Title	Fundamental frequency characteristics of esophageal and tracheoesophageal speech of Cantonese during speech and non-speech tasks
Author(s)	Mak, Tsz-may; 麥芷薇
Citation	
Issued Date	2011
URL	http://hdl.handle.net/10722/192898
Rights	

Fundamental frequency characteristics of esophageal and tracheoesophageal speech of Cantonese during speech and non-speech tasks

Mak Tsz May, May

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2011.

Abstract

The study investigated the fundamental frequency (F0) from standard esophageal (SE) and tracheoesophageal (TE) speakers of Cantonese in speech and non-speech tasks, and compared the result with normal laryngeal (NL) speakers when they share different speech and air reservoir mechanism. 10 speakers in each group performed spontaneous speech and standard passage reading in speech tasks, and pitch scaling in non-speech tasks. PRAAT was used in F0 extraction. 1 SE data was excluded due to failure of F0 extraction. Results showed all speaker groups produced significantly higher mean F0 in pitch scaling than spontaneous speech task. Significant differences between SE, TE and NL speakers were only observed in pitch scaling task. Result suggested SE and TE speakers with good speech proficiency could produce a F0 higher than habitual speaking F0, which suggested prosody training in future vocal rehabilitation.

Introduction

Many previous research studies have been focused on the acoustic characteristics of normal and pathological voices. Acoustic analysis is commonly done in combination of perceptual assessment of voice, as acoustical variables are able to provide an objective description of voice. Through acoustical analysis, voice production is interpreted in forms of mechanical waves that carry energy. Physically, voice production involves energy generation and propagation. According to the source-filter theory, the glottal sound produced by the vibration of vocal folds (source) is modified by the resonance effect of vocal tract (filter). During this resonance process, sound energy is being increased for some frequencies and decreased for other frequencies (Kuttruff, 2007). Human speech sound is a complex signal which carries energy at all frequencies known as the harmonics. The harmonics can be obtained through Fourier transformation (Kuttruff, 2007), and fundamental frequency (F0) corresponds to the lowest frequency among these harmonics. Physiologically, voice production is generated by pulmonary air that stored and expelled from the lungs (Sataloff, 2006). Air passes through the trachea and reaches the vocal folds in the larynx. The vocal folds adduct and close the glottis. The airflow thus set the vocal folds into vibration. According to Myoelastic-Aerodynamic Theory of Voice Production, this vocal folds vibration is the result of a cyclic interaction of forces from airflow and laryngeal muscles (van den Berg, 1958). Vibratory cycles begin when continued positive subglottal pressure blows apart the vocal folds. Cycles cease when the negative intraglottal surface pressure that built from the nature elasticity of vocal folds overcome reduced positive pressure and close the glottis (van den Berg, 1958). F0 can be altered by changing the tension of vocal folds and the associated laryngeal muscles (Sataloff, 2006). The tenser are the vocal folds, the higher is the F0. As a result, an elevated F0 is perceived as a raise in pitch, and vice versa. According to the assumption of the source-filter theory, F0 of human voice is independent of the vocal tract resonance. Such independence relationship between the source and filter indicates that failure of one system does not affect the other.

Total laryngectomy is a surgical procedure of removing a pathological larynx. It is a common surgery for patients with late stage laryngeal cancer. After the procedure, patients need to create a permanent tracheostoma in the neck to connect the trachea to the outside air for respiration. They also suffer the loss of phonation due to the removal of phonatory apparatus (Elmiyeh, Dwivedi, Jallali, Chisholm, Kazi, Clarke & Rhys-Evans, 2010). Laryngectomized patients therefore need to learn to use an alternative phonation method in order to regain verbal communication. This in fact is an important part of post-laryngectomy rehabilitation for laryngeal cancer survivors. Currently, four types of alaryngeal phonation are available. They are the standard esophageal (SE), tracheoesophageal (TE), electrolaryngeal (EL) and pneumatic artificial (PA) speech (Ng, Kwok & Chow, 1997). Regarding sound source, both SE and TE speech are considered internal as they make use of pharyngoesophageal (PE) segment, known as the neoglottis, as a new sound source. The neoglottis is set into vibration during SE and TE phonation. Sound energy generated by the vibrating neoglottis is then shaped by supralaryngeal vocal tract. Apparently, both SE and TE speech share a similar phonatory device. The only difference between the two types of alaryngeal phonations lies in the mechanism of air reservoir. For SE speakers, air is inhaled or injected and stored in the upper esophagus, while TE speakers make use of pulmonary air from the lungs for phonation (Diedrich, 1999). After an occlusion of tracheostoma, the pulmonary air is directed to TE speakers' esophagus through a one-way valve that situated in a fistula between the trachea and esophagus (Diedrich, 1999).

Because of the use of PE segment as a new sound source, the vocal characteristics associated with SE and TE speech have received particular interests from many researchers (e.g., Blood, 1984; Filter & Hyman, 1975; Ng et al., 1997; Max, Steurs & de Bruyn, 1996;

Robbins, 1984; Robbins, Fisher, Blom & Singer, 1984). Generally speaking, results of previous studies indicated that, with the use of PE segment as the neoglottis, both SE and TE speakers of English exhibited a significantly lower average F0 than NL speakers during speech. F0 values of English-speaking SE speakers were found to be lower than TE speakers by about 25 Hz (Blood, 1984). SE speakers approximately produced an average of 65 Hz and TE speakers produced 88 Hz, which was approximately one octave lower than that of laryngeal speakers (e.g., Blood, 1984; Robbins, 1984; Robbins et al., 1984). For Polish alaryngeal speakers, acoustical parameters including maximum phonation time (MPT) showed that TE speech was more similar to laryngeal speech when compared with SE speech (Olszański, Gieroba, Warchoł, Morshed & Gołabek, 2004).

However, previous studies reported that Mandarin and Cantonese SE speakers spoke with significantly higher average F0 than English SE speakers (Liu, Wan, Wang, Wang & Lu, 2005; Ng, Gilbert & Lerman, 2001). Cantonese SE speakers exhibited an average F0 of 155.2 Hz in reading (Ng et al., 2001), and Mandarin SE speakers had an average of 84.35 Hz (Liu et al., 2005). According to Ng et al. (2001), this difference in average speaking F0 between English SE speakers and Chinese SE speakers might be attributable to the language difference: English is a non-tonal language while both Mandarin and Cantonese are tonal. A tonal language is characterized by variation of F0 at the syllable level to convey different lexical meaning. For example, there are six contrastive tones in Cantonese. The syllable /ji/ can mean ‘clothes’, ‘chair’, ‘opinion’, ‘to move’, ‘ear’, or ‘easy’, depending on which tone is produced. Therefore, F0 control is more important to speakers of a tone language than those of a non-tonal language for effective communication with others. Due to this unique language requirements, SE and TE speakers of a tonal language such as Cantonese and Mandarin needed to acquire the ability in varying F0 to convey the correct meaning (Liu et al., 2005; Ng et al., 2001). In other words, they needed to learn to better control the tension of PE segment during speech production. To achieve this, SE and TE speakers may contract the neck muscles to tense the PE segment (Ng et al, 2001). This tensed PE segment can then be set into a faster vibration, resulted in higher F0. This may explain why Cantonese SE speakers produced F0 that appeared to be higher than English SE speakers.

