed quotient (SQ) associated with tracheoesophageal and laryngeal speech.

TE speech	Mean	Range	Laryngeal speech	Mean	Range
TE1	0.92 (0.31)	0.53-1.90	NL1	2.02 (0.31)	1.51-2.35
TE2	1.40 (0.30)	0.98-2.19	NL2	1.25 (0.34)	1.34-2.43
TE3	1.33 (0.62)	0.60-4.76	NL3	1.76 (0.36)	1.17-2.39
TE4	1.38 (0.32)	0.85-2.34	NL4	1.74 (0.49)	1.18-2.43
TE5	0.84 (0.16)	0.51-1.43	NL5	1.82 (0.46)	1.13-2.37
TE6	1.34 (0.37)	0.81-2.85	NL6	2.21 (0.29)	1.67-2.67
TE7	1.18 (0.18)	0.85-0.93	NL7	1.78 (1.24)	1.56-1.93
TE8	1.20 (0.23)	0.84-1.67	NL8	1.97 (0.34)	1.45-2.43
Mean	1.20 (0.20)			1.81 (0.26)	

Average area under curve (AUC)

Average area under curve (AUC) values associated with TE and laryngeal speech are depicted in Table 4. Mean AUC measures for TE and laryngeal speech ranged from 0.35 to 7.28 and 1.19 to 6.74, respectively. TE speech generally had comparable AUC with laryngeal speech. Results of the Mann Whitney U test revealed that there was no significant difference between two groups ($p \geq .05$).

Table 4. Mean, standard deviation, and range values of area under curve (AUC) (in mL) associated with tracheoesophageal and laryngeal speech.

TE speech	Mean	Range	Laryngeal speech	Mean	Range
TE1	7.28 (3.45)	1.19-12.53	NL1	6.74 (0.95)	4.39-8.9
TE2	3.78 (0.65)	2.21-5.27	NL2	1.57 (0.41)	0.87-2.46
TE3	2.47 (1.04)	1.19-4.39	NL3	1.19 (0.35)	0.38-1.84
TE4	4.38 (1.39)	2.25-7.25	NL4	1.50 (0.53)	1.27-1.89
TE5	0.51 (0.24)	0.15-1.00	NL5	2.89 (0.23)	2.56-3.33
TE6	3.16 (0.73)	2.13-4.82	NL6	2.26 (0.14)	2.02-2.54
TE7	0.33 (0.08)	0.13-0.48	NL7	1.89 (0.31)	1.52-2.14
TE8	0.35 (0.09)	1.12-0.51	NL8	2.11 (0.43)	1.25-3.01
Mean	2.84 (2.31)			2.52 (1.67)	

Opening, closing, and closed phases

Mean duration of the opening, closing, and closed phases relative to the entire vibratory cycle is depicted in Figure 3. It can be seen that the durational relationship among the three phases in neoglottic vibration is generally similar to that in glottic vibration. However, the opening phase in neoglottic vibration in TE speakers is shorter than that in glottic vibration, whereas the closing phase in neoglottic vibration is longer than that in glottic vibration. The mean opening phase in neoglottic vibration is 41% of the entire vibratory cycle whereas that in glottic vibration is 46%. The mean closing phase in neoglottic vibration is 37% of the entire vibratory cycle while that in glottis vibration is 31%. This implies that, for neoglottal vibration, the time interval during which the PE segment was open was shorter than vocal folds in glottal vibration.

Reliability measures

For the aerodynamic measurement, inter- and intra-examiner test-retest reliability was determined. For the intra-examiner reliability, percentage of agreement was calculated based on the data between the first and the second measurements made by the primary investigator. Pearson Product-Moment Correlation coefficient was 0.92 ($p < .05$).

For the inter-examiner reliability, percentage of agreement of measurements made by the first and second investigators was measured. Pearson Product-Moment Correlation coefficient was 0.87 ($p < .05$). The high inter- and intra-examiner test-retest values indicated high agreement of reliability among the raters and among the repeated repetition respectively, indicating that the measured aerodynamic values were reliable and consistent.

