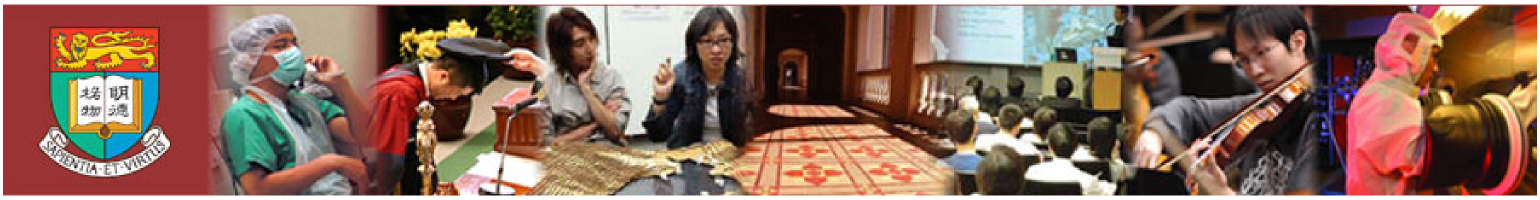

Title	Voice onset time (VOT) characteristics of esophageal, tracheoesophageal and laryngeal speech of Cantonese
Other Contributor(s)	University of Hong Kong.
Author(s)	Wong, Ching-yin, Juliana
Citation	
Issued Date	2007
URL	http://hdl.handle.net/10722/55506
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Voice onset time (VOT) characteristics of esophageal, tracheoesophageal and laryngeal speech of Cantonese

Wong, Juliana Ching-Yin

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, 2007

Abstract

The ability of esophageal (SE) and tracheoesophageal (TE) speakers of Cantonese to differentiate between aspirated and unaspirated stops in three places of articulations were investigated. Six Cantonese stops /p, p^h, t, t^h, k, k^h/ followed by the vowel /a/ produced by 10 SE, TE and laryngeal (NL) speakers were examined through perceptual judgement tasks and voice onset time (VOT) analysis. Results from perceptual experiment showed lower identification accuracy in SE and TE than NL speech for all stops. Misidentification of aspirated stops as their unaspirated counterparts was the dominant error. Acoustic analysis revealed that aspirated stops produced by NL, SE and TE speakers were associated with significantly longer VOT values than their unaspirated counterparts. Velar unaspirated stops showed significantly longer VOT values than bilabial and alveolar stops in NL and SE speech. In conclusion, SE and TE speakers were still able to use VOT to signal aspiration contrast, but TE was unable to differentiate among different places of articulation.

Introduction

Total laryngectomy is a surgical procedure involving the removal of the entire larynx. It is a treatment method to remove the carcinoma in person with laryngeal cancer. After the surgery, the patient's phonatory and respiratory system will be severely altered. Since swallowing function is always preserved after the procedures, restoration of voice and speech, which is critical to one's quality of life and psychosocial state, will be the focus of rehabilitation (Blom, 2000; Lewin, 2004). There are currently four major methods of alaryngeal speech: pneumatic artificial device (PA), electrolarynx (EL), standard esophageal (SE) and tracheoesophageal (TE) speech. While the first two methods depend on an external device, the latter two are internal. With SE and TE speech, the pharyngoesophageal (PE) segment replaces the removed larynx and become the neoglottis, functions as the new voicing source (Dworkin, Meleca, Simpson, Zormeier, Garfield, Jacobs & Mathog, 1999).

SE phonation involves the intake of air via injection and inhalation, into the cervical esophagus, which acts as the air reservoir. When air expelled, the PE segment is set into vibration (Salmon, 1994). The prevalence of TE phonation as a voice reconstruction started from the mid-20th century. To date, 77-100% of patients can use it successfully (Lam, Ho, Ho, Ng, Yuen, & Wei, 2005). In TE phonation, a valved silicon prosthesis and tracheoesophageal (TE) puncture is introduced to the person during the surgical procedures, the prosthesis allows pulmonary air to flow into the esophagus when the tracheostoma is occluded (Blom, Ronald & Hamaker, 1996). The PE segment is then set into vibration and ready for phonation. With the use of PE segment as the new voicing source, TE and SE sound characteristics should be affected by the different physiologies.

Listeners' confusions and misidentification of stops produced by SE speakers of English have been well investigated. Hyman (1955) reported that voiceless stops were more easily misidentified as its voice cognates than the opposites, and the stop categories were the second least correctly identified types of consonants. Such confusions were also reported by other researchers (Hirose, 1996; Nichols, 1876; Shames, Font & Matthews, 1963). In recent years, with the prevalence of TE speech, its perceptual stops voicing distinction was also reported (Jongmans, Hilgers, Pols & van As-Brooks, 2006; Lundstrom & Hammerburg, 2004; Searl, Carpenter & Banta, 2001). Similar to SE speech, confusions in perceptual stops voicing contrast was prevalent.

Voice onset time (VOT) has been described as significant acoustic correlates and perceptual cues to signify timing contrast between voiced and voiceless stops (Lisker & Abramson, 1964; 1967; Lisker, 1975; Klatt, 1975; Zlatin, 1974). It is defined as the time interval between the release of a stop closure and the onset of phonation (Lisker & Abramson, 1964). It physiologically reflects the coordination of timing between the articulatory system and the phonatory system. The release of stop closure is related to the supralaryngeal articulators, such as the lips, tongue tips and tongue dorsum; while the onset of phonation is related to the larynx. Along the voicing continuum, the onset of phonation of voiced stops happens preceding or simultaneously with the stop release, while the onset of voiceless stops is delayed by 20 ms. Based on the VOT value, voicing in a CV syllable can be classified into three types: (1) voicing lead, with voicing begins 75-125 ms before the release, (2) short lag, with voicing begins between 0-25 ms after the stop release, and (3) long lag, with voicing begins between 60-100 ms after the stop release (Lisker & Abramson, 1964). For NL speakers, the VOT values are longer in voiceless stops than their voiced cognates and aspirated stops are longer than their unaspirated counterparts (Lisker & Abramson, 1964; Zlatin, 1974). According to the classification, a stop will be perceptually identified as voiceless when VOT falls within the long lag range, and will be identified as voiced when voice lead or voice lag lies within the long lag range (Lisker & Abramson, 1964).

VOT characteristics of SE and TE of English and its contribution to voicing contrast had been described in the literature. SE speakers were generally reported to be associated with shorter VOT values than NL speakers. Christensen, Weinberg and Alfonso (1978) found that the VOT of 42% of the voiceless stops /p/, /t/, /k/ produced by SE speakers were less than 25 ms, and therefore being perceptually identified as voiced /b/, /d/, /g/. SE speakers had VOT values shorter than NL speakers. Connor, Hamlet and Joyce (1985) discovered significantly longer VOT values for voiceless bilabial stops /p/ than those for its voiced cognate /b/ in high intelligibility SE speakers, but contrasts were not shown in low intelligibility production. However, contradictory results were reported for TE speakers. Saito, Kinishi and Amatsu (2000) described that TE speakers with high intelligibility demonstrated longer VOT values than NL speakers for voiceless stops /p/ but had comparable values for voiced stops /b/. Searl and Carpenter (2002) also reported that voiceless stops produced by TE speakers demonstrated longer VOT values than voiced stops and stops in TE speech demonstrated longer VOT values than NL speakers for voiceless aspirated stops /p/ and /t/. Similar results were reported for their voiced cognate /b/ but not for /d/. A small number of studies compared stops produced by the three groups of speakers. Robbins, Christensen and Kempster (1986) reported that the overall VOT values of /p/ and /k/ for TE and SE speakers were shorter than those for NL speakers. TE speakers showed longer VOT value than SE speakers. However, a recent study on Hebrew by Most, Tobin and Mimran (2000) found that the mean VOT values of voiced labial stops /b/ and its voiceless cognate /p/ were not statistically significantly different among NL, good SE, moderate SE and TE speakers. They explained that the small number of subjects (each group was comprised of only five members) led to large individual variation, which might make the result less valid.

