
| | |
|---|---|
| **Title** | **Tone production using inspiratory phonation by Cantonese speakers** |
| **Other Contributor(s)** | **University of Hong Kong** |
| **Author(s)** | **Tsoi, Ka-ki, Iris; 蔡嘉琪** |
| **Citation** | |
| **Issued Date** | **2008** |
| **URL** | **http://hdl.handle.net/10722/123873** |
| **Rights** | **Creative Commons: Attribution 3.0 Hong Kong License** |

# Tone Production using Inspiratory Phonation by Cantonese Speakers

Tsoi Ka Ki, Iris

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, 2008.

ABSTRACT

The present study examined the Cantonese tones produced by native adult Cantonese speakers using expiratory (EP) and inspiratory (IP) phonation. Fundamental frequency (F0) values were measured from the Cantonese syllable /ji/ produced at six contrastive tones using EP and IP by forty native Cantonese speakers. Speech samples collected were identified in a perceptual experiment by ten naïve listeners. Results showed a higher accuracy of tone identification was found for rising tones (tone 2 and tone 5) when produced using IP. F0 contours revealed similarity patterns between IP and EP rising tones, while falling tones showed the most differences. The success in producing lexical tones using IP was likely to be related to the physiology of the vocal mechanism during IP.

# INTRODUCTION

Phonation is accomplished by setting the vocal folds into vibration in a largely controlled manner. Phonation is normally egressive as sounds are produced with an outward flow of air. This is known as expiratory phonation (EP) (Timcke, von Leden, & Moore, 1958). Stroboscopic findings showed that, during EP, vocal folds achieve controlled and periodic openings and closings (Zemlin, 1998). To initiate EP, vocal folds are adducted by increased activity of glottal adductors including interarytenoid and lateral cricoarytenoid muscles (Zemlin, 1998), combined with the building up of subglottal pressure (Raphael, Borden, & Harris, 2007). According to the Myoelastic-Aerodynamic Theory, opening and closing of the vocal folds are governed by different forces. The adducted vocal folds are “blown” open when the transglottal pressure differential is sufficiently high. As soon as the vocal folds are open, air escapes from the glottis, and subglottal pressure drops. When the transglottal pressure differential (TPD) falls below a certain value, the Bernoulli force caused by the narrowed glottis and muscle elasticity of the twisted vocal folds close the vocal folds. This is followed by another cycle of opening and closing (Raphael, Borden, & Harris, 2007). Following this sequence of events, vocal folds are set into periodic vibration and yield phonation. The minimum TPD required to initiate glottal vibration is known as the Phonation Threshold Pressure (PTP) (Jiang, O'Mara, Conley, & Hanson, 1999).

The flexibility and ingenuity of our phonatory system allow production of a wide range of different loudnesses and pitches. Physiologically, a change in pitch is associated with a change in the rate of vocal fold vibration (Zemlin, 1998). This is largely achieved by manipulating the longitudinal tension of the vocal folds via activity of the cricothyroid muscle, combined with an elevated subglottal pressure (Raphael et al., 2007). When cricothyroid muscle contracts, the distance between cricoid and thyroid cartilages reduces, resulting in an increase of vocal fold longitudinal tension (Zemlin, 1998). The subsequent thinning and elevation tension of the vocal folds facilitate the faster movement of the vocal muscles with smaller excursion. As vibration of vocal folds is initiated by the transglottal air pressure difference, a change in subglottal pressure and airflow rate may also change the vocal fold vibratory rate even with constant glottal tension (Zemlin, 1998).

In addition to EP, sounds can also be produced by a reverse flow of air, known as inspiratory phonation (IP) (Orlikoff, Baken, & Kraus, 1997; Robb, Chen, Gilbert, & Lerman, 2001). During IP, vocal folds are set into periodic vibration by the inward flow of air. However, EP and IP are different not only in the direction of airflow, but also the characteristics of glottal vibration. Maneuvering the laryngeal mechanism to yield IP phonation appears to be very different from EP (Robb et al., 2001). During IP, a negative pressure needs to be established in the oral cavity and subglottic region to generate an influx of airflow into the oral cavity. This is created by a negative intraoral and subglottal pressure. Soon after the vocal folds are adducted, the supraglottal and subglottal pressures start to differ. Because of the separation of the adducted vocal folds, subglottal pressure continues to drop while supraglottic intraoral pressure remains. Such pressure differential between the supraglottic and subglottal regions continues to increase. The vocal folds start to vibrate when the transglottal pressure differential (TPD) reaches a sufficient level. This is probably created by: (1) lowering of the larynx; and (2) negative pulmonary pressure generated by inhalatory muscular force (Orlikoff et al., 1997). Despite the difference in the vibratory pattern of vocal folds between EP and IP, the rules governing the opening and closing of vocal folds should be similar. Due to the unique vibratory pattern during IP phonation, IP voice tends to be partially periodic, with substantial amount of noise and significant frequency and amplitude perturbations (Orlikoff et al., 1997).