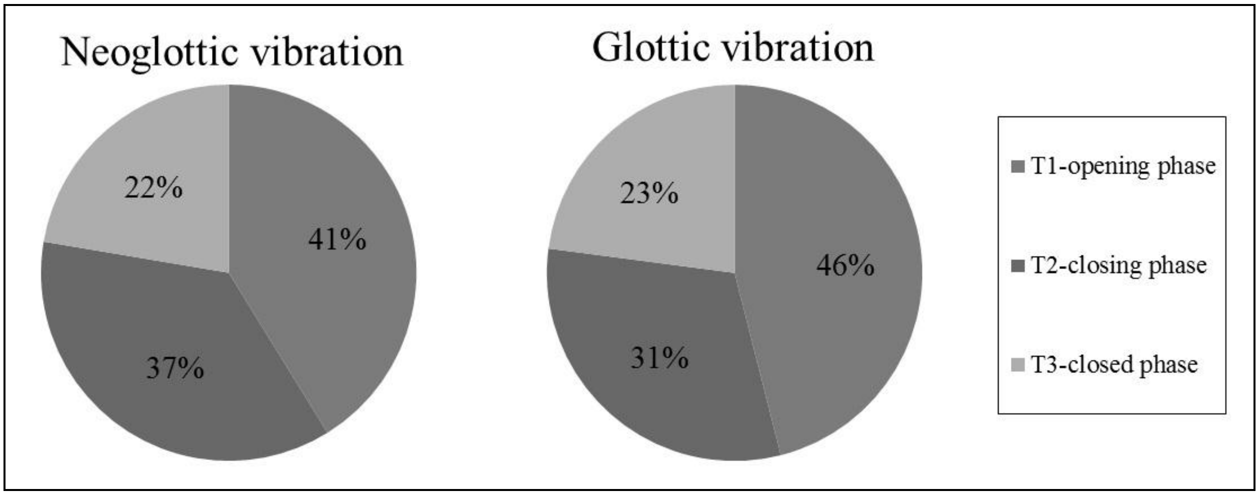

Figure 3. Relative duration of opening, closing, and closed phases in a vibratory cycle of neoglottic vibration and glottic vibration.

DISCUSSION

The average airflow rates in neoglottic vibration and glottis vibration were obtained from TE speakers and NL speakers respectively. The present results support the hypothesis that the neoglottal vibration demonstrated by Cantonese TE speakers was variable and exhibited comparable airflow volume with laryngeal speakers.

Shape and temporal characteristics of inverse-filtered waveforms

The inverse-filtered signals provided descriptive information regarding the signal of the voicing source, regardless of mode of phonation. Comparing the inverse-filtered waveforms between TE and laryngeal speakers, it is apparent that the highly-variable TE neoglottal waveform contrasts markedly with the relatively more consistent and homogeneous glottal waveforms associated with laryngeal speakers.

In an earlier study of esophageal and tracheoesophageal speakers, Qi and Weinberg (1995) highlighted the homogeneity of the neoglottal airflow signals, with appearance resembling vocal fold vibration. Examination of the inverse-filtered airflow waveform from a representative laryngeal speaker as shown in Figure 2 (right panel) showed comparable results, in terms of both its shape and temporal characteristics. According to Figure 2, both neoglottal and glottal waveforms are triangular in shape and quasi-periodic in nature. According to Kendall and Leonard (2010), the shape and movement of vocal folds are primarily determined by the fine coordination of intrinsic laryngeal muscles during phonation. The intrinsic laryngeal muscles include an array of voluntary skeletal muscles such as posterior cricoarytenoid, lateral cricoarytenoid, arytenoid, cricothyroid and cricoarytenoid muscles (Zemlin, 1997). During laryngeal phonation, the network of abductory and adductory laryngeal muscles serves as the control center for the adduction and abduction of vocal folds. As the intrinsic laryngeal muscles always function antagonistically (i.e., both abductory and adductory laryngeal muscles are active at all time), the fine-tuning and adjustments of laryngeal musculatures result in appropriate medial compression and longitudinal tension for glottal vibration that yields a voice that is pitch and loudness-appropriate (Zemlin, 1997). The subtle adjustment of laryngeal muscles simultaneously controls the ever-changing transglottal airflow rate and subglottic pressure, yielding the quasi-periodic vibration and a triangular shape waveform. The dichotomous pattern of vocal fold vibration has been considered as the characteristics of laryngeal vibration (Kendall & Leonard, 2010). Significant deviation from the asymmetry and regularity of the vocal fold vibratory pattern is regarded as an abnormal glottal vibration.