F0 control is an important part of prosody as it corresponds to the pitch listeners perceived. In verbal communication, prosody provides listeners important cues of the intended messages (Schindler, Canale, Cavalot, Albera, Capaccio, Ottaviani & Schindler, 2005). According to Ferreira, Anes, and Horine (1996), prosody can be more effective in conveying semantic information than syntactic structure. Therefore, studying the acoustical variables of F0 control of SE and TE speakers can shed lights to the understanding of their competence in pragmatics. Such information can also be used in deciding therapeutic approach. According to Schindler et al. (2005), TE speakers in English were able to produce higher pitch than comfortable pitch with significant difference. Also, Max et al. (1996) reported mean frequency range of Dutch SE and TE speakers as 56.3 Hz and 83.8 Hz respectively when compared their lowest and highest pitch. However, most studies of F0 control in SE and TE speakers in Cantonese focused on speech tasks only, such as reading a standard passage or model sentences. This limited the full picture of the performance of voluntary F0 control by both alaryngeal speakers in Cantonese. Since both speakers had used PE segment as the new sound source, it is not known how their F0 control in non-speech tasks differed from normal laryngeal speakers. Since SE and TE speakers used different air reservoir mechanism, it is not known if their F0 control differed from each other. Moreover, if the higher F0 associated with Cantonese and Mandarin alaryngeal speakers is originated from the fact that both Cantonese and Mandarin are tonal, they may produce comparable F0 values with non-tonal speakers in non-speech tasks.

The present project attempted to answer the questions of how speech and non-speech tasks affect the F0 characteristics of SE and TE speech of Cantonese, and how SE, TE and NL speakers differ from each other in the tasks when they share different speech and air reservoir mechanism. Based on the above discussion, it can be hypothesized that the highest F0 taken from non-speech tasks of SE and TE speakers was similar to speaking F0 taken from speech tasks. It was because their better control on PE segment that suggested by previous studies only limited to tone production and not in non-speech tasks such as singing. It can also be hypothesized that the F0 characteristics associated with speech and non-speech tasks produced by Cantonese SE and TE speakers are different from normal speakers if they use different sound source (vocal folds versus neoglottis). Since significantly different volume of air reservoir exists between SE and TE speakers, it is expected that SE speakers, who made use of positive intraesophageal pressure to move PE segment, had higher mean F0 values in speech tasks and higher mean of the highest F0 in pitch scaling in non-speech tasks than TE speakers.

Method

Participants

Thirty adult male native Cantonese speakers (10 SE, 10 TE, and 10 laryngeal speakers) participated in the present study. The normal laryngeal speakers (NL) were recruited to serve as controls. The alaryngeal speakers were superior SE and TE speakers selected from the New Voice Club of Hong Kong by a practicing speech therapist. All speakers were matched with age, ranged from 48 to 80 years. All of them had no history of speech-language and/or hearing problems except those associated with laryngectomy for alaryngeal speakers. All participants were literate and were able to read the speech materials used in the study.

Speech materials

Speech task. Spontaneous speech task and a passage reading task were performed by all speakers. To help elicit the spontaneous speech samples, the participants were asked to engage in casual conversation with the experimenter. They were asked probe questions such as "*How did you get here today? What did you have for dinner last night? Why did you choose SE/TE as your voice rehabilitation method?*", etc. Also, the Cantonese standard passage 'The North wind and the Sun' was used. See Appendix for the passage. It was printed on a white paper sheet with black words and medium font size to achieve clear vision on each Chinese characters.

Instrumentation and recording procedure

All speech recordings took place in a sound attenuated room. The conversation lasted for two minutes. After the conversation, passage reading was followed. The passage was presented to the participants. They were instructed to read aloud the passage with a comfortable loudness level, pitch level and speech rate. Practice period was provided in order to familiarize the participants with the recording materials and environment. When needed, modeling was provided. The speech samples were recorded using a high quality microphone (SM58, Shure) via a preamplification unit (PreMobile USB, M-Audio). A microphone-to-mouth distance of approximately 10 cm was maintained throughout the recording. The signals were digitized at a sampling rate of 20 kHz and quantization rate of 16 bits/sample using PRAAT (Boersma & Weenink, 2005; Wood, 2005). The digitized signals were stored in a computer for later acoustic analyses.

Data analysis

All digitized signals were specifically coded and encrypted. During average F0 measurements from reading passage, the third sentence of the reading passage was used to represent the reading F0 of the participants. PRAAT (Boersma & Weenink, 2005; Wood, 2005) was used to display waveform and F0 analysis. From the waveform display, the entire third sentence was marked by the experimenter, and aperiodic portion was excluded. Similar analysis was processed for measuring average F0 values in spontaneous speech using first utterance of the speech after one minute.

Non-speech task. To obtain non-speech sound samples, pitch scaling was performed by the participants. During the recording of non-speech activities, the participants were instructed to sing from a comfortable pitch to the highest pitch that they can reach for three times. Back low vowel ‘a’ was used as the key for pitch scaling. The recording set up and signal digitization was similar to the speech activities. When needed, explanation and modeling were provided. Regarding the F0 measurements of pitch scaling, the highest F0 values were obtained using PRAAT (Boersma & Weenink, 2005; Wood, 2005). Some of the F0 measurements from SE and TE voice samples in pitch scaling were not detected by PRAAT (Boersma & Weenink, 2005; Wood, 2005) when the vocalic portions were too short and unstably produced. From the waveform display, the short consecutive periodic portion of the vocalic part needed to be identified and marked manually by the experimenter. A pitch calculation routine was used to extract the F0 values from the marked segment.

Reliability measurement.

Reliability measurement of F0 value calculation was performed. A total of 30 speech samples (10 samples from each task) from spontaneous speech, the passage reading, and pitch scaling were randomly selected. Pearson product-moment correlation coefficient and absolute error values were calculated between the first and second F0 values to indicate intrajudge reliability.

Statistical Analysis

To assess the possible difference in measured F0 values among the three speaker groups and the different production tasks, a 3 (speaker group) x 3 (production tasks) mixed design Analysis of Variance (ANOVA) test was carried out, with speaker group as the between-subjects variable and production task as the within-subjects variable.

Results

Periodic portions of sound waves were automatically marked by PRAAT (Boersma & Weenink, 2005; Wood, 2005). Figures 1, 2 and 3 show the waveform associated with the spontaneous speech task produced by a typical normal speaker, a typical TE speaker and a typical SE speaker respectively. However, the automatic pitch extraction feature of PRAAT may become unreliable at times. For example, PRAAT failed to extract F0 information from the speech samples recorded from a SE speaker during pitch scaling due to his unstable and hoarse voice quality. In this case, manual identification of periods of the entire speech sample was followed. However, no periodic portion was identified from this SE speaker. As no F0 information could be obtained, his sample was excluded from the data analysis. For this SE speaker, the speech signal of the spontaneous speech task is presented in Figure 4. The corresponding signals for passage reading and pitch scaling are shown in Figures 5 and 6. In the end, a total of 10 samples from NL speakers, 9 from SE speakers and 10 from TE speakers were included in statistical analysis.