Despite the discomfort and lack of naturalness associated with IP, IP is found plentiful in neonatal cry and laughter (Grau, Robb, & Cacace, 1995). It is also a commonly used voice therapy technique for various types of dysphonia, such as spasmodic dysphonia and dysphonia of psychogenic origins (Boone, 1977; Colton & Casper, 1996). Though IP is used in laryngeal assessment (Wood, Jafek, & Chepniak, 1986), few researchers have examined the characteristics of IP voice. Orlikoff et al. (1997) investigated the general characteristics of IP using data collected from production of sustained /a/ for approximately two seconds by 16 English speaking adults. Their research revealed that IP was associated with higher fundamental frequency (F0), reduced vocal folds contact, increased in airflow, higher F0 variability (jitter), and greater amplitude perturbation (shimmer) as shown in electroglottographic (EGG) signals when compared with productions using EP. The findings of increased F0 and decreased vocal fold contact during phonation were further supported by Kelly and Fisher (1999), in which videostrobolaryngoscopic and acoustic data were obtained from female adults sustaining the vowel /i/. By comparing EP and IP productions of vowels /i/, /u/ and /a/, Robb et al. (2001) found that the average F0 was generally higher when using IP. This was consistent with the findings reported previously (Orlikoff et al., 1997).

Orlikoff et al. (1997) and Robb et al. (2001) partially revealed the control of vocal fold vibration during IP. According to Orlikoff et al. (1997), at least two vocal registers, fry and falsetto, were possible with an ingressive airflow. Robb et al. (2001) found that formant frequencies of vowels produced by using IP and EP were relatively well maintained. The presence of well-maintained formant frequencies was indirect evidence that vibration of vocal folds could still be manipulated volitionally when using IP.

Despite the previous reports on the unique vibratory characteristics of IP voice, a detailed acoustic and perceptual study of IP is not available. Knowledge of laryngeal control over the initiation and sustainment of IP is needed to better understand the mechanism of IP, especially in controlling the pattern and rate of vocal fold vibration. Rate of vocal fold vibration has been found to correlate with the perceived pitch (Zelmlin, 1998). Although both intonation and tone are perceptually associated with variation in perceived pitch, they are not mutually exclusive in their linguistic function (Bauer & Benedict, 1997). Intonation is closely associated with the grammatical structure and emotional status of the speaker while tone is a phonemic feature that plays an important role in determining the meaning of a word (Bauer & Benedict, 1997). For example, in a non-tonal language such as English, variation in intonation distinguishes a question from a statement, and emotions such as anger and irony. However, in a tonal language, difference in meaning can be signaled by use of different lexical tones. Therefore, tone perception is an important aspect of communication among speakers of a tonal language.

Many studies have reported listeners' perception of tone in relation to F0 (e.g. Gandour, 1983; Moore & Jongman 1997; Barry & Blamey, 2004). Speech sounds of different onset F0 and contours were presented to listeners. Findings from previous studies converged to an agreement that F0 location, including the onset and offset, and the specific F0 contour are the most important acoustic cues in tone perception regardless of linguistic experience (e.g. Gandour, 1988; Barry & Blamey, 2004; Khouw & Ciocca, 2007). Yet, despite the many studies in tone production and perception, none compared tone production using EP and IP. Seeing the significant physiological difference between EP and IP, it is valuable to investigate tonal production in a tone language using IP in an effort to obtain a more complete picture of IP mechanism. The use of IP as a voice therapy technique has been advocated in the literature, but the use of technique on tonal language is not well documented. With better understanding of IP in tone production, use of IP as a voice therapy technique for dysphonia and suppressing ventricular phonation in patients who are speakers of a tonal language can be objectively evaluated and more effective treatment regime can then be developed. This study (1) acoustically and perceptually examined tone production using IP by native speakers of Cantonese; and (2) compared the tone contours of the six Cantonese tones produced using IP with those using EP.

Cantonese is a variation of the Yue dialect family (Matthews & Yip, 1994). It is used mainly in Hong Kong, Macau, and the Guangdong province of mainland China, as well as in many Chinese communities in the Western countries. Cantonese is a tone language, in which there are three basic levels of tone: high, mid and low, and three different contours: level, rising and falling; yielding a total of six lexical tones: high-level (e.g.詩), high-rising (e.g. 史), mid-level (e.g. 試), low-falling (e.g. 時), low-rising (e.g. 市), and low-level (e.g. 事) (Mathews & Yip, 1994). Tone letters were introduced as a tool to describe lexical tones for Chinese. According to Chao (1947), tone letters are ‘simplified time-pitch graph of the voice’ which have been used for a long time (p. 24). However, for convenient notation, the six Cantonese tones are more easily indicated by using tone numbers from 1 to 6; namely tones 1 to 6 (Bauer & Benedict, 1997).

## METHOD

Participants

*Speakers*

Forty (20 male and 20 female) adults who were native speakers of Cantonese participated in the present study. The speakers were of ages ranged from 19 to 29 years, and had been residing in Hong Kong for at least 10 years. Only those who were capable of producing the speech materials using inspiratory phonation (IP) successfully as perceived by the investigator who was experienced in IP phonation were recruited. To maintain naturalness of productions, none of the speakers had received voice or phonetics training prior to the experiment. They had no reported history of speech, voice and/or language problem.

*Listeners*

Ten (5 male and 5 female) naïve listeners who were native speakers of Cantonese participated in the perceptual experiment. The listeners were also adult native speakers with reported normal hearing and had no prior speech and/or language problem.