Meanwhile, the inverse-filtered airflow waveforms associated with TE speech contrast with those with laryngeal speakers. Figure 2 above compares the inverse-filtered waveforms between a typical TE speaker and a laryngeal speaker. A greater amount of irregularities can be observed in the TE waveform, in which some cycles appear to be more rectangular and others more rounded and slightly skewed. Recall that the inverse-filtered airflow waveform reflects the opening and closing of the glottis/neoglottis during phonation, resembling the time-varying area formed by the vibrating glottis/neoglottis. Physically, such awkwardness in the TE waveform can be translated into the reduced regularity and consistency of neoglottal (PE segment) vibration in TE speakers, suggesting the poor control over vibration of PE segment in TE speakers. According to Edels (1983), the PE segment is composed of fibers of the inferior pharyngeal constrictor, cricopharyngeus muscles and the most proximal cervical esophagus (muscular layer of esophagus). They consist of both skeletal and smooth muscle fibers which imply the lack of 100% voluntary control of the PE segment in TE speakers. In addition, muscles for neoglottal vibration are not in parallel in position and do not act in pairs in executing the movement. Omori, Kojima, Nonomura, and Fukushima (1994) reported that the anatomical findings for the vibrating neoglottis with the use of roentgenography. While the medial-lateral and vertical movement was seen in vocal fold vibration, the thyropharyngeus muscle in TE speakers contracted anteroposteriorly and cricopharyngeus muscle contracted circularly. The difference in the pattern of muscle contraction and lack of combined action of the muscles further explained the irregularity and inconsistency of the neoglottal waveform of TE speakers (see Figure 2).

Aerodynamic characteristics of neoglottal vibration

As shown in Tables 2, 3, and 4, average OQ and AUC values of TE speech appeared to be comparable with those of laryngeal speech. However, average SQ values of TE speakers were lower than those of laryngeal speakers. As discussed previously, OQ refers to the relative time when the glottis/neoglottis is open, and AUC corresponds to the amount of airflow escaping the glottis or neoglottis in each cycle. Comparable average OQ values found here implies that the glottis and the PE segment were open for a similar amount of time within each cycle during TE and laryngeal phonation. With a similar duration of opened glottis and PE segment within each cycle, the comparable AUC values between TE speech and laryngeal speech implied that the air was leaking at a similar rate, regardless of the mode of phonation. This finding is consistent to Moon and Weinberg (1987), which suggested that the TE speech exhibited comparable trans-source airflow rate with the laryngeal speech.

From the present findings, there was no significant difference in OQ values but significant difference was shown in SQ values between neoglottic vibration and glottis vibration. These results indicated that the opening phase (T1) of PE segment was shorter than that of glottis, or the closing phase (T2) of PE segment was longer than that of glottis, yielding a similar duration of opened phase (T1 + T2). Further investigation was done on the duration of phases between neoglottic vibration and glottis vibration (see Figure 3), and the finding confirmed that the opening phase (T1) was shorter while the closing phase (T2) of the neoglottic vibration was longer than that of glottis vibration during each vibratory cycle. This implied that the PE segment opened in a faster rate but required longer time in closing when compared with vocal folds, resulting in a smaller SQ value. According to the myoelastic-aerodynamic theory of vocal fold vibration, sufficient subglottal pressure is responsible for the opening of the vocal folds, and the major contributor to the closing of vocal folds lies in the Bernoulli force (van den Berg, 1958; Titze, 1976, 1980, 2006). Similar analog could be drawn for the neoglottal vibration (Moon & Weinberg, 1987). The quick opening of the PE segment might be explained by high trans-neoglottal pressure differential. According to Moon and Weinberg (1987), TE speech exhibited higher source driving pressure (i.e., sub-neoglottal pressure) in comparison with laryngeal speech, resulting a higher pressure below the vibratory source (neoglottis). Meanwhile, the supra-neoglottal pressure in TE speakers appeared to be comparable to the supraglottic pressure in laryngeal speakers, at least during an open vowel production (Schutte & Nieboer, 2002). It follows that the trans-neoglottal pressure differential is higher than the transglottal pressure differential. The greater sub-neoglottal pressure, according to the myoelastic-aerodynamic theory, results in a faster opening of the PE segment. This explains the shorter duration of opening phase (T1) in neoglottic vibration.