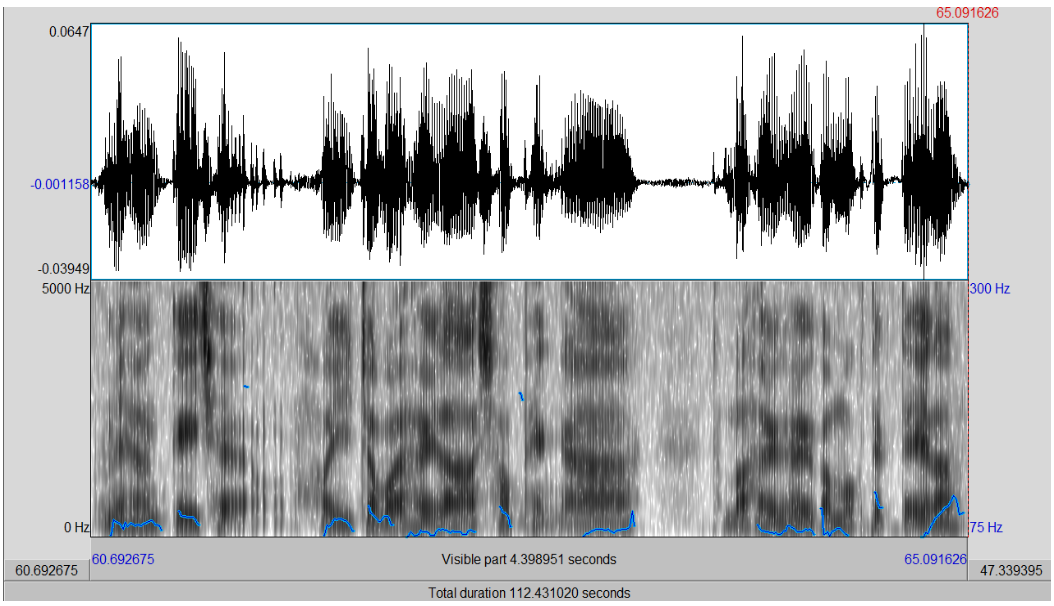

Figure 1. Example waveform and wide-band spectrogram of a sentence extracted from spontaneous speech task produced by a typical normal laryngeal speaker.

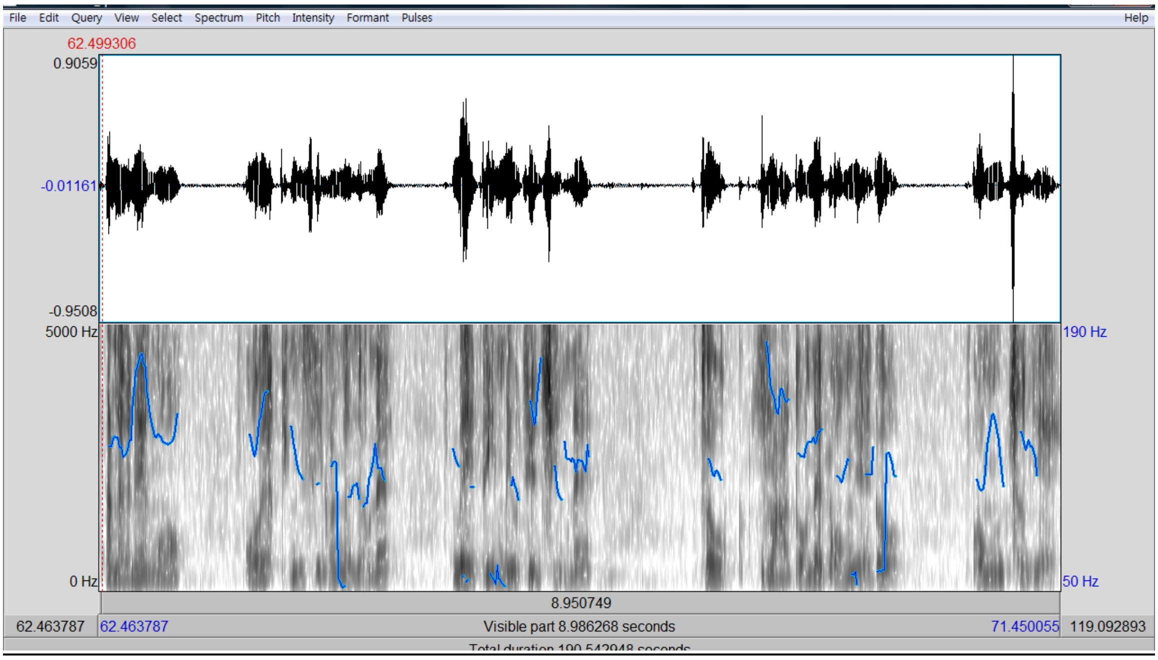

Figure 2. Example waveform and wide-band spectrogram of a sentence extracted from spontaneous speech task produced by a typical tracheoesophageal speaker.

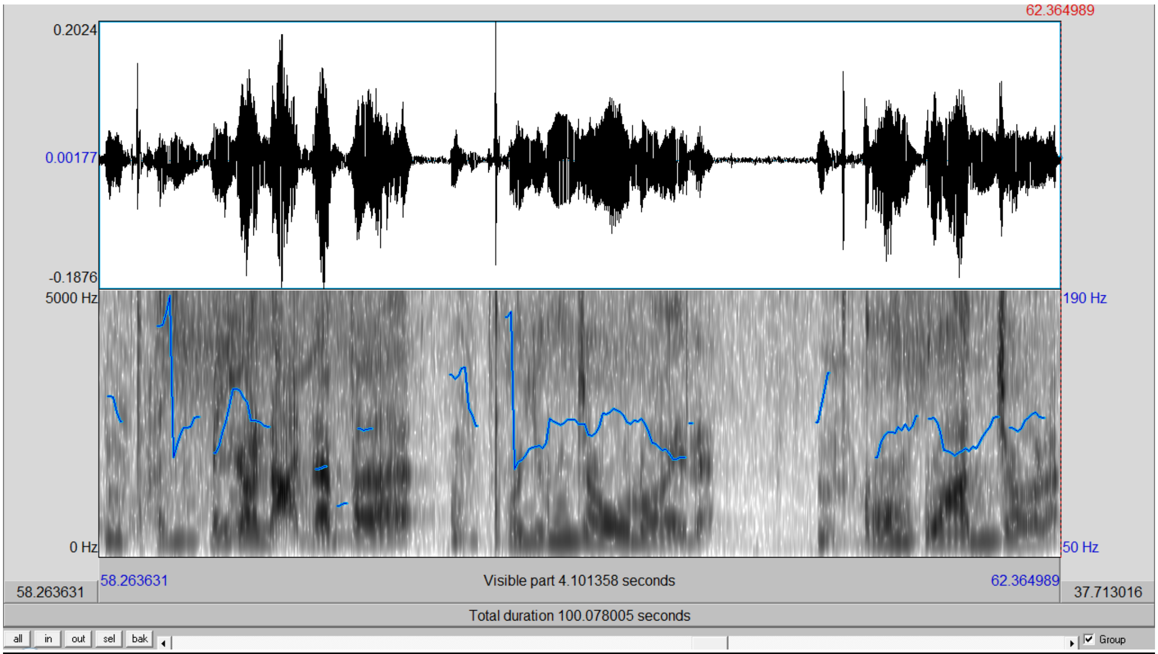

Figure 3. Example waveform and wide-band spectrogram of a sentence extracted from spontaneous speech task produced by a typical standard esophageal speaker.