Several researchers hypothesized about the earlier onset of voicing in English voiceless stops in SE speech in relation to the limited abductor-adductor properties of the PE segment (Christensen, et al, 1978; Christensen & Dwyer, 1990; Gandour, Weinberg, Petty & Dardarananda, 1987). Robbins et al. (1986) provided further evidences in support of this hypothesis by stating that, using the PE segment as the new phonatory source, both SE and TE speakers, showed significantly shorter lag time than NL speakers. However, further evidences would be required to understand the mechanism of PE segment.

In addition to voicing contrast, VOT also varies with the place of articulation. This is true at least in NL speakers. Comparing among bilabial, alveolar and velar stops, velar stop generally exhibited the longest and bilabial the shortest VOT values for voiced and voiceless stops (Klatt, 1975; Lisker & Abramson, 1964). This characteristic was not only exhibited in English stops, but was proposed to be universal by cross-linguistic study (Cho & Ladefoged, 1999). The explanation for the VOT variations had been discussed by several literatures with respect to the areas of laws of aerodynamics (Hardcastle, 1973), speed of movement of various articulators (Hardcastle, 1973), extent of articulator contact (Stevens, 1998) and temporal adjustment between stop closure duration and VOT (Maddieson, 1997). VOT characteristics in relation to place of articulation in SE speakers were only reported by Christensen et al. (1978), who stated SE speakers also produced stops with VOT values sensitive to the place of articulation.

As VOT reflects the timing between supralaryngeal articulator and vocal fold activities, it is associated with the variation among places of articulation and the voicing continuum. A comprehensive study about VOT characteristics among SE, TE and NL speakers should include comparison of stop voicing contrasts among different places of articulation. However, only Christensen et al. (1978) had compared voiced and voiceless stops among the three places of articulation, but TE speakers were not included in their subject samples. Though Robbins et al. (1986) had compared VOT values of both SE and TE speakers, the types of stops included was only /p/ and /k/. A comprehensive study about VOT values of stops produced by the three groups of speakers among different places of articulation was lacking.

While previous researches mainly focused on English initial stops, which were distinguished by voiced-voiceless contrast, study about language with stops distinguished by voiceless aspiration contrast was rare. The study in Thai alaryngeal speech conducted by Gandour et al. (1987) showed that there was considerable overlap in VOT values of the voiceless aspirated and voiceless unaspirated stops over the three places of articulation in their female SE speaker. In addition, their perceptual analysis revealed that listeners could only identify the voiceless aspirated stops /p^h/, /t^h/, /k^h/ with the percent of correct-identification below 17%. By comparing the data from English, the authors concluded that, out of the three stop contrasts, Thai SE speakers could only distinguish between voiced-voiceless contrast and not aspiration contrast. Nevertheless, owing to the small sample of subjects in their study (only two SE speakers were involved), and one datum could not be analyzed due to technical errors, generalization of the results of SE speakers' difficulty in signifying aspiration contrast cannot be made. Therefore, an investigation of VOT characteristics in Cantonese, which is a language with initial stop contrasted by aspiration, in SE and TE speakers will provide more evidences in support of this conclusion.

Cantonese initial stops are all voiceless, which are contrasted by aspiration (Clumeck, Barton, Macken & Huntington, 1981; Hashimoto, 1972; Lisker & Abramson, 1964) at three places of articulation, which are bilabial (/p/, /p^h/), alveolar (/t/, /t^h/) and velar (/k/, /k^h/) respectively. Based on the data from eight Cantonese NL speakers, Clumeck et al. (1981) reported that the mean VOT values for unaspirated labial, dental and velar stops produced in isolated words are 9.98, 10.46 and 25.99ms respectively. For the aspirated series, the values are 74.22, 83.22 and 90.75 ms respectively (Clumeck et al., 1981). The VOT values were sensitive to place of articulation, with labial shortest and velar the longest, which was similar to the pattern of English stops (Lisker & Abramson, 1964).

A considerable amount of research in Cantonese alaryngeal speech had been conducted in area such as tone perception (Ching, Williams & Van Hasselt, 1994; Ng, Lerman & Gilbert, 1998), overall speech intelligibility (Ng, Kwok & Chow, 1997; Yiu, van Hasselt, Rhys Williams & Woo, 1994), perceptual characteristics (Ng et al, 1997), fundamental frequency, intensity and vowel duration (Ng, Lerman & Gilbert, 2001). However, knowledge of VOT characteristics of alaryngeal speech is still lacking.

Based on the above discussion, the present study attempted to compare the VOT characteristics of Cantonese SE, TE and NL speakers, and to determine, if SE and TE speakers, who both used PE segment as the phonatory source, can distinguish aspiration contrast and place of articulation in stops when comparing with the performance of NL speakers.

Method

Speakers and Listeners

Ten SE, 10 TE and 10 NL speakers participated in the study. All speakers were male and age-matched. The SE speakers were from 52 to 75 years old (mean =65.5 years), with post operative period of 2 to 21 years (mean = 10 years) and duration of using the speaking method ranged from 0.6 to 20 years (mean = 8 years). The TE speakers were ranged from 49 to 84 years old (mean = 69.4 years), with post operative period 2-18 years (mean = 6years), with duration of using TE speech ranged from 2 to 18 years (mean = 6 years). The NL speakers were with ages ranging from 53 to 83 years, with a mean age of 66 years. The alaryngeal speakers were selected from the New Voice Club of Hong Kong, which is a non-profit self-help organization for the laryngectomees in Hong Kong. Proficiency criteria on the use of the alaryngeal speaking method were not imposed as to include a wider range of speech proficiency to facilitate generalization. The only inclusion criteria for the alaryngeal speakers were having no history of speech and language disorders except those associated with laryngectomy, and the speaking methods were their primary mode of verbal communication. All participants were literate and were able to read the speech material used in the study.

Twenty native Cantonese speakers who had no previous exposure to any alaryngeal speech were recruited as listeners. Their ages ranged from 18 to 50 years old (mean = 23 years). They had passed the hearing screening at 25 dB for 0.5, 1, 2 and 4 kHz.

Speech Materials

The speech materials included six CV syllables produced at the high-level tone. Formed by the six Cantonese initial stops followed by the vowel /a/, the actual syllables and their meanings (enclosed and parentheses) were /pa/ (“father”), /p^{h}a/ (“to prostrate”), /ta/ (“a dozen”), /t^{h}a/ (“he”), /ka/ (“home”) and /k^{h}a/ (“compartment”). The vowel /a/ was chosen because, according to Robbins et al. (1986), SE, TE and NL speakers differ more significantly in VOT for the vowel /a/ than other vowels /i/ and /u/. During the experiment, each syllable was produced embedded in a carrier phrase [/ŋɔ3/ /jiu3/ /tʌk6/ ____ /pei2/ /nei3/ /t^{h}eŋ1/], meaning “I want to read _____ to you”. The carrier phrase was used to eliminate the possible contextual effect. When producing this carrier phrase in Cantonese, each citation word was preceded by a voiceless unreleased stop consonant /k/, and followed by a voiceless unaspirated stop /p/. This facilitated the identification of the target syllable during acoustic analysis.