On the other hand, the longer duration of the closing phase in neoglottic vibration might be explained by differences in mechanism to facilitate the closing of glottis and neoglottis. According to the myoelastic-aerodynamic theory (van den Berg, 1958; Titze, 1976, 1980, 2006), closing of the vocal folds is achieved by the elasticity of twisted vocal fold tissue, the negative Bernoulli pressure, and the gravity. When the vocal folds are abducted, air escapes through the glottis, and the transglottal pressure differential decreases. When the negative pressure between the medial edges of the vocal folds is attained, the inferior muscle fibers in vocal fold are twisted and forced to adduct from a parallel course by the Bernoulli effect in order to equalize the pressure. However, the elasticity of muscles might be different for PE segment vibration. The muscles in PE segment are thicker and more massive than the vocal folds. More time is needed for the less elastic PE segment to close, yielding a weaker "closing force" to close the opened PE segment, as compared with the vocal folds. This might explain the longer duration for PE segment to adduct resulting a longer closing phase found in the present study. Yet, further studies are needed to confirm this suggested explanation.

Similarities between English-speaking norms and the present findings

Durational relationship among different phases for modal register has been investigated by researchers previously. Recall that modal register is the most efficient mode of phonation used in daily conversation. Holmberg, Hillman, and Perkell (1988) reported normative data of OQ and SQ for English-speaking laryngeal speakers. OQ and SQ values for male laryngeal speakers phonating at a comfortable level of pitch and loudness ranged from 0.46 to 0.77, and 1.32 to 2.58, respectively. The positive SQ values indicated that the opening phase is always longer than the closing phase; opening is always slower than closing of the vocal folds. The present data of laryngeal speakers showed comparable results. Zemlin (1997) reported the average values for the opening phase (T1), closing phase (T2) and closed phase (T3) for English speakers as 50%, 31%, and 19% of the entire vibratory cycle, respectively. The present data of laryngeal speakers also showed comparable findings. This consistency in T1, T2, and T3, and OQ and SQ values implies that, despite the difference in language being investigated, vocal folds of Cantonese laryngeal speakers were vibrating in a similar fashion as English speakers, yielding similar proportions in the duration of phases. Language does not appear to affect these glottal measurements. In a way, this also shows that the present findings for laryngeal phonation could reliably serve as references for the TE data to compare with.

Limitations of the present study and suggestions for future studies

The present study investigated the vibratory pattern of the neoglottis in TE speakers of Cantonese. One limitation of the study is related to the small data set available. This seems unavoidably due to the limited number of TE speakers in Hong Kong, combined with the fact that only superior TE speakers were chosen in the present study. During the experiment, they were asked to phonate at a comfortable level of loudness and pitch. As the vocal characteristics and the voice qualities are determined by the vibratory of the neoglottis during phonation (Omori & Kojima, 1999), findings of the present study may not generalize to the entire TE population. Further research on different voice qualities and phonation models among the population of tracheoesophageal speakers is needed to better understand the relationship between the vibratory patterns of the PE segment and various parameters.

CONCLUSION

The present study derived the neoglottal vibratory waveform by inverse-filtering the airflow signal obtained from the Cantonese tracheoesophageal speakers during phonation. Results on aerodynamic measures indicated TE speech exhibited comparable open quotient values and volume of airflow with laryngeal speech whereas showed significantly smaller speech quotient values than laryngeal speech. During the vibratory cycle, the PE segment of TE speakers appeared to open faster than vocal folds in laryngeal speakers while it tended to close more slowly than vocal folds. Based on this finding, it can be concluded that the unique vibratory nature of the neoglottis and the difference in voicing mechanism between the PE segment and vocal folds attribute to the difference in the vibratory patterns.