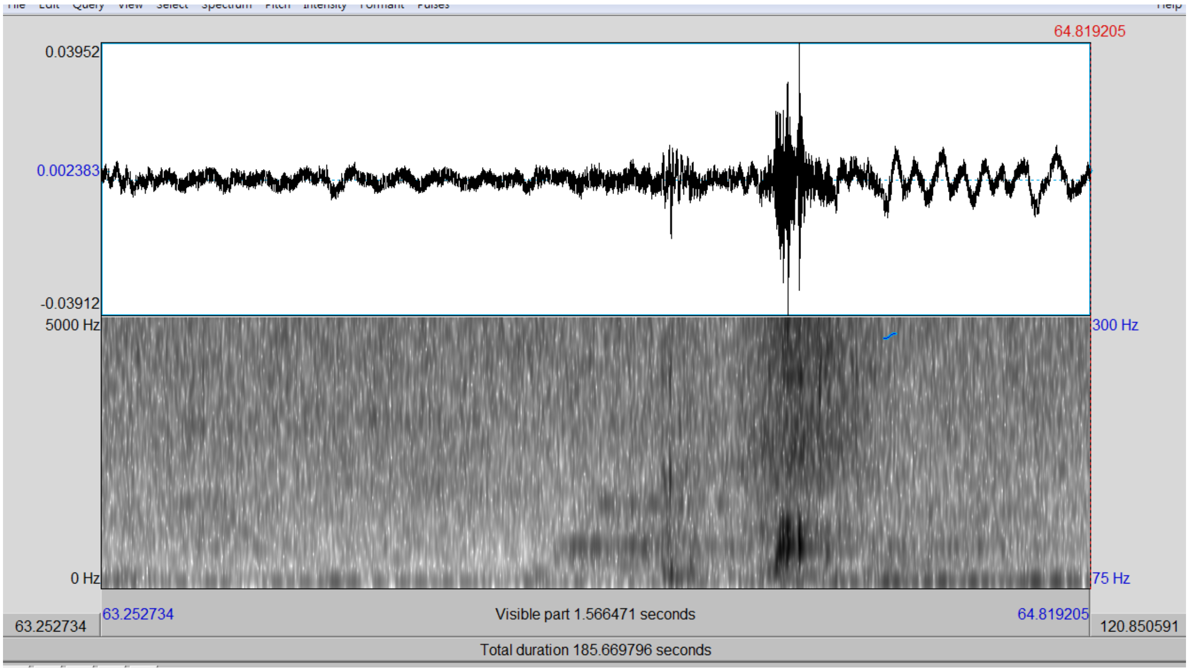

Figure 4. Example waveform and wide-band spectrogram of a sentence extracted from spontaneous speech task produced by the excluded standard esophageal speaker.

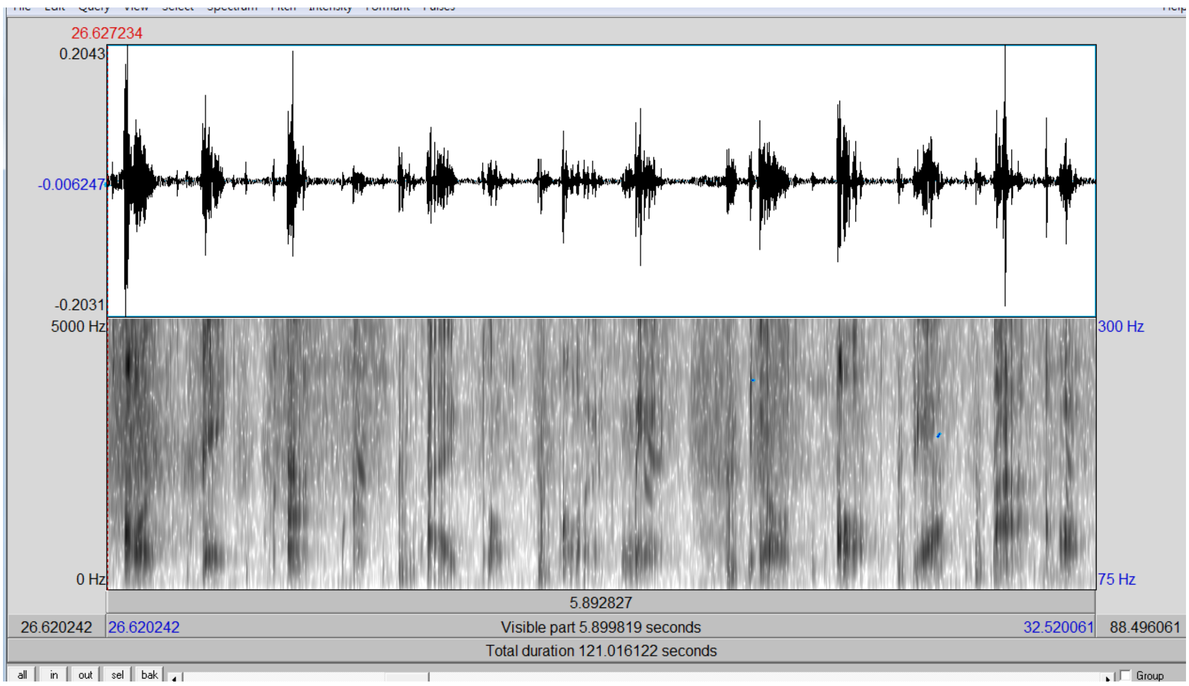

Figure 5. Example waveform and wide-band spectrogram of the third sentence extracted from passage reading task produced by the excluded standard esophageal speaker.