Equipment and Recording procedures

All recordings were conducted in a silent room located in the New Voice Club of Hong Kong. The speech samples were recorded on high-quality mini-disc (Sony MDW74CRX) via a minidisk recorder (Sony MZ N710) via a unidirectional microphone

(Bruel & Kjaer Type 2812 MKII Two Channel Microphone preamplifier). The mouth-to-microphone distance was maintained at 15 cm. During the recording, the speakers were instructed to read aloud each syllable in the carrier phrase three times at a normal speaking rate and a comfortable loudness level. Each carrier phrase was printed on separate cards, and the cards were presented individually in a random order. Instructions and practice time was provided to the participants before each recording in order to familiarize the speakers with the experimental procedures. Upon completion of the recording, the 540 speech samples (30 speakers × 6 stops × 3 repetitions) were digitized at a sampling rate of 20 kHz and stored in a computer as WAVE files for the listening task.

Listening Task

The listening task was conducted in a sound-treated booth at the University of Hong Kong. The listeners were seated in the booth and the randomized speech samples were presented at a comfortable loudness level (approximate 70-75 dB SPL) via high quality headphones (Sony MDR-EX32LP). The listeners were provided with an answer sheet on which they circled the citation word they had heard. The listeners were instructed to choose among a closed set of six choices provided in the answer sheet. The type of speakers was not identified for the listeners. Inter-stimulus pauses of three seconds were provided during the listening tasks. Five trials were provided as practice to the listeners in order to familiarize themselves with the experimental procedures.

VOT measurements

VOT (in milliseconds) values were measured from the digitized speech samples using a speech analysis software (Pratt v. 4.3). VOT was defined as the time interval between the burst of a stop and the onset of the following vowel (Lisker & Abramson, 1964). Lisker and Abramson (1964) obtained VOT measurements based on wide-band spectrograms. However, a recent study by Francis, Ciocca and Yu (2002) suggested that waveform-based measurement of VOT values appeared to be more accurate and reliable than spectrogram-based measurement. Therefore, waveform and wide-band spectrogram (window size = 5 ms) were used together in the analysis. During analysis, the waveform was visually inspected, and cursors were carefully placed on the waveform. The first cursor marked the abrupt change in waveform to indicate the release of burst, and the second cursor was placed at the beginning of the first identifiable period of the second formant (Lisker & Abramson, 1964). The time interval marked by the two cursors indicated the VOT value.

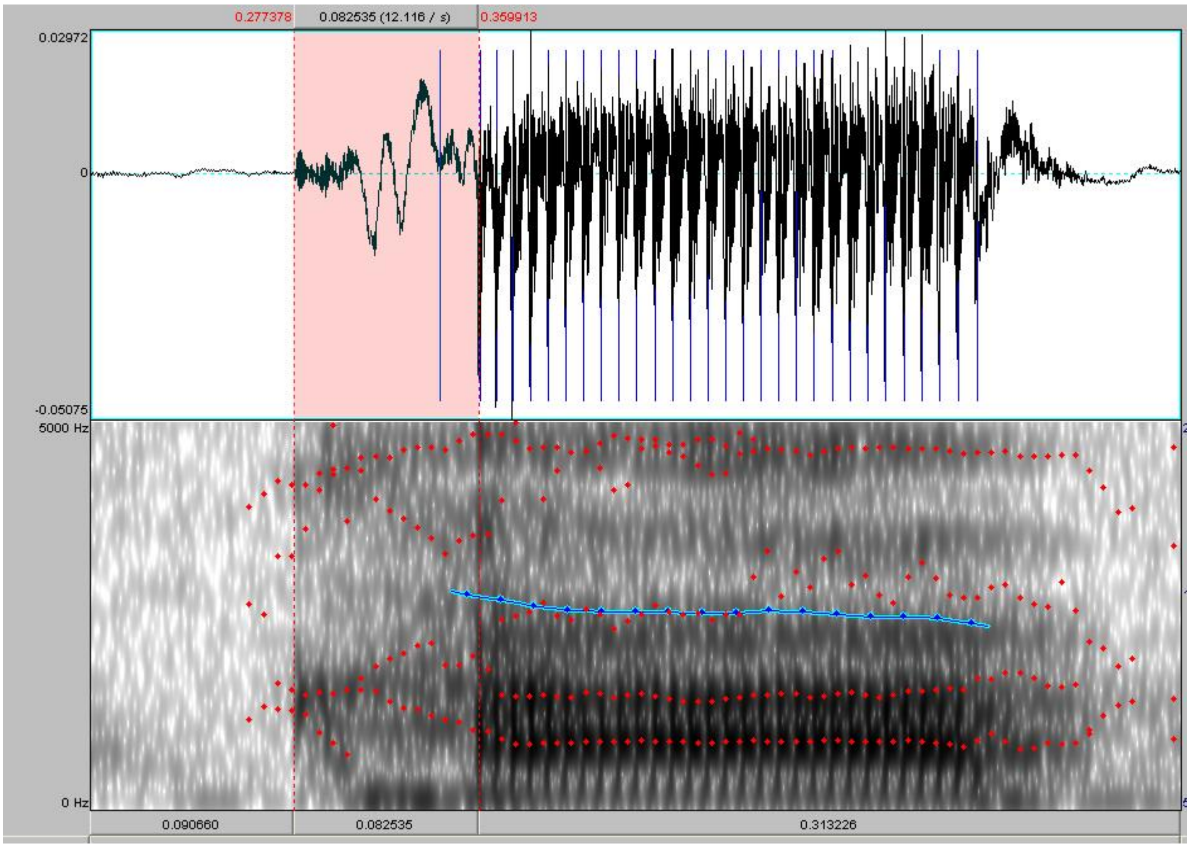

Figure 1. An example of VOT measurement obtained from an aspirated stop produced by a NL speaker. The upper panel shows the waveform whereas the lower panel shows the wide-band spectrogram.

Reliability Measures

For the listening task, intra-listener reliability was obtained to ensure consistency in their responses. The listeners listened to a set of 108 speech samples (20% of the entire data corpus) that were randomly selected a second time. They were not informed that the speakers were repeated. Percentage of agreement was calculated based on the data obtained from the first and second responses. All listeners demonstrated an agreement with higher than 72% (mean = 83.366%; SD = 5.079). A high consistency in their judgement was indicated and therefore their responses were acceptable to be included in the perceptual analysis.

For VOT measurement, intrajudge and interjudge reliability was obtained to determine the reliability of VOT measurements. To measure intrajudge reliability, a set of 108 speech samples (20% of the total samples) was randomly selected and measured a second time by the primary investigator. The Pearson Product-Moment Correlation coefficient (r) obtained between the first and second VOT measurements made by the primary investigator was 0.94 ($p < 0.01$).

Similar to intrajudge reliability measure, a second set of 108 speech samples (20% of the total samples) was randomly selected and measured by another investigator for interjudge reliability. Pearson Product-Moment Correlation coefficient (r) between the VOT measurements made by the first and second investigators was 0.89 ($p<0.01$). The high correlations for both intrajudge and interjudge reliability measure indicate that the VOT values measured by the primary investigator were consistent and reliable.

Results

Perception

Tables 1-3 summarize the percent of correct identifications of the word-initial stops produced by NL, SE and TE speakers of Cantonese. As shown in Table 1, all stops produced by NL speakers were identified with high level of accuracy (≥ 80%), with /t/, /k/ and /k^h/ a near-perfect level (≥ 96%). According to Tables 2 and 3, the unaspirated stops /p, t, k/ produced by SE and TE speakers achieved acceptably high accuracy (68%-85%). However, the accuracy was much lower for aspirated stops (29.4%-55.5%) as compared with their unaspirated counterparts and that produced by NL speakers.