ACKNOWLEDGEMENTS

I would like to expressive my sincere gratitude to my supervisor, Dr Lawrence Ng, for his invaluable comments, suggestions and kind concerns. I would also like to express my thanks to the participants of this study. Furthermore, I would like to thank my classmates and friends for their love, friendship and encouragements throughout the process of the completion of this study, including Tracy Yau, Thomas Siu, Noel Bell, Kammy Lee, Sunny Wong, Moli Ng, Connie Ho and Yamada Yukiho. Last but not least, I would like to dedicate my most sincere gratitude to my parents and two sisters, Zoe and Becky, for their endless love and unconditional support.

References

van As, C. J., Hilgers, F. J. M., Verdonck-de Leeuw, I. M., & Beinum, F. J. (1998). Acoustical analysis and perceptual evaluation of tracheoesophageal prosthetic voice. *Journal of voice, 12*(2), 239-248.

van den Berg, J. W. (1958). Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research, 1*, 227-244.

Edels, Y. (1983). Pseudo-voice: its theory and practice. In Y. Edels (ed.), *Laryngectomy: Diagnosis to Rehabilitation* (pp. 107-142). London: Croom-Helm.

Eskenazi, L., Childers, D. G., & Hicks, D. M. (1990). Acoustic correlates of vocal quality. *Journal of Speech and Hearing Research, 33*(2), 298-306.

Hertegård, S., Gauffin, J., & Karlsson, I. (1992). Physiological correlates of the inverse filtered flow waveform. *Journal of voice, 6*(3), 224-234.

Holmberg, E. B., Hillman, R. E., & Perkell, J. S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *Journal of the Acoustical Society of America, 84*(2), 511-529.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P., & Goldman, S. L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research, 38*, 1212-1223.

Lehto, L., Airas, M., Björkner, E., Sundberg, J., & Alku, P. (2007). Comparison of two inverse filtering methods in parameterization of the glottal closing phase characteristics in different phonation types. *Journal of Voice, 21*(2), 138-150.

Löfqvist, A. (1991). Inverse filtering as a tool in voice research and therapy. *Logopedics Phoniatrics Vocology, 16*(1-2), 8-16.

Moon, J. B., & Weinberg, B. (1987). Aerodynamic and myoelastic contributions to tracheoesophageal voice production. *Journal of Speech and Hearing Research, 30*(3), 387-395.

Ng, M. L. (2011). Aerodynamic characteristics associated with oesophageal and tracheoesophageal speech of Cantonese. *International Journal of Speech-Language Pathology, 13*(2), 137-144.

Ng, M. L., & Chu, R. (2009). An acoustical and perceptual study of vowels produced by alaryngeal speakers of Cantonese. *Folia Phoniatrica et Logopaedica, 61*(2), 97-104.

Ng, M. L., Gilbert, H. R., & Lerman, J. W. (1997). Some aerodynamic and acoustic characteristics of acute laryngitis. *Journal of Voice, 35*(6), 356-263.

Ng, M. L., Kwok, C. L., & Chow, S. F. (1997). Speech performance of adult Cantonese-speaking laryngectomees using different types of alaryngeal phonation. *Journal of Voice, 11*(3), 338-344.

Ng, M. L., & Wong, J. (2009). Voice onset time characteristics of esophageal, tracheoesophageal, and laryngeal speech of Cantonese. *Journal of Speech, Language, and Hearing Research, 52*(3), 780-789.

Omori, K., & Kojima, H. (1999). Neoglottic vibration in tracheoesophageal shunt phonation. *European Archives of Oto-Rhino-Laryngology, 256*(10), 501-505.

Omori, K., Kojima, H., Nonomura, M., & Fukushima, H. (1994). Mechanism of tracheoesophageal shunt phonation. *Archives of Otolaryngology - Head and Neck Surgery, 120*(6), 648-652.

Pickett, J. M. (1999). *The Acoustics of Speech Communication - Fundamentals, Speech Perception Theory, and Technology*. Boston, MA: Allyn and Bacon.