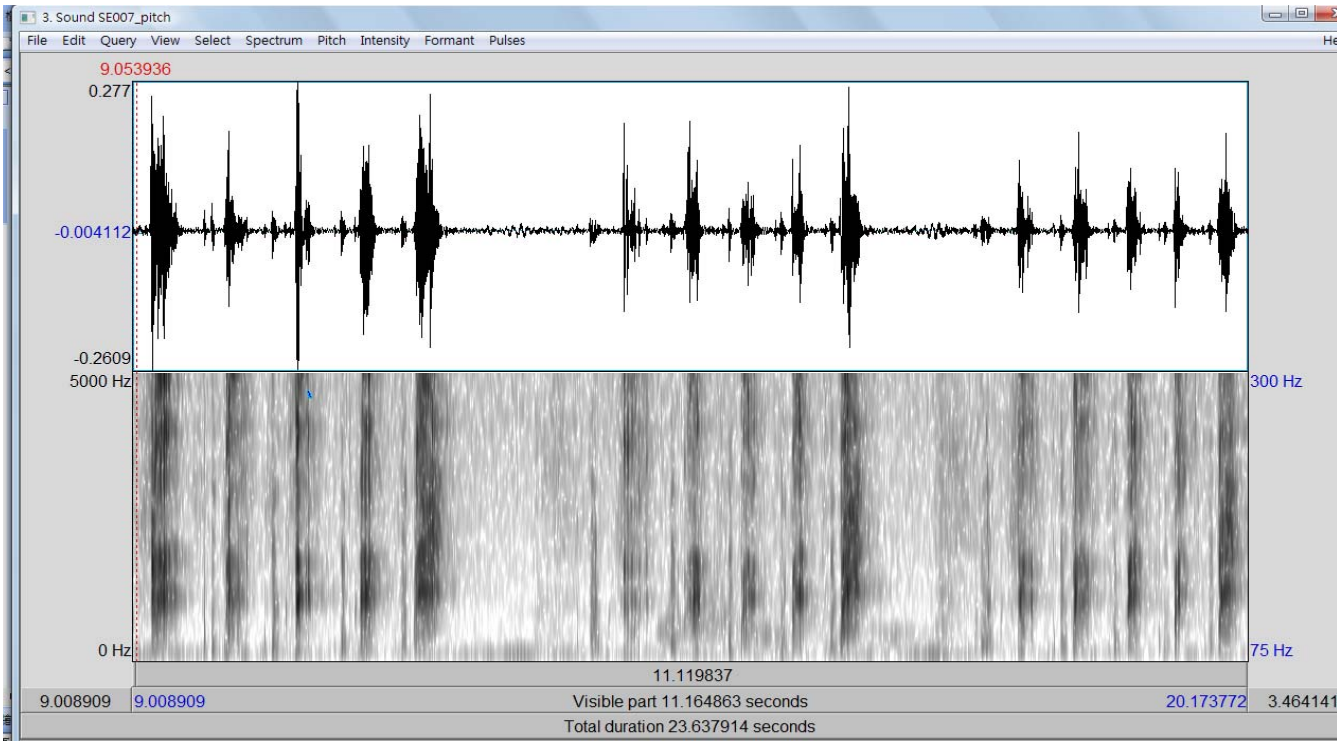

Figure 6. Example waveform and wide-band spectrogram associated with pitch scaling by the excluded standard esophageal speaker.

Reliability of Measurement

Intrajudge reliability was used to represent the accuracy of the F0 extraction in the present study. Average absolute error values and Pearson product-moment correlation coefficients of F0 values between the first and second measurements during spontaneous speech, reading passage and pitch scaling tasks are shown in Table 1. The specific correlation coefficients are: $r(10) = .967$, $p < .001$ for spontaneous speech task, $r(10) = .828$, $p = .003$ for passage reading task, and $r(10) = .998$, $p < .001$ for pitch scaling task. These results consistently indicated F0 measurements are consistent and reliable.

Table 1.

Intrajudge reliability for F0 measurement in spontaneous speech, passage reading and pitch scaling tasks.

Tasks	Average Absolute Error Values	Pearson Product-Moment Correlation
Spontaneous Speech Task	3.99 Hz	.967 ($p < .001$)
Passage Reading Task	19.40 Hz	.828 ($p = .003$)
Pitch Scaling Task	2.70 Hz	.998 ($p < .001$)

Mean F0 values in spontaneous speech, passage reading, and pitch scaling tasks across three speaker groups.

The average F0 and standard deviation values associated with the three production tasks produced by the NL, SE, and TE speakers are presented in Table 2. According to Table 2, generally speaking, the average F0 values among the three tasks were different. NL, SE and TE speakers consistently exhibited higher average F0 values in pitch scaling (non-speech) tasks than spontaneous speech and passage reading tasks. Differences in average F0 values also existed among speaker groups. SE speakers produced higher average F0 values in all the tasks than TE and NL speakers except pitch scaling. SE speakers exhibited the highest average F0 values in spontaneous speech (124 Hz for SE, 111 Hz for NL, and 109 Hz for TE) and passage reading (136 Hz for SE, 114 Hz for NL and 109 Hz for TE). However, in pitch scaling task, NL speakers exhibited the highest average F0 values than SE, and followed by TE speakers (212 Hz for SE, 335 Hz for NL, and 203 Hz for TE).

Table 2.

Average F0 Values (in Hz) and Standard Deviation (in Hz) of NL, SE and TE speakers of Cantonese in Spontaneous Speech, reading passage and pitch scaling

Tasks	NL Mean (S.D.)	SE Mean (S.D.)	TE Mean (S.D.)
Spontaneous Speech	111 (22.0)	124 (45.8)	109 (22.1)
Reading Passage	114 (16.5)	136 (37.6)	109 (25.3)
Pitch Scaling	335 (101.6)	212 (119.4)	203 (103.4)

Regarding standard deviation of F0 values across different production tasks, SE speech demonstrated the highest SD of F0 values than NL and TE speakers. This indicated that SE speech is associated with the most variability in F0 values than NL and TE speakers during different production tasks.

Results of a 3 (speaker groups) x 3 (tasks) repeated-measures ANOVA indicated a significant interaction between speaker group and production task [$F(4, 52) = 6.31$, $p < .001$]. A one-way repeated-measures ANOVA was then performed for each group to determine if there were significant differences among production tasks. A one-way ANOVA was also performed to assess for significant F0 values among different production tasks.

Among different speech tasks

The ANOVA results revealed that, for NL speakers, significant main effect was found [$F(2, 18) = 49.5$, $p < .001$]. Post-hoc tests using Least Significant Difference (LSD) showed that the average F0 from the spontaneous speech task (M = 111, SD = 22.0) was significantly lower than that from pitch scaling task (M = 335, SD = 102) ($p < .001$), and that from passage reading (M = 114, SD = 16.5) was significantly lower than that from pitch scaling task (M = 335, SD = 102) ($p < .001$).

For TE speaker group, ANOVA results indicated significant main effect [$F(2, 18) = 9.57$, $p = .001$]. LSD post-hoc tests showed that the average F0 associated with spontaneous speech task (M = 109, SD = 22.1) was significantly lower than that with pitch scaling task (M = 203, SD = 103) ($p = .015$), and that with passage reading task (M = 109, SD = 25.3) was significantly lower than that with pitch scaling task (M = 203, SD = 103) ($p = .011$). This does not seem to support the present hypothesis that TE speakers cannot produce significant difference between speech and non-speech tasks.

For SE speaker group, significant main effect was found [$F(2, 16) = 6.77$, $p = .007$]. The average F0 associated with spontaneous speech task (M = 124, SD = 45.8) was significantly lower than that with pitch scaling task (M = 212, SD = 119) ($p = .017$). This does not support the present hypothesis that SE speakers cannot produce significant difference between speech and non-speech tasks. However, the pattern of how tasks differ from each other in SE was different from NL and TE speakers.