Table 1

Percent of identification of stops produced by NL speakers

	Stimuli					
Response	/p/	/t/	/k/	/p^h/	/t^h/	/k^h/
/p/	92.5	2.3	0.2	0.2	0.3	0
/t/	6.5	96.3	3.3	0	0.2	0.2
/k/	0.5	1.2	96.3	0	0.2	1.3
/p^h/	0.5	0	0	82.2	2.7	0
/t^h/	0	0.2	0	17	86.5	1.7
/k^h/	0	0	0.2	0.6	10.1	96.8

Table 2

Percent of identification of stops produced by SE speakers

	Stimuli					
Response	/p/	/t/	/k/	/p^h/	/t^h/	/k^h/
/p/	77.4	14	0.3	49	7.2	0.2
/t/	7.3	68.2	4.4	4.9	36	0
/k/	3	3	82	3.9	6.2	41.5
/p^h/	9	2.3	0.3	29.4	0.3	0
/t^h/	1.3	10.3	1.8	7.2	36	2.8
/k^h/	2	2.2	11.2	5.6	7.8	55.5

Table 3

Percent of identification of stops produced by TE speakers

	Stimuli					
Response	/p/	/t/	/k/	/p^h/	/t^h/	/k^h/
/p/	84	3.8	0.5	62.5	4.2	0
/t/	1.3	67.7	1	0.8	38.1	0.5
/k/	0.8	2	84.7	0.2	5	47.7
/p^h/	13.4	3.2	0.5	35	2.5	0
/t^h/	0.3	22.5	0.5	1.3	48	1.8
/k^h/	0.2	0.8	12.8	0.2	2.2	50

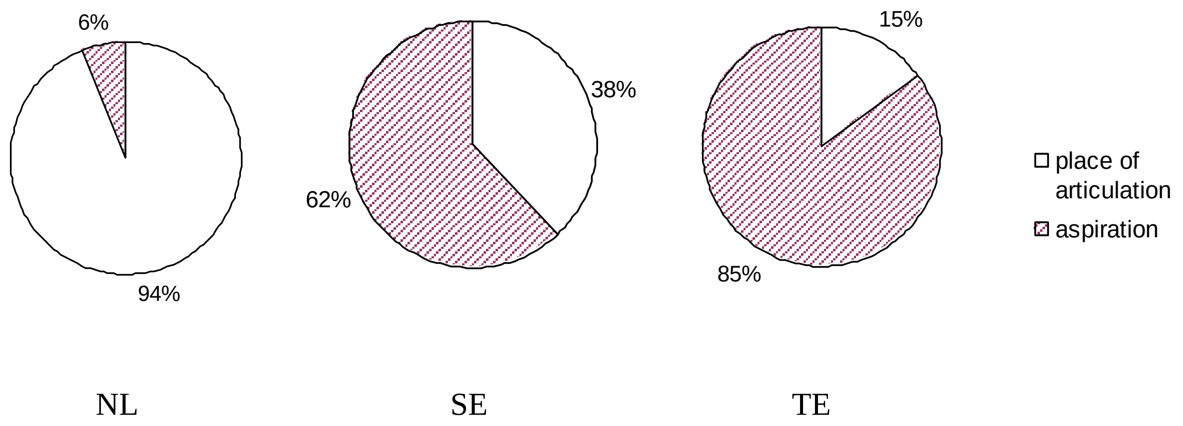

Figure 2. Comparison of types of errors in identification of stops produced by NL, SE and TE speakers

As shown in Figure 2, the dominant error pattern of misidentification for stops produced by NL speakers was the place of articulation (94%). Aspiration errors were the predominant type of error for the SE (62%) and TE (85%) groups.

Table 4

Mean and Standard Deviation of VOT values (ms) of stops in NL, SE and TE speech

	NL		SE		TE	
	Mean (ms)	SD (ms)	Mean (ms)	SD (ms)	Mean (ms)	SD (ms)
/p/	17.92	5.78	26.56	19.44	34.99	32.57
/t/	17.89	2.97	23.66	8.64	35.16	19.98
/k/	33.08	7.59	44.77	15.75	44.20	11.97
/p^h/	77.62	21.42	75.61	52.20	67.53	30.64
/t^h/	84.09	20.20	69.01	45.04	66.06	38.44
/k^h/	91.89	26.69	86.06	56.89	86.96	48.44

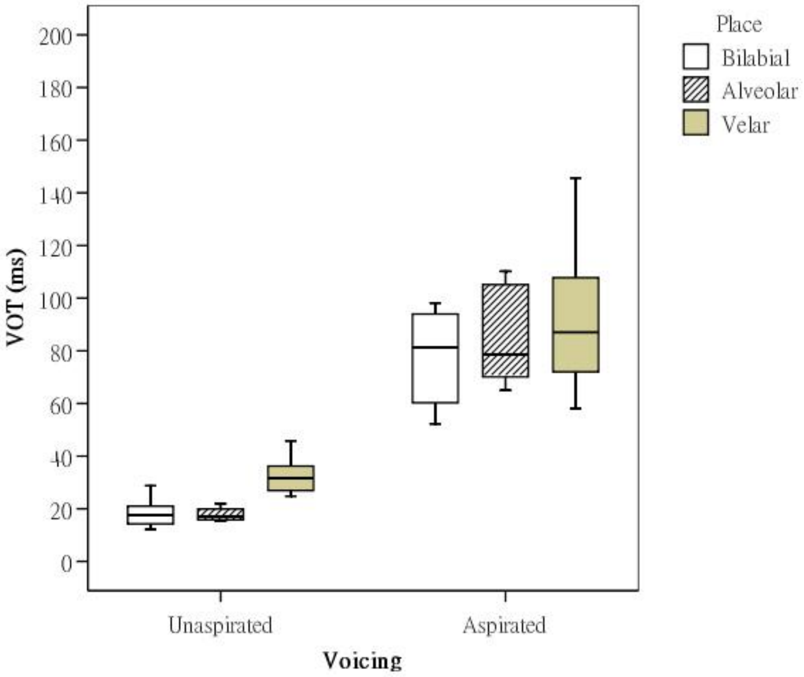

Figure 3. Average VOT values for aspirated and unaspirated stops for different place of articulation produced by NL speakers

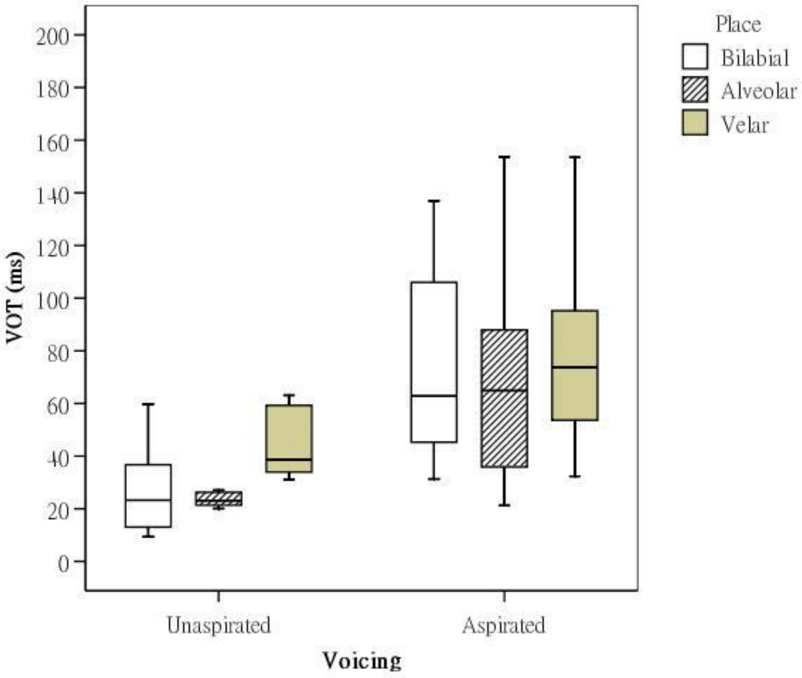

Figure 4. Average VOT values for aspirated and unaspirated stops for different place of articulation produced by SE speakers