Speech Materials

The speech materials used in the study included the CV syllable /ji/ produced at six Cantonese tones (high-level, high-rising, mid-level, low-falling, low-rising, and low-level) denoted as tones 1 to 6 respectively. This yielded six distinctive meaningful Cantonese words: 衣, 椅, 意, 移, 耳, 二, meaning “clothes”, “chair”, “idea”, “to move”, “ear”, and “two”.

Due to the limitation associated with IP, each Cantonese word was produced in isolation during the experiment.

Recording Procedure

Prior to the experiment, speakers were provided with training on using IP in producing the 6 Cantonese words. Only those speakers who were able to produce the different Cantonese words using EP and IP comfortably and intelligibly as judged by the investigator were recruited. During the recording, the speakers were presented with cards on each of which the Cantonese character corresponding to the target word was printed. They were instructed to produce the speech samples at a comfortable loudness level.

EP and IP speech samples were obtained by using a pre-amplification system (M-Audio, MobilePre USB) with a high-quality microphone (Shure, SM58) which was placed at approximately 10 cm away from the speakers' mouth. To eliminate extraneous noise, all recordings were carried out in a sound booth of the Speech Science Lab.

Perceptual Experiment

The perceptual experiment was carried out in a sound treated room. During the experiment, speech samples were randomized and presented to the listeners at a comfortable loudness level (around 60 dB) via high-quality headphones. Prior to the listening experiment, the listeners were provided with answer sheet on which the six Cantonese words were printed. Upon presentation of each speech stimulus, the listeners were instructed to choose the word they had perceived by circling the answers on the preprinted answer sheet. To provide sufficient time for the listeners to respond, an inter-stimulus of approximately five seconds was provided. The average percent correct identification of tones was calculated and correlated with F0 contours associated with IP and EP voices.

Measurements and Analysis

Acoustic analysis was carried out by using Praat. To measure F0 values, each acoustic signal was displayed in a waveform, and cycle-to-cycle F0 measurement was carried by using the automatic pitch estimation routine of Praat. Vertical cursors were automatically placed to indicate glottal cycles. The automatic pitch estimation was monitored by visual examination by the investigator in order to rule out erroneous measurement. The vertical cursors were visually checked by the experimenter before F0 was calculated. Only those that were correctly marked by the routine were used for F0 analysis. Upon completion of F0 measurements, F0 contours were generated for IP and EP voices associated with different tones in a F0 space. The results were compared with findings obtained from perceptual experiments (see discussion).

Perceptual data collected from the listening experiment were used to calculate the average percent correct identification of tones produced using EP and IP. Confusion matrices were generated for comparison of perceptual accuracy between IP tones and EP tones.

RESULTS

Perceptual Experiment

Cantonese speech samples produced using IP and EP were randomly presented and perceived by listeners. Percent correct identification of tones was averaged for each of the six Cantonese tones. It was found that the average identification of EP tones was better than IP tones (55.1% for EP vs. 35.4% for IP). T1 to T6 were used to represent tones 1 to 6 for simpler notation and easier understanding.

Table 1. Confusion Matrix of Responses to T1- T6 Produced by using Expiratory Phonation (EP).

| | Presented tones | | | | | |
|---|---|---|---|---|---|---|
| | T1 | T2 | T3 | T4 | T5 | T6 |
| Perceived tones | | | | | | |
| T1 | **36.0%** | 2.3% | 16.3% | 3.0% | 2.8% | 5.3% |
| T2 | 1.3% | **59.3%** | 2.5% | 1.8% | 14.0% | 5.5% |
| T3 | 42.0% | 2.3% | **38.5%** | 4.8% | 2.3% | 22.3% |
| T4 | 2.8% | 5.0% | 3.3% | **69.0%** | 3.3% | 10.0% |
| T5 | 4.0% | 27.8% | 3.8% | 3.5% | **75.0%** | 4.3% |
| T6 | 14.0% | 3.5% | 35.8% | 18.0% | 2.8% | **52.8%** |

Table 2. Confusion Matrix of Responses to T1-T6 Produced by using Inspiratory Phonation (IP).

| | Presented tones | | | | | |
|---|---|---|---|---|---|---|
| | T1 | T2 | T3 | T4 | T5 | T6 |
| Perceived tones | | | | | | |
| T1 | **34.0%** | 6.5% | 30.7% | 12.3% | 6.8% | 20.3% |
| T2 | 4.8% | **46.5%** | 3.8% | 3.5% | 28.0% | 3.8% |
| T3 | 27.3% | 7.3% | **33.4%** | 25.3% | 11.5% | 27.0% |
| T4 | 5.0% | 3.8% | 4.8% | **26.5%** | 5.0% | 9.5% |
| T5 | 10.5% | 28.5% | 4.5% | 6.0% | **41.3%** | 9.0% |
| T6 | 18.5% | 7.5% | 22.9% | 26.5% | 7.5% | **30.5%** |

Tables 1 and 2 show the confusion matrices for the identification of Cantonese tones produced using EP and IP respectively. Confusion matrices are constructed to indicate the error pattern of tone perception, with the diagonals showing the tones that are correctly identified. As shown in Tables 1 and 2, similar patterns of tone identification were seen in EP and IP productions except for tone 3.