Across three production tasks

In addition to examining the difference between different speaker groups, between-group comparison was also made across the three production tasks. For both spontaneous speech and passage reading tasks, no significant main effect was found (for spontaneous speech: $F(2, 26) = .570$, $p = .572$; for passage reading task: $F(2, 26) = 2.57$, $p = .096$). This indicates that the three speaker groups exhibited comparable average F0 during spontaneous speech and passage reading. This does not appear to support the hypothesis that TE and SE speakers would produce different F0 than NL speakers.

For pitch scaling task, significant main effect was found [$F(2, 26) = 4.62$, $p = .019$]. The subsequent post-hoc multiple comparisons showed that the average F0 associated with NL speech (M = 335, SD = 102) was significantly higher than TE speech (M = 203, SD = 103) ($p = .011$) and SE (M = 212, SD = 119) ($p = .020$). This finding supports the hypothesis that TE and SE speakers would produce significantly different F0 from NL speakers in pitch scaling task.

Discussion

The NL, SE, and TE speakers showed different patterns of F0 values across the three different speech tasks. Two unpredicted outcomes were obtained. The first was that TE and SE speakers could produce significant differences in F0 between speech and non-speech tasks. Another finding was that TE and SE speakers exhibited similar F0 values with NL speakers during spontaneous speech and passage reading tasks. Discussion on performance on the three production tasks for each speaker group, and discussion on the F0 extracted from the three speaker groups for each tasks is followed in an attempt to explain the two outcomes.

Comparison of F0 across NL, SE and TE speakers

According to the present study, NL speakers exhibited comparable average F0 of 111 Hz and 114 Hz in spontaneous speech and passage reading tasks, respectively. The comparable F0 may be related to the fact that both speech tasks required the participants to use their habitual and comfortable pitch level. During the experiment, the participants were instructed to speak into the microphone using the most comfortable loudness and pitch levels. This finding is comparable to what previous studies have reported. For instance, average reading F0 exhibited by Cantonese and Mandarin speakers were reported to be 120.5 Hz and 111.51 Hz, respectively (Liu et al., 2005; Ng et al., 2001). Both studies used a third grade Chinese passage as the speech materials. This indicates that the use of different reading materials and dialectal difference does not affect average reading F0. The finding is also comparable to studies reported from English-speaking population. For example, an average F0 value of 116.65 Hz in passage reading has been reported by Baken (1987). Again, this shows language difference does not appear to affect the average F0 in reading, at least for laryngeal speakers. However, their average F0 in pitch scaling task was significantly higher than spontaneous speech and reading passage tasks, indicating that NL speakers can voluntarily modulate F0 from habitual speaking pitch to a higher pitch level (M = 335 Hz) (for as much as nearly 1.6 octaves higher). According to Sulter, Schutte, and Miller (1995), 335 Hz accounted for about 60% of maximum F0 that exhibited by young Dutch-speaking males. According to Ptacek, Sander, Maloney, and Jackson (1966), the pitch range of geriatric adults reduced when compared to young adults who under 40 years old. So, the present finding of NL's highest pitch was comparable to previous studies on males' vocal capacities. Such a change of F0 from habitual pitch to the highest pitch was contributed by an increased tension on vocal folds and contraction of related laryngeal muscles such as cricothyroid muscles (Sataloff, 2006).

For TE speakers, discrepancy between the present study and the previous literature renders its difficulty to directly compare the present findings to previous studies. In the study, average speaking F0 in spontaneous speech and passage reading tasks were both around 109 Hz, which was comparable to findings based on three European countries that reported an average F0 of 110 Hz for good TE speakers in sustained vowel (van Gogh, Festen, Verdonck-de Leeuw, Parker, Traissac, Cheesman & Mahieu, 2005), and 101.7 Hz in reading first paragraph of the English *Rainbow Passage* as reported by Robbins et al. (1984). However, a study based on Hebrew reported average F0 value 85.48 Hz when reading short and long sentences (Most, Tobin & Mimran, 2000), and 88.3 Hz reported by Blood (1984) in reading the second sentence of *Rainbow Passage*. The difference of about 14 - 25 Hz in the average

F0 values across different studies can be the result of different sampling methods used including the use of different sampling sizes and gender consideration. For example, van Gogh et al. (2005) recruited six female laryngectomees, whereas Most et al. (2000) recruited five TE male speakers. According to Bellandese (2009), male and female TE speakers had significant difference in average F0 values obtained from reading the first paragraph of *Rainbow Passage*. Female TE speakers tended to exhibit higher F0 (Bellandese, 2009). Different sampling methods using the same gender or not, and large or small sample size, may result in discrepant F0 measurements. Another reason for the observed discrepancy could be the different voice qualities of TE speakers recruited in different studies. According to van Gogh et al. (2005), TE speakers produced complex waveforms that usually had high variability, which contained aperiodicity and noise that leaded to unreliable measurements. Different categories of voice quality resulted in significant different fundamental frequency stability and harmonic-to-noise ratio (van Gogh et al., 2005). Therefore, the possible contributing factors by sampling method and voice qualities variation in TE speakers made it difficult to directly compare the present finding to previous findings that investigated non-tonal language speakers on habitual speaking F0.

The average of the highest F0 in pitch scaling task was 203 Hz, about 94 Hz or 0.90 octaves higher than habitual speaking F0. The speakers produced significantly higher F0 in pitch scaling task than spontaneous speech and passage reading. The Cantonese TE speakers in this study could increase F0 not only in speech tasks for tone production, but also in non-speech tasks for high pitches production. This finding was comparable to a study of Dutch-speaking TE speakers that exhibited a mean frequency range of about 83.8 Hz (Max et al., 1996). The present finding also agrees with the functional pitch variation reported for TE speakers by Desschler, Doherty, Reed, and Singer (1999). It also agrees with Schindler et al. (2005) that TE speakers could voluntarily, but also poorly change F0 for an interval of fifth from comfortable pitch with significant difference. It is suggested that good TE Cantonese speakers performed similarly to TE speakers of a non-tonal language. They may apply the aerodynamic-myoelastic event in F0 variation during non-speech task such as pitch scaling task (van den Berg, 1958). When they are asked to increase pitch to their largest extent, they tended to achieve a F0 that higher than habitual speaking F0 (Fujimoto, Kinishi, Mohri & Amatsu, 1994). Intra-tracheal pressure, neoglottal closure resistance and airflow rate may tend to inflate by contracting the inferior constrictor muscle of the pharynx (Fujimoto et al., 1994; Grolman, Eerenstein, Tan, Tange & Schouwenburg, 2007), which yielded a faster vibration of PE segment.