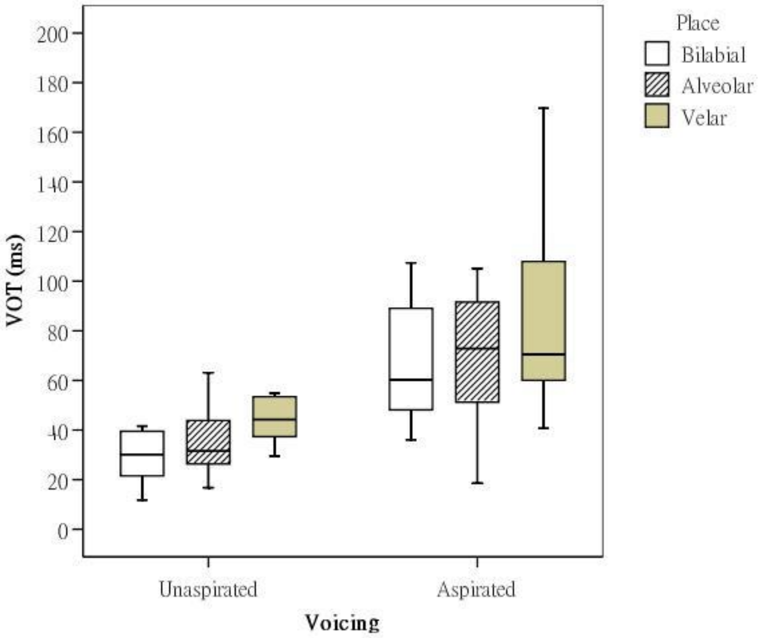

Figure 5. Average VOT values for aspirated and unaspirated stops for different place of articulation produced by TE speakers

Production

The mean VOT values for the word-initial stops produced by the NL, SE and TE speakers are summarized in Table 4. Results of a 2 × 3 × 3 (aspiration × phonation type × place of articulation) three-way Analysis of Variance (ANOVA) indicated significant interaction effects between phonation type and aspiration [$F\ (2, 179) = 4.129, p < 0.05$]. Therefore, one 2 × 3 (aspiration × place of articulation) ANOVAs and two one-way (place of articulation) ANOVAs were carried out to test for statistically significant differences in VOT values between aspirated and unaspirated stops, and among bilabial, alveolar and velar stops within each group of speakers.

The mean VOT values for aspirated and unaspirated stops produced at different places of articulation by NL speakers are illustrated in Figure 3. As illustrated, all aspirated stops were associated with a longer VOT than its unaspirated cognates in all three places of articulation. The average VOT values of aspirated and unaspirated stops were 84.533 ms and 22.963 ms respectively. A 2 × 3 (aspiration × place of articulation) two-way ANOVA revealed no significant interaction between aspiration and place of articulation [$F\ (2, 59) = 0.341, p > 0.05$]. Yet, significant main effects were found in aspiration [$F\ (1, 59) = 236.659, p = 0.000$] and in place of articulation [$F\ (2, 59) = 4.966, p = 0.01$]. Aspirated stops displayed VOT values that were significantly longer their unaspirated cognates, regardless of the place of articulation. Results of a Tukey HSD post-hoc test revealed that velar stops were associated with significantly longer VOT values than alveolar and bilabial stops. Two one-way ANOVAs were performed for each voicing type. For unaspirated stops, significant differences were found between different places of articulation [$F\ (2, 29) = 29.988, p = 0.000$]. Tukey HSD post hoc test revealed significantly longer VOT values for the velar stop /k/ than the bilabial stop /p/ and the alveolar stop /t/. For aspirated stops, VOT values associated with different places of articulation were not statistically significant [$F\ (2, 29) = 1.123, p > 0.05$].

Average VOT values for aspirated and unaspirated stops produced at different places of articulation by SE speakers were illustrated in Figure 4. Similar to NL speakers, the VOT values of aspirated stops were longer than their unaspirated cognates regardless of the place of articulation. The mean VOT values of SE aspirated and unaspirated stops were 76.895 ms and 31.670 ms, respectively. A 2 × 3 (aspiration × place of articulation) two-way ANOVA revealed no significant interaction effect between aspiration and place of articulation [$F\ (2, 59) = 0.076, p = 0.927$]. No significant main effect was found for place of articulation [$F\ (2, 59) = 2.020, p = 0.143$]. However, significant main effect was found for aspiration [$F\ (1, 59) = 31.434, p = 0.000$], revealing that aspirated stops were associated with significantly longer VOT values than unaspirated stops. Two one-way ANOVAs were performed within each voicing type. For unaspirated stops, significant differences were found for places of articulation [$F\ (2, 29) = 8.071, p < 0.05$]. A Tukey HSD post-hoc test revealed significantly longer VOT values for velar /k/ than the bilabial /b/ and the alveolar /t/. For aspirated stops, differences in place of articulation was not statistically significant [$F\ (2, 29) = 0.413, p > 0.05$].

The mean VOT values for aspirated and unaspirated stops produced at different places of articulation by TE speakers are illustrated in Figure 5. Similar to NL speakers, the aspirated stops were longer than their unaspirated cognates regardless of the place of articulation. The mean VOT values of TE speech for aspirated and unaspirated stops were 74.627 ms and 38.118 ms respectively. A 2 × 3 (aspiration × place of articulation) two-way ANOVA revealed no significant interaction effect between aspiration and place of articulation [$F\ (2, 59) = 0.215, p = 0.807$]. No significant main effect was found for place of articulation [$F\ (2, 59) = 1.836, p = 0.169$]. However, significant main effect was found in aspiration [$F (1, 59) = 28.747, p = 0.000$], indicating that aspirated stops were associated with significantly longer VOT values than unaspirated stops, which was similar to the results obtained from NL and SE speech. Two one-way ANOVAs were performed within each voicing type. No significant differences were found in place of articulation for both aspirated [$F (2, 29) = 1.248, p > 0.05$] and unaspirated stops [$F (2, 29) = 0.982, p > 0.05$]. Therefore, places of articulation were not significantly distinguishable in TE speech for both aspirated and unaspirated stops.

Two F-tests were carried out to test for significant differences among phonation types for aspirated and unaspirated stops. For aspirated stops, no significant differences were found among the three groups of speakers [$F (2, 89) = 0.873, p > 0.05$]. Yet, for unaspirated stops, significant differences were found [$F (2, 89) = 8.647, p = 0.00$]. A Tukey post hoc HSD test revealed that TE speech was associated with longer VOT values than NL speech.

Discussion

The present study aimed at investigating the VOT characteristics of word-initial stops produced Cantonese NL, SE and TE speakers. VOT values associated with voiceless aspirated and unaspirated stops produced at different places of articulation were compared. In addition, the effect of using PE segment as the new voicing source in SE and TE speech on the VOT values was investigated.

Perception

As indicated by the results from the perceptual judgement task, Cantonese SE and TE speech failed to signal voiceless aspirated and unaspirated stops contrast as successfully as NL speech. The major error was misidentification of aspirated stops as their unaspirated counterpart for both types of speakers (see Figure 2). Similar results were reported in alaryngeal speech research on language with aspiration contrast such as in Thai (Gandour et al., 1987) and Mandarin (Liu et al., in press). However, as results on perception of stops produced by TE speakers are not available in language with aspiration contrast, comparison could not be made. Although the distinction among English stops is different (they differ from each other by voicing) from that in Thai, Mandarin and Cantonese (they are different from each other by aspiration), both can be distinguished by VOT. Reports on English stops showed perceptual voicing confusions in SE speech (Hyman, 1955; Nichols, 1976) and TE speech (Jongmans et al., 2006; Lundström & Hammerburg, 2004; Searl et al., 2000). The confusion in voicing contrast was related to the new voicing source, the neoglottis, adopted by SE and TE speakers. It has been suggested by researches that the motor control of the neoglottis is limited when compared with control of the larynx (Christensen, et al., 1978; Christensen & Dwyer, 1990; Gandour, et al., 1987; Robbins et al., 1986). In order to produce aspirated or voiceless stops, the glottis has to stay open for a longer period of time after the release of stops. With abductor-adductor properties in the larynx, NL speaker have better control over the timing of glottal opening, thus able to produce the aspiration contrast without much effort. In contrast, SE and TE speakers would exhibit greater difficulty in controlling the timing of neoglottal opening due to the limited control. Thus, confusion in aspiration or voicing contrast was resulted. The relatively high percent of identification for unaspirated stops than aspirated stops in SE and TE speech was in agreement with Gandour et al. (1987) and Liu et al. (in press). This indicates that SE and TE speakers of Cantonese were able to produce voiceless unaspirated stops which were intelligible to listeners.