Robbins, J., Fisher, H. B., Blom, E. C., & Singer, M. I. (1984). A Comparative Acoustic Study of Normal, Esophageal, and Tracheoesophageal Speech Production. *Journal of Speech and Hearing Disorders, 49*(2), 202-210.

Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal airflow waveform during voicing. *Journal of the Acoustical Society of America*, *53*, 1632-1645.

Searl, J. P. (2007). Bilabial contact pressure and oral air pressure during tracheoesophageal speech. *Annals of Otology, Rhinology and Laryngology, 116*(4), 304-311.

Searl, J. P. (2002). Magnitude and variability of oral pressure in tracheal speech. *Folia Phoniatrica et Logopaedica*, *54*, 312-328.

Searl, J. P., & Evitts, P. M. (2004). Velopharyngeal aerodynamics of /m/ and /p/ in tracheoesophageal speech. *Journal of voice*, *18*(4), 557-566.

Singer, M. I., & Blom, E. D. (1980). An endoscopic technique for restoration of voice after laryngectomy. *Annals of Otology, Rhinology and Laryngology, 89*(6 Pt 1), 529-533.

Smitheran, J. R., & Hixon, T. J. (1981). A clinical method for estimating laryngeal airway resistance during vowel production. *Journal of Speech and Hearing Disorders, 46*(2), 138-146.

Timcke, R., Leden, H., & Moore, P. (1958). Laryngeal vibrations: Measurements of the glottic wave: Part I. The normal vibratory cycle. *Archives of Otolaryngology - Head and Neck Surgery*, *68*(1), 1-19.

Timcke, R., Leden, H., & Moore, P. (1959). Laryngeal vibrations: Measurements of the glottic wave: Part II. - Physiologic variations. *Archives of Otolaryngology - Head and Neck Surgery, 69*(4), 438-444.

Titze, I. R. (1976). On the mechanics of vocal fold vibration. *Journal of the Acoustical Society of America, 60*(6), 1366-1380.

Titze, I. R. (1980). Comments on the myoelastic-aerodynamic theory of phonation. *Journal of Speech and Hearing Research, 23,* 495-510.

Titze, I. R. (2006). *The Myoelastic-Aerodynamic Theory of Phonation.* Iowa City, IA: National Center for Voice and Speech.

Qi, Y., & Weinberg, B. (1995). Characteristics of voicing source waveforms produced by esophageal and tracheoesophageal speakers. *Journal of Speech Hearing Research, 38*(3), 536-548.

Zemlin, W. R. (1997). *Speech and Hearing Science: Anatomy and Physiology* (4th edition). Boston, MA: Allyn & Bacon.

Appendix A: Chinese consent form for subjects

參與<空氣管食道活塞瓣發聲法之氣流動力學特性研究>同意書

香港大學言語及聽覺科學部誠意邀請您參與由四年級學生陳銘華主理的研究調查。這是一項關於氣管食道活塞瓣發聲法的學術研究，旨在探討其發聲法之發聲來源假聲門的氣流動力學特性。

您將要以自然的聲調及聲量:

1. 朗讀 “(呀/ 衣/ 污/ 柯/ 呃)”；
2. 每個音節朗讀三次, 每次維持大約五秒。
3. 朗讀/$p^{h}ap^{h}ap^{h}a$/音節三次。

朗讀的時候會戴上特製的氧氣口罩以紀錄氣流的數據，這不會對身體有影響。紀錄過程中您可能會有少許不適或疲倦，所以在有需要時您可以隨時稍作休息。整個過程會於香港新聲會進行，需時約三十分鐘。

是次研究並不為閣下提供個人利益，但所搜集數據將對研究假聲門提供寶貴的資料。是次參與純屬自願性質，您可隨時終止參與是項行動，有關決定將不會引致任何不良後果。所收集的資料只作學術及研究用途，原始數據不會包含任何名字及身份的信息，個人資料將絕對保密·所收集之數據將會妥善保存在香港大學言語及聽覺科學部的實驗室內，而只有研究者本人及其研究督導員能看到這些數據. 這些數據將會保存五年，最後，數據將會銷毀。