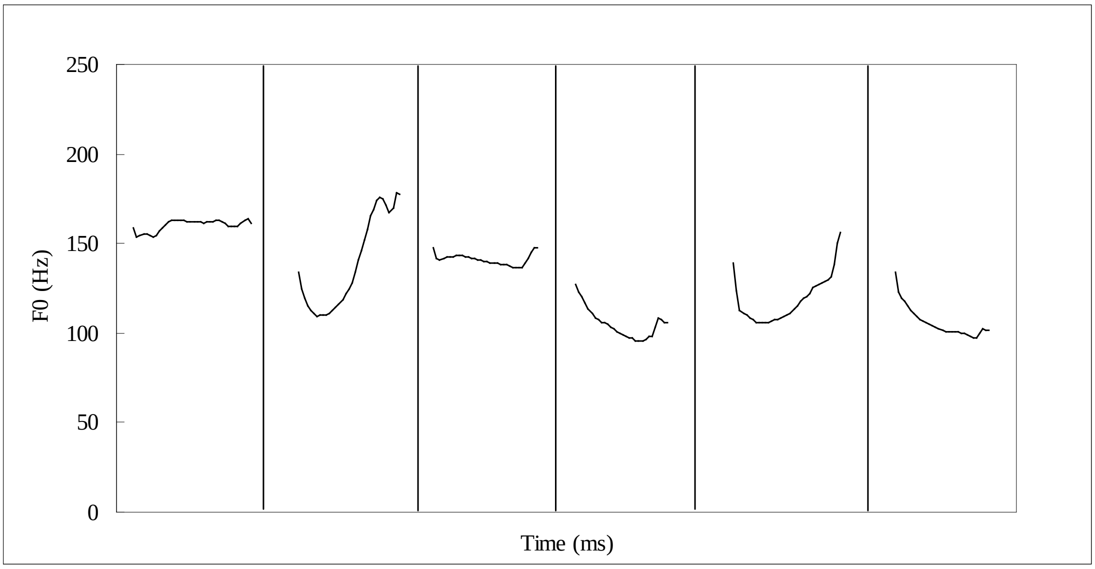


Figure 1. Typical F0 contours of T1 – T6 produced by a male using expiratory phonation (EP)

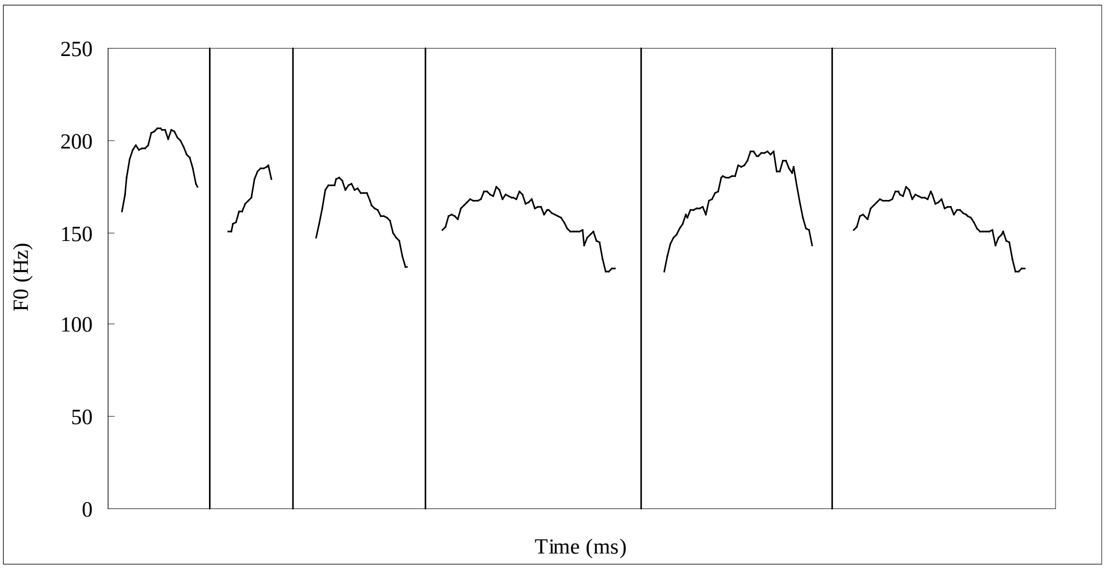


Figure 2. Typical F0 contours of T1- T6 produced by a male using inspiratory phonation (IP)

Production of T1 – T6 using EP

Cycle-to-cycle F0 values were measured from the vowel segment and F0 contours were developed to indicate the tone contours. Figures 1 and 3 show typical F0 contours associated with the six Cantonese tones produced by a male and a female speaker using EP. F0 contours of T1 to T6 are represented from left to right. Visual inspection of the EP F0 contours indicates that 29 out of 40 speakers produced T1 with a level contour and a difference between onset and offset F0 of less than 30 Hz. These contours ascended from the onset to the peak then dropped gradually to their endpoint, or dipped at the beginning and gradually ascended to the endpoint. Some speakers demonstrated a high-falling pattern of F0 contour for T1, yielding a F0 difference of as much as 20 Hz. The high falling feature of T1 has been considered rare in Cantonese speakers of Hong Kong (So, 1997).