The percent of correct identification of stops produced by NL speakers were not 100%. An average of 8% of identification error was found and the predominant errors were associated with the misperception of place of articulation (see Figure 2). This was consistent with the findings reported by Searl et al. (2001), who reported that NL speakers committed with 12% identification error and the predominant perceptual error type was manner of articulation, such as misidentified stops /p/ as the fricatives /f/, and those errors committed by

TE speakers were mainly voicing contrast. In the present study, though the type of possible errors was restricted to either place of articulation or aspiration contrast due to the closed choice in the listening task, possible reasons could be proposed to explain the error pattern. The intelligibility of laryngeal speakers could be affected by factors such as speech stimuli and listening task (Yorkston & Beukelman, 1980). The CV syllables were embedded in a meaningless carrier phrase with which contextual support was lacking. In addition, with the inter-stimulus silence interval of 3 seconds, listeners had to make a quick judgement of the target phonemes. These made the tasks more demanding. However, rare occurrences of aspiration confusion made by the listeners signalled that NL speakers, with their larynx present, rarely resulted in perceptual errors on the voicing continuum. The aspiration confusion was distinctive to SE and TE speakers.

Production

Phonation Type. In the present study, TE speech was associated with significantly longer VOT values than NL speech for voiceless unaspirated stops. Similar results were reported by Saito et al. (2000), who compared the production of Japanese voiceless unaspirated bilabial stop /b/ by 40 TE speakers and 10 NL speakers. They reported that the average VOT values for TE speakers of high intelligibility (26.2 ms) was longer than NL speakers (17.2 ms). However, different pattern was shown in aspirated stops for the present study. The mean VOT values of aspirated stops produced by NL, SE and TE speakers were 84.53 ms, 76.89 ms and 73.52 ms, respectively. Although the present data failed to show significant difference in aspirated stops produced by the three speaker groups, mean VOT values showed that NL speakers had longer VOT than the alaryngeal speakers. This is in line with data reported previously. Previous studies examining voicing distinction of English stops showed that voiceless aspirated stops in SE speech were associated with significantly shorter VOT values than NL speech (Christensen et al., 1978; Robbins et al., 1986). Liu et al (in press) observed that VOT values associated with NL speakers were significantly longer than SE speakers of Mandarin in aspirated stops. However, opposing results were reported by Searl and Carpenter (2002). Although the differences were not substantial, TE speakers were found to be associated with a longer VOT than NL speaker (NL: 30 ms vs. TE: 22 ms for /p/; NL: 41 ms vs. TE: 37 ms for /t/).

In general, for stop production, intraoral pressure is established when articulators, such as lips and tongue, come to an occlusion. The vocal folds are probably abducted during this period of time, allowing intraoral pressure to build up. At this stage, the pressure in the supraglottic (intraoral) region should be similar to that in the subglottal region. The pressure difference between supraglottal and subglottal regions is known as the transglottal pressure differential. Depending on the type of stop being produced, the vocal folds may become adducted before or after the release of oral occlusion begins. Upon release of the oral occlusion, intraoral air pressure starts to drop. Transglottal pressure differential begins to increase (Stevens, 1998). The vocal folds will be set into vibration once the transglottal pressure differential reaches a critical level (about 3-5cm H_2O). This is called the phonation threshold pressure (PTP) (Fisher & Swank, 1997).

The contradictory pattern of VOT values in aspirated and unaspirated stops produced by three groups of speakers suggested different mechanisms in PE segment in producing these two stops. It is suggested that PE segment is normally constricted (Robbins et al, 1986). Unlike the vocal fold, which has to be adducted to begin vibration, it is hypothesized that the PE segment has to be relaxed. In the production of unaspirated stops, the time for the release of intraoral pressure is short. Therefore, the onset of vibration depends largely on the timing of relaxation of the PE segment. Since the motor control of the neoglottis is limited (Christensen, et al., 1978; Christensen & Dwyer, 1990; Gandour, et al., 1987; Robbins et al., 1986), it is hypothesized that the time taken for relaxing the neoglottis in SE and TE speakers is longer when compare with the adduction of the vocal folds in NL speakers. Therefore, longer VOT values are resulted. In the production of aspirated stops in NL speakers, Stevens (1998) stated that the vocal folds are abducted before the release of the oral occlusion to allow for aspiration. An adductory force is later applied to enable the glottis to reach a size ready for phonation after the stop released (Stevens, 1998). When compared with abducting in vocal folds, the time taken for relaxation of the PE segment would be shorter than the adduction of the vocal folds. Therefore, shorter VOT values are resulted in aspirated stops produced by SE and TE when compared with NL speech. However, the relative degree of opening and control, and other physiology related to the PE segment was still obscure. Future study making use of advanced imaging techniques should provide further information on the capability of the neoglottis in making articulatory adjustments.

Despite the explanation above, there is also another reason to account for this. For unaspirated stops, several researchers had reported higher intraoral pressures in TE speakers than NL speakers (Motta, Galli & Di Rienzo, 2001; Saito et al., 2000). Accordingly, a stop with higher intraoral pressure during the occlusion phase should take a longer time for transglottal pressure differential to reach the PTP for onset of vocal fold vibration, and a higher VOT values results. Though the present data failed to indicate significant differences in VOT values between SE and NL speech, the mean VOT values was greater in SE (31.66ms) than in NL (22.96ms). This was consistent with data reported by Hirose (1996), who noted a higher VOT values in Japanese voiceless stops produced by SE speakers than NL speakers. Hirose’s result was supported by a higher intraoral pressure values measured in SE speakers when compared with NL subjects. However, contradictory data were observed by Liu et al. (in press). Liu et al. found that VOT values of voiceless unaspirated stops exhibited by NL (13.74 ms) and SE (16.85 ms) speakers of Mandarin were comparable. No significant difference was present in the VOT values of voiceless unaspirated stops produced by NL and SE speakers of Mandarin. In addition, study comparing intraoral air pressure in aspirated stops produced by SE and TE speakers was lacking. Future aerodynamics studies the comparing intraoral pressure, transglottal airflow, subneoglottal air pressure in aspirated and unaspirated stops produced by SE and TE speakers would shed light to this issue and help us better understand the mechanism of neoglottal vibration.

Aspiration. For NL speech, aspirated stops were associated with longer VOT values than their unaspirated counterparts (Lisker & Abramson, 1964). Results from the present study were in agreement with this finding (Figures 3-5). As compared with the results reported by Clumeck et al. (1981) in Cantonese speaking laryngeal adults, a longer mean VOT was found for voiceless unaspirated stops (22.96 ms vs. 15.48 ms) and a comparable VOT for voiceless aspirated stops (84.53 ms vs. 82.73ms). Referring to Figures 3, 4, and 5, all three groups of speakers exhibited longer VOT values for aspirated stops than unaspirated stops. Results of ANOVAs also revealed significant differences in VOT values between the aspirated and unaspirated stops. It follows that, using PE segment as the new sound sources, SE and TE speakers could still vary the VOT values in signalling aspiration contrast.