您有權要求聽回個人的錄音樣本，也有權要求銷毀所有或部分樣本。如您對是項研究有任何問題，請現在提出。

如日後您對是項研究有任何查詢，請與陳銘華聯絡(電話號碼: 9016-0717 /電郵地址: mingwa@hkusua.hku.hk)。你亦可以與其研究之督導員 吳民華教授聯絡(電郵地址:manwa@hku.hk)聯絡如您想知道更多有關研究參與者的權益，請聯絡香港大學非臨床研究操守委員會(2241-5267)。如您明白以上內容，並願意參與是項研究，請在下方簽署。

研究參與者姓名：________________________

研究參與者簽署：________________________ 簽署日期：_______________________

Appendix B: Chinese case history form for subjects

個人資料

姓名: ____________________(中文) __________________ (英文)

性別:__________

年齡:__________ 歲

出生日期: ____________________

你有沒有聲線的問題? □有 □沒有

你有沒有呼吸的疾病? □有 □沒有

你有沒有聽覺的疾病? □有 □沒有

接受手術時間: _____________________________

手術病因(腫瘤位置: 如食道, 咽喉, 鼻咽): _______________

手術後所接受的復康治療 (例:放射治療/化學治療): __________

次數/時間:______________

使用空氣管食道活塞瓣發聲法之年期:_________

人工聲瓣種類: Provox/ Blom-singer/ Voice master

Appendix C: An Informal English consent form for subjects

Examining the neoglottal vibratory pattern of Cantonese tracheoesophageal speakers - A preliminary aerodynamic study using inverse-filtering

You are invited to participate in a research study conducted by the fourth year Speech and Hearing Science student, Chan Ming Wa Joanna in the Division of Speech and Hearing Sciences at the University of Hong Kong.

PURPOSE OF THE STUDY

The proposed research will investigate vibratory behavior of neoglottis with the use of inverse filtering.

PROCEDURES

The recording procedure will take place in the New Voice Club of Hong Kong. It will take about 30 minutes. Participants will be asked to produce 5 Cantonese words (呀/ 衣/ 污/ 柯/ 呢/) for three times for each word. In addition, participants will be asked to produce syllables /pʰapʰapʰa/ for three times. All the productions will be at a comfortable loudness level and speech rate. During speech production, an anesthesia mask will be held tightly against the speaker's face.

POTENTIAL RISKS / DISCOMFORTS AND THEIR MINIMIZATION

The procedures are non-invasive. You may experience some mild fatigue and discomforts during the procedure. Such fatigue and/or discomforts will be kept to a minimum because the tasks are self-paced and you are free to take short breaks.

POTENTIAL BENEFITS

There are no direct benefits to you. However, the research project can provide valuable information on aerodynamic characteristics of Cantonese tone. This information in turn could help better understand production of tones in a tone language.

COMPENSATION FOR PARTICIPATION

Participants will not be given any compensation.

CONFIDENTIALITY

Any information obtained in this study will remain very strictly confidential, will be known to no-one, and will be used for research purposes only. The raw data will be securely transported to the Division of Speech and Hearing Sciences, Faculty of Education at the University of Hong Kong and kept in a secure filing cabinet in a locked room at the University of Hong Kong. No one but the investigator and her supervisor will have access to it. The raw data will be kept for five years after the results are published. The data will be destroyed at the end of the project. Codes, not names, are used on all test instruments to protect confidentiality.

You can review the audio-recording of the procedure. We will erase the entire audiotape or parts of it if you want us to do so.

PARTICIPATION AND WITHDRAWAL

Your participation is voluntary. This means that you can choose to stop at any time without negative consequences

QUESTIONS AND CONCERNS

If you have any questions or concerns about the research, please feel free to contact Chan Ming Wa Joanna at HKU, telephone: 90160717; Email: mingwa@hku.hk. You may also contact my supervisor Dr Lawrence Ng; email: manwa@hku.hk. If you have questions about your rights as a research participant, contact the Human Research Ethics Committee for Non- Clinical Faculties, HKU (2241-5267).

SIGNATURE
I ___________________________________ (Name of Participant) understand the procedures described above and agree to participate in this study.

Signature of Participant:_______________________ Date: ______________________________