A pattern of a slight drop of F0 from onset then increased sharply was noted in most speakers' production (38 out of 40) for T2. The sharp increase in F0 ranged from 23.3 Hz to 114.7 Hz. Only two of them showed a tone contour which straightly rose from the onset to its peak. The onset F0 was generally much lower than that of T1 which is also a high register tone (T1: 113.6Hz – 306.0 Hz; T2: 88.8 Hz – 256.6 Hz). However, a more diverted contour was noted in the mid-level tone (T3) produced using EP. Most speakers were able to maintain the onset and endpoint F0 within a range of 20 Hz, giving a level contour with or without a peak or a dip. Nine out of 20 speakers demonstrated a falling pattern with a drop of 25 Hz or more from its onset to its endpoint.

Although T4 is considered a falling tone, about 65% of the speakers demonstrated a V-shaped contour: F0 began with a drop from the onset and followed by a slight rise to its endpoint. Eleven speakers demonstrated a downward sloping tone contour in their productions. The drop in F0 contour varied from 7 Hz to 65 Hz. A typical T5 tone contour started from a mid-low F0 position, followed by a rise to its peak at the endpoint. The onset F0 ranged from 69.8 Hz to 251.7 Hz. A subsequent drop of F0 from the onset was common. The F0 increased by 9.2 Hz to 102.6 Hz from its dip.

T6 is considered as a level tone, and its onset and offset F0 appeared to be comparable (within a difference of 20 Hz). Twenty-four out of 40 speakers demonstrated such F0 contour for T6, some with a small elevation or depression within the contour. The termination of F0 contour tended to be lower than the onset or the peak in the majority of the speakers, and yielded a general declination in the overall contour. Meanwhile, a smooth straightly downward sloping pattern was noted in the tone contours by some other speakers.

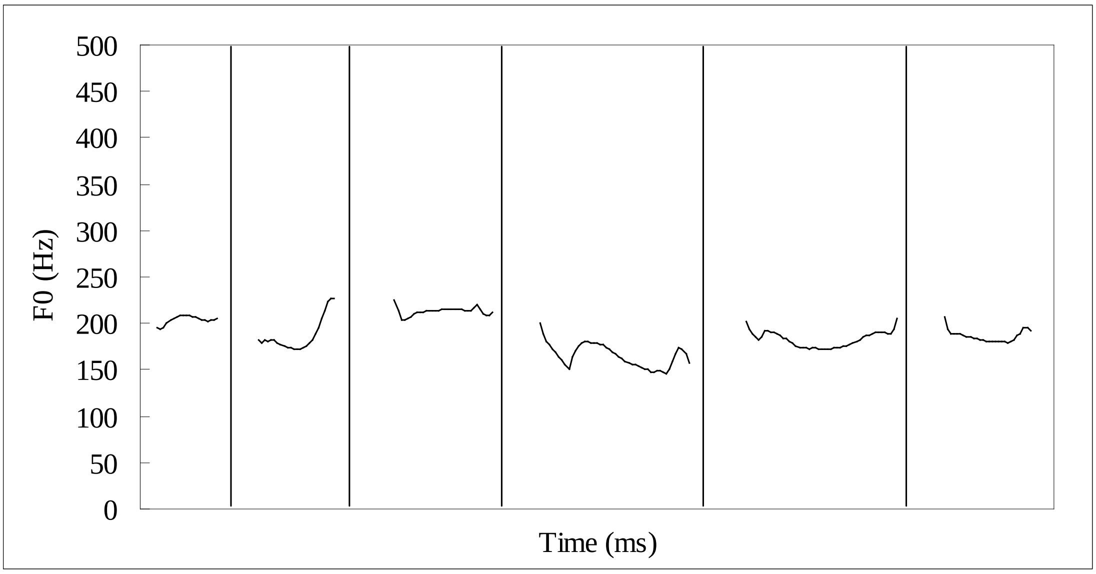


Figure 3. Typical F0 contours of T1 – T6 produced by a female using expiratory phonation (EP).