For SE speakers, the average F0 values in spontaneous speech and passage reading tasks were 124 Hz and 136 Hz, respectively. The reading F0 was lower than the mean F0 value 155.2 Hz obtained in a previous study using 10 Cantonese SE speakers (Ng et al., 2001). However, this was still higher than the F0 value reported in non-tonal languages such as English or Hebrew (Blood, 1984; Most et al, 2000; Robbins et al., 1984). For example, 77.1 Hz reported by Robbins et al. (1984), and 70 Hz reported by Blood (1984) for SE speakers. As higher speaking F0 correlated to better speech proficiency (Most et al., 2000), it follows that the Cantonese SE speakers recruited in the study were mainly superior speakers. It should be noted that a relatively large standard deviation of F0 was associated with the SE speaker group (45.8 Hz in spontaneous speech; 37.6 Hz in reading passage; 119.4 Hz in pitch scaling task), indicating that the SE speakers had relatively large variability on the average rate of PE segment vibration during SE phonation. This variability may reflect different duration and frequency of use of SE speech and individual physical strength after laryngectomy (Robbins et al., 1984). As lexical tone production was crucial for effective communication in tonal languages, these good SE speakers may have better learnt a more effective aerodynamic-myoelastic event and vocal tract control such that an increase of tension of PE segment and tracheal pressure for a higher speaking F0 was achieved (Liu et al., 2005; Ng et al., 2001). Results of the present study also indicated higher average of highest F0 in pitch scaling (M = 212 Hz) than spontaneous speech but not in passage reading. But, noted that the difference between passage reading and pitch scaling approached statistical significance (p = .051). The averages of the highest F0 was about 88 Hz or 0.77 octaves higher than the averages of speaking F0 in spontaneous speech, and 78 Hz or 0.64 octaves higher than that in reading passage. The differences in mean F0 between pitch scaling and the two speech tasks were larger than the frequency range of 56.3 Hz reported by Max et al. (1996). This suggested SE Cantonese speakers in the present study can produce higher than habitual speaking F0 by voluntarily increased the rate of vibration of PE segment. It may suggest Cantonese SE speakers had better F0 control than Dutch speakers due to the unique nature of the language. Such an effect on pitch scaling may be generated by similar aerodynamic-myoelastic event in speech tasks, but with a larger effort to further increase the tension of PE segment by contracting the intrinsic and/or extrinsic neck muscles, and increasing pressure under the neoglottis (Liu et al., 2005; Ng et al., 2001; Schutte & Nieboer, 2002; van den Berg, 1958).

Comparison across spontaneous speech, passage reading and pitch scaling tasks

As both speech tasks required the participants to use their comfortable pitch, all speakers exhibited comparable F0 in the two speech tasks. In the spontaneous speech and passage reading tasks, both TE and SE speakers were not significantly different from NL speakers. The finding was inconsistent with previous studies on Cantonese that a significant differences across speaker groups was obtained (Ng et al., 2001), but noted that the mean F0 obtained from SE speakers was still higher than NL speakers and followed by TE speakers, which was consistent with previous literature using Cantonese participants (Ng et al., 2001). The failure to obtain statistical significance in the present study may reflect some good TE and SE speakers were recruited who had practiced TE or SE speech for many years and were able to modulate a stable, efficient PE segment contraction and produced F0 that can achieve a better acceptability and intelligibility (Most et al., 2000). In that case, the average F0 values obtained from TE and SE may both approach to NL speakers. So, statistical significance of between group differences in speech tasks was not observed here. Moreover, tone production was crucial in communication for tonal speakers. According to Liu et al. (2005) and Ng et al. (2000), Cantonese and Mandarin SE speakers may better learn to modulate the PE segment for F0 variation within a syllable to note different lexical tones. Similar theory may apply to Cantonese TE speakers. So, language difference may help to explain why present finding did not consistent with previous literature that reported significant differences among speaker groups in non-tonal speakers such as American English or European languages (Blood, 1984; Most et al., 2000).

The present finding was consistent to previous literatures that TE's speech was closer to normal speakers than SE speech (Blood, 1984; Max et al., 1996; Olszański et al., 2004; Robbins et al., 1984; Schindler et al., 2005), as the mean F0 values of TE speakers were closer to NL than SE speakers in both speech tasks. SE speakers produced higher than normal speaking F0 in both speech tasks. This agreed with higher F0 of SE than TE speakers reported by Olszański et al. (2004). As TE and SE's phonatory system differed by mechanism of air reservoir (Diedrich, 1999), the present finding may support an over-riding effect of aerodynamics to choice of sound source on speaking F0 production. According to Ng (2011), SE speakers produced a higher pressure below the neoglottis than TE speakers. Also, the neoglottal resistance measured was greater in SE speakers than in TE speakers (Ng, 2011). The increased measurements of these parameters from SE speakers accounted for a higher F0 production (Fujimoto et al., 1994; Grolman et al., 2007; Ng, 2011). Such a difference in aerodynamic-myoelastic effect between SE and TE speakers may be contributed by different air reservoir mechanisms (Ng, 2011). As TE speakers used pulmonary air as the driven source, which was the same as normal speakers, the air expulsion was independent from PE segment (Ng, 2011). However, SE speakers used air that stored in the upper esophagus. Air expulsion to neoglottis and PE segment vibration may be generated at the same time, thus increased the pressure and F0 values in speech tasks (Ng, 2011).

For pitch scaling tasks in non-speech tasks, NL speakers produced significantly higher F0 than SE and TE speakers. Both SE and TE speakers produced similar mean F0 of the highest pitch, and comparable differences between mean habitual F0 and mean of the highest F0 exhibited (e.g. 88Hz for SE; 83.8 Hz for TE). This agrees with Max et al. (1996) that SE and TE speakers had no significant differences for frequency range. This also agrees that TE speakers had poor control in voluntary frequency inflation (Schindler et al., 2005). The results may be explained by the elasticity differences between vocal folds and neoglottis. Although NL and TE speakers used lung as air reservoir, their performances in maximum pitch production was different. This suggested that, when extreme F0 was intended during voice production, vocal folds used by NL speakers should be regarded as a better phonatory tool over neoglottis which allowed a steadier, more dynamic interaction with Bernoulli force produced by airflow (van den Berg, 1958). It also allowed a greater extent of change of tension than neoglottis. Also, physical constraint on the pseudophonatory systems after laryngectomy may contribute to reduced frequency range by TE and SE speakers (Robbins et al., 1984).

Clinical value of the findings and limitation of the study

Since the present finding has indicated that good Cantonese TE and SE speakers were able to produce high pitch voluntarily during non-speech tasks, this may suggest prosody training in vocal rehabilitation for these clients. Although they may not be able to produce similar frequency range as normal speakers, prosodic changes can still be made as long as they had significant differences from habitual F0 (Schindler et al., 2005). Prosody was important in marking appropriate pragmatics during verbal communication. Training prosody was functional for these laryngectomees to increase their language competence by providing additional contextual cues of the intended messages to interlocutors (Ferreira, Anes & Horine, 1996). However, the limitation of the present study was that mainly good or excellent TE and SE speakers were recruited. It may not be representative for poor TE and SE speakers who usually exhibited higher harmonic-to-noise ratio, lower F0 values, and slower speech rate (Most et al., 2000, van Gogh et al., 2005). It is suggested by Most et al. (2000) that alaryngeal speakers should be categorized into good and moderate speech proficiency before acoustical analysis or speech rehabilitation. So, further study may investigate the F0 measurements of different speech proficiencies of TE and SE speakers to see if there is valuable result. Also, only fundamental frequency was considered in the acoustical analysis of non-speech tasks. It was not known if there was other change of acoustical parameters such as intensity level, phonation time, number of pauses in phonation that help to illustrate a more specific profile of NL, TE and SE speakers in pitch scaling in non-speech tasks, and whether these unknown parameters interacted with F0 values during the tasks. Further study should consider these acoustic variables in non-speech tasks and match with speech tasks to investigate a full picture of TE and SE speakers' speech mechanism. Last, only male participants were included. The result cannot generalize to female speakers. As female SE and TE speakers exhibited higher speaking F0 (Bellandese, 2009), their performance in non-speech tasks such as pitch scaling should be different from male participants. Whether they will exhibit higher or larger range of F0 is left for further investigation.