Though significantly longer VOT values were found for aspirated stops, percent of correct identification of aspirated stops for both SE and TE speakers were fairly low (29.4%-55.5%), with the predominant error being confusing it with its unaspirated counterparts. As indicated by Lieberman and Blumstein (1988), perception of VOT was categorical. Discrimination of sounds was made only if listeners can identify them to be in distinct categories. According to Lisker and Abramson (1964), the perceptual category of VOT of unaspirated stops was in the short lag range (between the 0-25 ms), and that of aspirated stops in the long lag range (between 60-100 ms). By comparing the VOT values in Table 2, there are overlapping of VOT values for both SE and TE speaker groups. For examples, the mean VOT values for unaspirated /p/ in SE was 26.56 ms (S.D. = 19.44 ms); and aspirated /p^h/ was

75.61 ms (S.D. = 52.20 ms). Thus, the VOT values ranged from 24.41 ms to 46.00 ms were confusing to listener's perception. Similar overlapping was seen for all aspirated stops produced by alaryngeal speakers. These overlappings lead to the low percent of identification. This finding was consistent with the data reported by Gandour et al. (1987). In addition, in a Cantonese study investigating the use of acoustic cues to perception of aspiration and place of articulation, Tsui and Ciocca (2000) stated that, comparing with the presence of aspiration noise, VOT was a weak cue to the identification of aspiration. Therefore, though SE and TE speakers were able to produce aspirated stops with longer VOT values, other acoustic cues, such as aspiration noise might be lost. Future research on the presence of aspiration noise in aspirated stops produced by Cantonese SE and TE speakers would provide more information on aspirated-unaspirated confusion in stops produced by those speakers.

Place of articulation. Christensen et al. (1978) compared VOT values in NL and SE speakers in production of English stops. They reported that, for both NL and SE speakers, VOT values increased from labial to velar stops, this was observed in both voiceless and voiced stops. The present data showed similar findings that voiceless unaspirated velar stops in NL and SE speakers were associated with VOT values significantly longer than alveolar and bilabial stops.

Though significant differences were not demonstrated among the three places of articulation for aspirated stops produced by NL and SE speakers, the result was comparable to the Cantonese data reported by Clumeck et al (1981). (77.62 ms, 84.09 ms and 91.89 ms vs. 74.22 ms, 83.22 ms and 90.75 ms). As seen in Table 4, and Figures 3 and 4, similar to data from Thai and Mandarin alaryngeal speakers reported by Gandour et al. (1987) and Liu et al. (in press), Cantonese velar stops were also associated with longer VOT than alveolar and bilabial stops. Though significant differences were not demonstrated for voiceless aspirated and unaspirated stops produced by the TE speakers, the pattern of VOT values increased from labial stops to velar stops was still apparent (see Figure 5).

It is not surprising that voiceless unaspirated velar stops /k/ were associated with significantly longer VOT values than alveolar and bilabial stop in NL speakers. According to the previous discussion (Stevens, 1998), the voicing source will be set into vibration once PTP is reached. The smaller cavity behind the oral occlusion associated with a velar stop yields a rapid build-up of higher intraoral pressure when compared with bilabial and alveolar stops (Hardcastle, 1973, Maddieson, 1997). Therefore, a longer time would be needed to allow air pressure to fall to a sufficient level for onset of vibration. In addition, the articulator associated with the production of velar stops is the tongue dorsum. Different articulators have different speed of movement (Hardcastle, 1973; Kuehn & Moll, 1976). Since the lips and tongues tip have a smaller mass than the tongue dorsum and rotational movement of the jaw opening affected movement of lips and tongue tips more than the back of the tongue, the release of bilabial closure appears to be faster than that of a velar closure. Therefore, air escapes at a slower rate after the release of velar stops, resulting in a longer time to dissipate the intraoral pressure for critical transglottal pressure difference to attain onset of vibration at the voicing source. Furthermore, the contact area between the tongue dorsum and the soft palate is larger than that between the two lips. Thus, Bernoulli force is greater in this narrow constriction to pull the articulators together. The rate of increase in velocity of airflow is slow, making a slower rate in the drop of intraoral pressure (Stevens, 1999). All of these may explain the longer VOT values observed in a velar stop when compared with a alveolar or bilabial stop.

The present data showed that, voiceless aspirated velar stop /k^h/ demonstrated a longer VOT than alveolar and bilabial stops regardless of speaker type. However, the speed of articulator movement is not applied in explaining VOT differences for aspirated stops, as aspiration ends longer after the separation of the articulators (Maddieson, 1997). Therefore, Stevens (1998) suggested that in the production of velar aspirated stops, the build-up of intraoral pressure induces an outward-pulling force on the wall of the vocal tract, together with an increase in the stiffness in the vocal folds, acting as an inhibiting force to the initiation of vibration. Thus, a longer time for phonation onset is resulted. This phenomenon is not observed in the production of bilabial and alveolar aspirated stops as the intraoral pressure of them are smaller, thus adduction forces forms more rapidly and shorter VOT values are resulted.

Limitations of the present study

The inter-subject variation for SE and TE speakers is great as the PE segment varied greatly in the length, mass, passive compliance (Moon & Weinberg, 1987), and configuration (Damste & Lerman, 1969). In this study, only 10 subjects are recruited for each speaker group. Future studies should include larger number of participants to improve external validity. Proficiency criteria were not imposed in the present study. Thus, the relationship between overall speech intelligibility and ability to produce aspirated-unaspirated contrast could not be determined. Additional research should be carried out to include a larger number of participants and to categorize them according to their speaking proficiency within each speaker type. In addition, only CV syllables with the vowel context /a/ embedded in a carrier phrase was used as the speech material. This might not faithfully reflect speakers' performance and listener's perception in spontaneous production. Further studies should involve target syllables in various vowels contexts and linguistic structures such as monosyllabic word, phrase, sentence and conversation to facilitate generalization to spontaneous speech.

Clinical Implications

The present study raises clinicians’ attention to the importance of aspiration distinction in the assessment and treatment of intelligibility of SE and TE speakers of Cantonese. Articulatory accuracy such as aspiration distinction in stop production should also be emphasized on the development of assessment protocols and therapy for alaryngeal speakers. Given that the difficulty in signalling aspirated-unaspirated contrast in Cantonese was prevalent (near or below 50% of identification rate for aspirated stops), a therapy regimen facilitating the production of aspiration stops should be implemented. Results from the present study indicated that, with the use of PE segment as the new voicing source, successful production of aspiration contrast was proven possible as demonstrated by the alaryngeal speakers. Approaches such as “pushing hard” on voiceless aspirated consonants (Christensen & Dwyer, 1990) were proposed to increase the intraoral air pressure, burst and aspiration noise for production of voiceless consonants in order to distinguish them from their voiced cognates. Instrumental monitoring (Connor et al., 1985) were also suggested to be effective in training SE speakers with low intelligibility to vary intraoral pressure, such as producing higher pressure for voiceless stops. These approaches should also be adopted and modified for the application for the production of Cantonese voiceless aspirated stops in alaryngeal speakers to improve their articulatory proficiency.

References

Benki, J. R. (2001). Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics, 29,* 1-22.

Blom, E. D. (2000). Tracheoesophageal voice restoration: Origin-evolution-state-of-the-art. *Folia Phoniatrica Logopaedia, 52,* 14-23.

Blom, E. D., Ronald, C., & Hamaker, M. D. (1996). Tracheoesophageal voice restoration following total laryngectomy. In E. N. Myers & J. Suen (Eds.), *Cancer of the head and neck (Chapter 42).* WB: Saunders Publishers.

Ching, T. Y-C., Williams, R. & van Hasselt, A. (1994). Communication of lexical tones in Cantonese alaryngeal speech. *Journal of Speech and Hearing Research, 37,* 557-571.