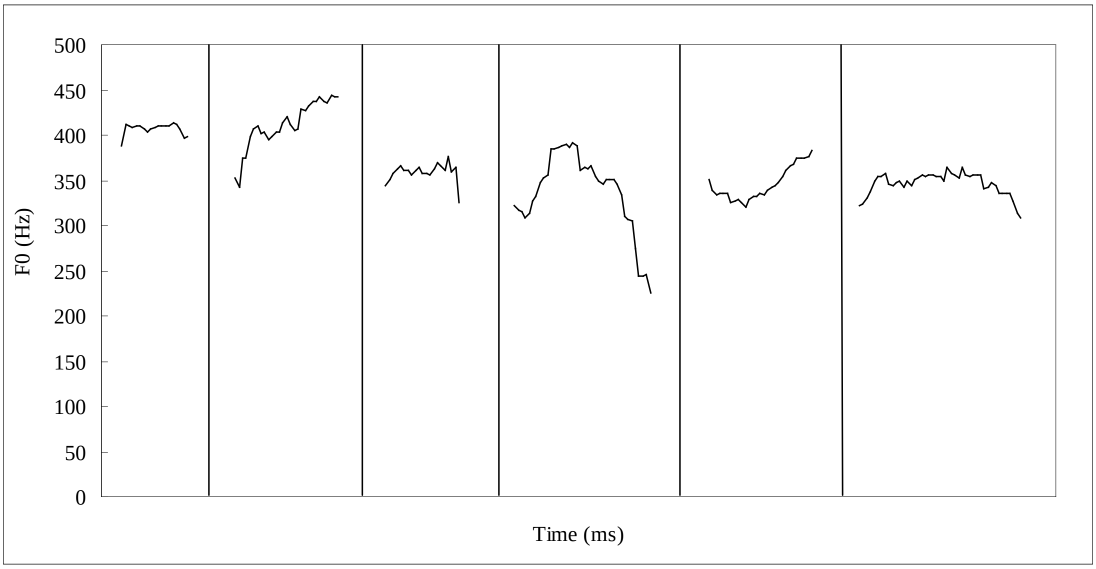


Figure 4. Typical F0 contours of T1 – T6 produced by a female using inspiratory phonation (IP).

Production of T1 – T6 using IP

F0 contours of the six Cantonese tones produced using IP are shown in Figures 2 and 4. Generally speaking, the F0 contours of IP tones only partially resembled the contours of EP tones. For T1 produced using IP, it was noted in most of the speakers that F0 increased from the onset to a peak well above other tones, and then dropped to its offset which was about 20 Hz above or below the onset F0. Yet, a high falling contour was noted in two speakers. The F0 contour dropped sharply after reaching the peak. Again, this high falling pattern of T1 contour was not common in Hong Kong Cantonese (Bauer & Benedict, 1997; So, 1997).

In 20 out of 40 speakers, the tone contour of T2 showed a rise of the F0 contour from the onset until where the contour terminated at the endpoint. A number of speakers demonstrated an inverted V-shaped F0 contour for T2; the contour rose from onset and fell to its endpoint. This was similar to that found in some of T1 contours. However, the rise from the onset to its peak in T2 was generally greater than that found in T1 contour.

The IP tone contour of T3 was similar to that of T1. The increase in F0 from its onset to the peak ranged from 166.7 Hz – 521.7 Hz, and then gradually dropped to its terminal point. The termination was slightly higher than the onset, usually by about 20-30 Hz. For T4, the F0 contour rose and fell sharply to its endpoint. It appears that it was the extent of F0 drop that discriminated a falling tone from a level tone which sometimes also had a downward sloping contour. The amount of F0 drop varied quite widely, from 30 Hz to over 100 Hz. However, 10 out of 40 subjects produced a contour with an endpoint higher than the onset.

When produced using EP, T5 is described as low rising with an onset F0 similar to that of T2. However, the IP contour of T5 showed a general trend of rising to a peak by 24 Hz -

310 Hz followed by a drop to its endpoint. The endpoint was usually higher than the onset and thus resembled an overall rise in F0, if considering only the onset and offset F0. About 10 speakers showed a straightly rising contour which was not common in EP. Similar to other level tones described previously, T6 usually reached a peak before falling back to its endpoint. The difference between the onset and offset varied greatly from 10 Hz to 50Hz. About 50% of the speakers had an F0 difference between the onset and offset of more than 15 Hz.

# DISCUSSION

## Perception

In the present study, tones produced using IP and EP were perceived by naïve listeners. Data indicated a lower percent correct identification of tones even produced using EP when compared with the perceptual results reported in previous studies (Khouw & Ciccoa, 2007; So, 2000). The lower percent correct identification of EP tones may be due to the limitation of the perceptual experiment. In the present study, all EP and IP speech samples were collectively randomized and presented to the listeners for identification. As mentioned by Gandour (1983), tone perception relied on both the location of F0 and tone contour. The range of onset F0 of EP and IP tones should have an effect on overall perception accuracy. The inclusion of both EP and IP tones produced by male and female speakers in the current study extended the F0 range for perception. The expanded F0 range covered by EP and IP tones may have created extra difficulty for the listeners to correctly perceive Cantonese tones. Moreover, gender information of the speech samples was not provided to the listeners. Previous researchers suggested that, tonal normalization is necessary for accurate tone perception, which depends on relative F0 values, rather than absolute F0 values (Leather, 1983; Moore & Jongman, 1997). With the known F0 difference between males and females, speech samples of mixed genders may affect the effectiveness and accuracy of tonal normalization and tone identification. Finally, as only isolated Cantonese words were used, contextual cues were unavailable to the listeners. All of these factors negatively affected listeners' perception of the stimuli and resulted in lower percent correct identification of IP and EP Cantonese tones.

Generally speaking, the pattern of listeners' identification of IP and EP tones was similar to that reported in previous studies (Khouw & Ciccoa, 2007; So, 2000). Identification of rising tones (T2, T5) and falling tones (T4) was generally better than that of level tones (T1, T3, and T6). Listeners tended to confuse tones with similar contour. They confused level tones with each others and between the two rising tones. T4 and T6 were confused probably because of the falling feature found in T6 as identified previously (Bauer & Benedict, 1997).