References

Baken, R. J. (1987). *Measurement of Speech and Voice*. Boston: College Hill.

Bellandese, M. H. (2009). Fundamental frequency and gender identification in standard esophageal and tracheoesophageal speakers. *Journal of Communication Disorders, 42* (2), 89-99

Blood, G.W. (1984). Fundamental frequency and intensity measurements in laryngeal and alaryngeal speakers. *Journal of Communication Disorders*, *17* (5), 319-324.

Boersma, P., & Weenink, D. (2005). Praat: doing phonetics by computer (Version 4.3.01) [Computer program]. Retrieved from http://www.praat.org/

Desschler, D.G., Doherty, E.D., Reed, C.G., & Singer, M.I. (1999). Effects of sound pressure levels on fundamental frequency in tracheesophageal speakers. *Otolaryngology-Head and Neck Surgery, 121* (1), 23-26.

Diedrich, W. M. (1999). Anatomy and physiology of esophageal speech. In Salmon SJ (ed.): *Alaryngeal Speech Rehabilitation* (2nd e.d.). Austin, TX: Pro-Ed.

Elmiyeh, B. B., Dwivedi, R. C., Jallali, N. N., Chisholm, E. J., Kazi, R. R., Clarke, P. M., & Rhys-Evans, P. H. (2010). Surgical voice restoration after total laryngectomy: An overview. *Indian Journal of Cancer*, *47* (3), 239-247.

Ferreira, F., Anes, M. D., & Horine, M. D. (1996). Exploring the use of prosody during language comprehensive using the auditory moving window technique. *Journal of Psycholinguistic Research, 25* (2), 273-290.

Filter, M. D. & Hyman, M. (1975). Relationship of acoustic parameters and perceptual ratings of esophageal speech. *Perceptual and Motor Skills*, *40* (1), 63-68.

Fujimoto, T., Kinishi, M., Mohri, M., & Amatsu, M. (1994). Mechanism of neoglottic adjustment for voice variation in tracheoesophageal speech. *Nippon Jibiinkoka Gakkai Kaiho*, *97* (6), 1009-1018.

Grolman, W, Eerenstein, S. E., Tan, F. M., Tange, R. A., & Schouwenburg, P. F. (2007). Aerodynamic and sound intensity measurements in tracheoesophageal voice. *Journal for oto-rhino-laryngology and its related specialties*, *69* (2), 68-76.

Kuttruff, H. (2007). *Acoustics: An introduction.* London: Taylor & Francis.

Liu, H., Wan, M., Wang, S., Wang, X., & Lu, C. (2005). Acoustic characteristics of Mandarin esophageal speech. *Journal of the Acoustical Society of America, 118* (2), 1016-1025.

Max, L, Steurs, W, & de Bruyn, W. (1996). Vocal capacities in esophageal and tracheoesophageal speakers. *Laryngoscope, 106* (1), 93-96.

Most, T., Tobin, Y., & Mimran, R. C. (2000). Acoustic and perceptual characteristics of esophageal and tracheoesophageal speech production. *Journal of Communication Disorders*, *33* (2) 165-181.

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

Olszański, W., Gieroba, R., Warchoł, J., Morshed, K., & Gołabek, W. (2004). Acoustic analysis of tracheoesophageal speech in comparison to esophageal speech after total laryngectomy. *Otolaryngologia polska. The Polish otolaryngology, 58* (3), 473-477.

Ptacek, P. H., Sander, E. K., Maloney, W. H., Jackson, C. C. R. (1966). Phonatory and Related Changes with Advanced Age. *Journal of Speech and Hearing Research*, *9*, 353-360.

Robbins, J. (1984). Acoustic differentiation of laryngeal, esophageal, and tracheoesophageal speech. *Journal of Speech and Hearing Research*, *27*, 577-585.

Robbins, J., Fisher, H. B., Blom, E. D., & Singer, M. I. (1984). A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production. *Journal of Speech and Hearing Disorders, 49*, 202-210.

Sataloff, R. T. (2006). *Vocal health and pedagogy* (2nd ed.). San Diego, CA: Plural Pub.

Schindler, A., Canale, A., Cavalot, A. L., Albera, R., Capaccio, P., Ottaviani, F., & Schindler, O. (2005). Intensity and fundamental frequency control in tracheoesophageal voice. *Acta otorhinolaryngologica Italica*, *25* (4), 240–244.

Schutte, H. K. & Nieboer, G.J. (2002) Aerodynamics of Esophageal Voice Production with and without a Groningen Voice Prosthesis. *Folia phoniatrica et logopaedica*, *54* (1), 8-18.

Sulter, A. M., Schutte, H. K., & Miller, D. G. (1995). Differences in phonetogram features between male and female subjects with and without vocal training. *Journal of Voice*, *9* (4), 363-377

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

van Gogh, C.D.L., Festen, J.M., Verdonck-De Leeuw, I.M., Parker, A.J., Traissac, L., Cheesman, A.D., & Mahieu, H.F. (2005). Acoustical analysis of tracheoesophageal voice. *Speech Communication, 47* (1-2), 160-168.

Wood, S. (2005). Praat for beginners [Manual]. Retrieved from http://www.ling.lu.se/persons/Sidney/praate/

Appendix

北風和太陽

有一次，北風同太陽喺度拗緊邊個叻啲。佢哋啱啱睇到外面有個人行過，哩個人著住件大褸同恤衫。佢哋就話嘞，邊個可以整到哩個人除咗件褸呢，就算邊個叻啲。於是，北風就拼命咁吹。點知，佢越吹得犀利，嗰個人就越係會執實件褸。最後，北風冇晒符，唯有放棄。跟住，太陽出嚟晒咗一陣，嗰個人就即刻除咗件褸嘞。於是，北風唯有認輸啦。

Acknowledgement

The dissertation was regarded as a unique, individual project for students who study Speech and Hearing Sciences in University of Hong Kong. However, it can never be completed without sincere help and support from different parties. The author would like to thank Dr. Lawrence Ng Man Wa from Speech and Hearing Sciences Division in University of Hong Kong for his close supervision and support throughout the process. A lot of knowledge about acoustics and characteristics of alaryngeal speakers, and critical thing were gained. The author would also like to acknowledge the staffs and participants in New Voice Club in Hong Kong for their kind assistance and genuine support during data collection. Special thanks given to my classmate Florence Yip who accompanied with me during data collection and gave support to my work. Last but not least, the author would like to deliver her greatest thanks to parents and her friends who deliver continual support and love.