Cho, T., Ladefooged, P (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics, 27,* 207-229.

Christensen, J.M., & Dwyer, D. E. (1990). Improving alaryngeal speech intelligibility. *Journal of communication disorder, 23 (6),* 445-451.

Christensen, J. M., Weinberg, B., & Alfonso, P. J. (1978). Productive voice onset time characteristics of esophageal speech. *Journal of Speech and Hearing Research, 21(1),* 56-62.

Clumeck, H., Barton, D., Macken, M.A., & Huntington, D. A. (1981). The aspiration contrast in Cantonese word-initial stops: Data from children and adults. *Journal of Chinese Linguistics,* 9, 210-224.

Connor, N. P., Hamlet, S. L., & Joyce, J.C. (1985). Acoustic and physiologic correlates of the voicing distinction in esophageal speech. *Journal of Speech and Hearing Disorders, 50 (4),* 378-384.

Damste, P.H., & Lerman, J. W. (1969). Configuration of the neoglottis: an X-ray study. *Folia Phoniatrica, 21:*247-258

Dworkin, J. P., Meleca, R. J., Simpson, M. L., Zormeier, M., Garfie,d I., & Mathog, R. H. (1999). Vibratory characteristics of the pharyngoesophageal segment in total laryngectomees. *Journal of Medical Speech-Language Pathology, 7,* 1-18.

Francis, A. L., Ciocca V., & Yu, J. M-C. (2003). Accuracy and variability of acoustic measures of voicing onset. *Journal of the Acoustic Society of America, 113,* 1025-1032.

Fisher, K. V., & Swank, P. R. (1997). Estimating phonation threshold pressure. *Journal of Speech, Language, and Hearing Research, 40,* 1122-1129.

Gandour, J. Weinberg, B., Petty, S. H., & Dardarananda, R. (1987). Voice onset time in Thai alaryngeal speech. *Journal of Speech and Hearing Disorders, 52(3),* 288-294.

Hardcastle W. J. (1973). Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics, 1,* 263-271.

Hashimoto, O. Y. (1972). *Studies in Yue dialexts 1: Phonology of Cantonese.* Cambridge: University Press.

Hirose, H. (1996). Voicing distinction in esophageal speech. *Acta Otolaryngologica Supplement, 524,* 56-63.

Hyman, M. (1955). An experimental study of artificial larynx and esophageal speech. *Journal of Speech and Hearing Disorders, 20,* 291-299.

Jongmans, P., Hilgers, F. J. M., Pols, L. C. W., & van As-Brooks, C. J. (2006). The intelligibility of trachesoesophageal speech, with an emphasis on the voiced-voiceless distinction. *Logopedics Phoniatrics Vocology, 31,* 172-181

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research, 18(4),* 686-706.

Kuehn D.P. & Moll, K. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of phonetics, 4,* 303-320.

Lam, K-Y, P., Ho, W-K., Ho, C-W, A., Ng, W-M. R., Yuen, P-W. A., & Wei W. I. (2005). Long-term performance of indwelling tracheoesophageal speaking valves in Chinese patients undergoing laryngectomy. *Archives Otolaryngoloy Head Neck Surgery, 131,* 954-958

Lewin, J. S. (2004). Advances in alaryngeal communication and the art of tracheoesophageal (TE) voice restoration. *The ASHA leader, 20,* 6-21

Lisker, L. (1975). Is it VOT or first formant detector? *Journal of the Acoustical Society of America, 57,* 1547-1551.

Lisker, L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word,* 20: 384-422.

Lisker, L. & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and speech, 10,* 1-28.

Liu, H., Ng, M. L., Wan, M., Wang, S. & Zhang, Y. (in press). Effects of place of articulation and aspiration on voice onset time in Mandarin esophageal speech. *Folia Phoniatrica et Logopaedica.*

Lundström, E. & Hammarberg, B. (2004). High-speed imaging of the voicing source in laryngectomees during production of voiced-voiceless distinctions for stop consonants. Logopedics *Phoniatrics Vcologica, 29,* 31-40.

Maddieson, I. (1997). Phonetic universals. In J., Laver, W. J., Hardcastle (Eds.). *The Handbook of Phonetic Sciences.* (pp. 619-639) Oxford: Blackwell.

Matta, S., Galli, I., & DiRienzo, L. (2001). Aerodynamic findings in esophageal voice. *Archives Otolaryngology Head Neck Surgery, 127,* 700-704.

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

Most, T., Tobin, Y., & Mimran, R. C. (2000). Acoustic and perceptual characteristics of esophageal and tracheoesophageal speech production. *Journal of Communication Disorders, 33,* 165-181.

Nichols, A. (1976). Confusions in recognizing phonemes spoken by esophageal speakers: in initial consonants and clusters. *Journal of Communication Disorders, 9*, 27-41.

Ng, M. L., Gilbert, H. R., Lerman, J. W. (2001). Fundamental frequency, intensity, and vowel duration characteristics related to perception of Cantonese alaryngeal speech. *Folia Phoniatrica et Logopaedica, 53*, 36-47.

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

Ng, M. L., Lerman, J. W., & Gilbert, H. R. (1998). Perceptions of Tonal changes in normal laryngeal, esophageal, and artificial laryngeal male Cantonese speakers. *Folia Phoniatrica et Logopaedic, 50,* 64-70.

Robbins, J., Christensen, J. & Kempster. G. (1986). Characteristics of speech production after tracheoesophageal puncture: voice onset time and vowel duration. *Journal of Speech and Hearing Research, 29(4),* 499-504.

Salmon, S. J. (1994). Methods of air intake and associated problems. In R. L. Keith & F. K. Darley (Eds.), *Laryngectomee rehabilitation* (3rd Ed.). Austin, Texas: PRO-ED.

Saito, M, Kinishi, M. & Amatsu, M. (2000). Acoustic analyses clarify voiced-voiceless distinction in tracheoesophageal speech. *Acta Otolaryngologica*, *120,* 771-777

Searl, J. P. (2002). Magnitude and variability of oral pressure in tracheoesophageal speech, *54:* 312-328. Retrieved October 30, 2006, from ProQuest Medical Library.

Searl, J. P., & Carpenter, M. A. (2002). Acoustic cues to the voicing feature in tracheoesophageal speech. *Journal of Speech, Language and Hearing Research, 45,* 282-294

Searl, J. P., Carpenter, M. A, & Banta, C. L. (2001). Intelligibility of stops and fricatives in tracheoesophageal speech. *Journal of Communication Disorders, 34,* 305-321

Shames, G. H., Font, J.M., & Matthews, J. (1963). Factors related to speech proficiency of the laryngectomized. *Journal of Speech and Hearing Disorders, 28,* 273-278.

Seikel, J. A, King, D. W., & Drumright, D. G. (2000). *Anatomy and physiology for speech, language and hearing.* San Diego: Singular Publishing Group

Stevens, K. N. (1998). *Acoustic Phonetics.* Cambridge: MIT Press.

Tsui, I. Y. H., & Ciocca, V. (2000). Perception of aspiration and place of articulation of Cantonese initial stops by normal and sensorineural hearing-impaired listeners. *International Journal of Language and Communication Disorders, 35 (4),* 507-525.

Williams, S. E. & Watson, J. B. (1987). Speaking proficiency variations according to method of alaryngeal voicing. *Laryngoscope, 97,* 737-740

Yiu, M-L. E. (1994). Speech intelligibility in tone language (Chinese) laryngectomy speakers. *European Journal of Disorder of Communication, 29,* 339-347.

Yorkston, K., & Beukelman, D. (1980). A clinically judged technique for quantifying dysarthric speech based on single-word intelligibility. *Journal of Communication Disorders, 13,* 15-31.

Zlatin, M. A. (1974). Voicing contrast: perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America. 56,* 981-994