*EP vs. IP*

Results from the listening experiment indicated that perception of IP tones was poorer than that of EP tones. The listeners were unable to identify IP tones as accurately as EP tones. Yet, the pattern of errors in identification of IP tones indicated listeners were able to perceive the difference in tone contours while having difficulty in perceiving different location of F0 (both onset and offset). Among all six IP tones, T2 had the highest accuracy for identification, followed by T5. A slightly different identification pattern was found for EP tones. EP productions of T5 showed the highest accuracy in tone identification, followed by those of T4. The discrepancy in tonal identification accuracy of T2 and T5 may be due to the different amount of F0 contour deflection noted in EP and IP contours. IP productions of T4 showed the lowest identification accuracy while EP productions of T4 were the second highest. Such drastic difference in identification accuracy appears to be due to the different phonation methods. It is noted that the F0 contour of T4 produced using IP demonstrated an increase to a peak before falling and such pattern was absent from production by EP.

The present study revealed that, for EP, T3 was commonly recognized as T6. This is consistent with data reported by So (2000), and Khouw and Ciccoa (2007). In addition, the better perception of tonal contrast of T1-T3 than that of T3-T6 also supports that there is a better identification between T1 and T3 (Ciccoa & Lui, 2003). Therefore, the higher identification rate of perceiving T3 produced in IP as T1 should be related to the great overlapping of onset F0 by T3 in IP and T1 in EP.

Production

For each /ji/ syllable, cycle-to-cycle F0 values were measured and used to develop the corresponding F0 contour. Comparing the F0 contours of T1-T6 obtained from the present study with those from previous studies (Bauer & Benedict, 1997; Khouw & Ciocca, 2007; Zee, 1991), a number of similarities were identified. First of all, all level tones (T1, T3, and T6) demonstrated a general falling trend with a slight change of F0 of less than 30 Hz. This small difference in tone contour distinguished the level tones from three other tones, as suggested by Khouw and Ciocca (2007). The finding of small drop or rise in the F0 contour of level tones was consistent with those reported previously from adolescent (Khouw &

Ciocca, 2007) and adult speakers (Bauer & Benedict, 1997). The range of onset F0 among the three level tones differed systematically within 20 Hz which was used to differentiate the three tones as high, mid, and mid-low register as suggested in previous studies.

In addition to the rising contour noted in T2 and T5 produced using EP, it was also noted that T2 and T5 had a similar onset F0 value (88.8 Hz to 256.6 Hz for T2 and 69.8Hz to 251.7 Hz for T5). Such similarity in onset F0 value has been reported in previous studies based on which similar tone letter representation between T2 and T5 was suggested (Bauer & Benedict, 1997). The present data also suggested that T2 and T5 were different in the rise of F0 within the contour, in a way similar to data reported previously (Zee, 1991). T4 produced using EP showed a falling contour with a dip along the contour. Such unique contour matched the general agreement described previously (Bauer & Benedict, 1997; Khouw & Ciocca, 2007; Matthews & Yip, 1994; Zee, 1991).

Comparison of the Cantonese tones produced using EP

The six lexical tones produced using EP can be divided into two groups according to the onset and offset location of F0. The first group of tones was the high register group. This included T1 and T2. T1 showed the highest onset F0 while T2 had a lower onset F0 value. Yet, both T1 and T2 terminated similarly at a high F0 level. The rise of F0 from the onset to the endpoint was the greatest in T2 which yielded a distinctive F0 contour for T2. Perhaps due to such distinctive F0 pattern, T2 yielded greater perceptual accuracy (see Table 1). The second group of tones included those regarded as mid or low registers. Their onset and / or offset F0 values were relatively lower than those of the high register group and usually reached to half way of T2 contour. T3 generally had onset and offset F0 values slightly higher than those of T2 with an offset F0 only up to half way of T2 contour. Another level tone, T6, had lower onset than T1 and T3. The lowest F0 was usually found at the offset of T4. Despite the similar onset F0 of T4 and T6 and the falling feature demonstrated by T6, the range of F0 fall from the onset towards the endpoint differentiated T4 from T6 as a falling tone perceptually. As another rising tone, T5 had an onset F0 comparable with T2 but a smaller increase of F0 toward its endpoint differentiated the two rising tones from each other.

Tone production using IP

In general, the F0 values associated with IP were higher than that with EP. This is similar to previous findings reported by Kelly and Fishers (1999), and Orlikoff et al. (1997). Greater F0 variability within the syllable was seen in IP when compared with EP productions (Bauer & Benedict, 1997; Khouw & Ciocca, 2007). T1, T3, and T6 produced using IP were found to have greater F0 variability within tone contour and F0 onset when compared with the EP counterparts. Greater variation in F0 within tone contour, as noted in production using IP, suggested an inverted V-shaped contour in all three level tones instead of a general falling trend as noted in production using EP. Although the present perceptual data indicated greater confusion among level tones, some level tones were still perceived as other non-level tones (i.e. T2, T4, and T5). This indicates that listeners had difficulty recognizing the ‘level’ characteristics of T1, T3 and T6 produced using IP. As tonal identification depends significantly on F0 location (Gandour, 1983; 1988), the great overlapping between the onsets of T1 produced using EP and T3 produced using IP may account for the higher tendency of perceptual confusion between T3 and T1 as shown in present study (see Table 2).

Despite the use of IP, the rising component of both T2 and T5 contours was maintained. A comparison between EP and IP tones reveals a general widening of F0 range in IP tones. When contrasting T2 and T5 produced using IP, a different pattern was found. First of all, the range of F0 increase from onset to offset was greater in T5 using IP which was opposite to the fact that T2 had a greater increase when using EP. Additionally, the range of offset F0 of T5 was also higher than that of T2 produced in IP. This finding may in general indicate a steeper slope in tone contour of T5 using IP and a higher chance of perceiving T5 as T2. This is further confirmed by the perceptual results that T5 had a lower perceptual accuracy than that of T2 in IP production.

T4 exhibited a falling contour when produced using EP, but it weakly manifested some falling characteristics when produced using IP. A rise to a peak before falling to the end was common in most of the speakers. Ten speakers produced T4 using IP with F0 rising to its peak and then falling to the endpoint which was higher than the offset. Those T4 tones failing to maintain a falling contour commonly resulted in confusion, as indicated by the lowest identification accuracy. Some of the T4 tones were produced with a level contour, resulting in confusion with other level tones, especially T3 and T6 which had onset F0 at similar location.

In summary, tone production using IP is possible with varied performance. Beside the increase in F0 as reported in previous studies (Orlikoff et al., 1997; Robb et al., 2001), a trend of increasing the range of F0 within the contour was noted in this study. Moreover, it is noted that speakers had a better performance in maintaining a rising contour than a level or falling contour using IP. It is supposed that the better performance in increasing F0 is due to the unique mechanism of IP.

In normal EP phonation, F0 is increased primarily by increasing the longitudinal tension within the vocal folds, as a result of the interplay of cricothyroid muscle and thyroarytenoid muscle. By contracting the cricothyroid muscle and combining with the oppositional force of thyroarytenoid muscles, the vocal folds are lengthened and tensed, resulting in an increase in F0. As suggested by Kelly and Fisher (2001), cricothyroid muscle is contracted even before IP phonation as in normal inhalation in order to create an ingressive airflow. Therefore, IP is associated with a higher F0, as reported in the studies by Orlikoff et al. (1997), Kelly and Fisher (1999), and the present study. The involvement of cricothyroid muscle in the initiation of IP also implies the need to maintain vocal fold tension for phonation. Increase in subglottal pressure only produce negligible change in F0 in EP, its effect in F0 is supposed to be similar in IP also (Zemlin, 1998). Activity of the cricothyroid muscle mainly alters the longitudinal tension of vocal folds; it should be independent of the direction of airflow during phonation. It follows that the role of cricothyroid muscle in increasing F0 should not be affected even during IP. This suggested the performance in increasing F0 should be the same in IP and EP. This conclusion may explain the better performance in rising tone production in IP noted in the present study.

During EP, lowering of F0 is achieved mainly by the thyroarytenoid muscles which advance the arytenoid cartilages towards the thyroid cartilage (Zemlin, 1998). Vocal fold tension is then reduced due to the reduction in anteroposterior distance. Subsequently, a decrease in subglottal pressure is resulted (Zemlin, 1998). The same mechanism can be applied to the lowering of F0 during IP. However, as the vocal folds need to be tensed to yield an adequate TPD for phonation in IP, the decrease of vocal fold tension during IP may be limited. As a result, the effectiveness of lowering F0 during IP is hindered. In the present study, the poorer performance in lowering F0 in tone contour produced using IP supports this hypothesis. In the present data, difficulty was observed mainly in lowering the F0 from the onset or falling to the endpoint from the peak. The most significant effect was noted in the production of falling tone (T4). A number of speakers failed to maintain a falling contour and some failed to produce the tone with an offset F0 lower than onset F0.

# CONCLUSION

The present study investigated the production of Cantonese tones using EP and IP. Identification of the six Cantonese tones produced using IP and EP were compared. Despite the generally poorer identification accuracy for IP tones, native speakers of Cantonese were able to produce different lexical tones by varying F0 to produce distinguishable level and rising contours for linguistic purposes. Further investigation of the F0 changing mechanism during IP is needed to better understand the difference between production of EP and IP tones.

REFERENCES

Barry, J. G., & Blamey, P. J. (2004). The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese. *Journal of the Acoustical Society of America, 116*, 1739-1748.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology.* New York, NY: Mouton de Gruyter.

Boone, D. (1977). *The Voice and Voice Therapy* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall Inc.

Chao, Y. R. (1947). *Cantonese Primer.* Cambridge, MA: Harvard University Press.

Ciocca, V., & Lui, J. Y. K. (2003). The development of the perception of Cantonese lexical tones. *Journal of Multilingual Communication Disorders, 1*, 141-147.

Colton, R., & Casper, J. (1996). *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment.* New York, NY: Williams & Wilkins.

Finger, L. S., & Cielo, C. A. (2007). Reverse phonation - physiologic and clinical aspects of this speech voice therapy modality. *Brazilian Journal of Otorhinolaryngology, 73*, 271-277.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics, 11*, 149-175.

Gandour, J. (1988). Perception and production of tone in aphasia. *Brain and Language, 35*, 201-240.

Grau, S. M., Robb, M. P., & Cacace, A. T. (1995). Acoustic correlates of inspiratory phonation during infant cry. *Journal of Speech, Language, and Hearing Research, 38,* 373-381.

Jiang, J., O'Mara, T., Conley, D., & Hanson, D. (1999). Threshold Pressure Measurements During Phonation by Airflow Interruption. *Laryngoscope, 109,* 425-432.

Kelley, C. L., & Fisher, K. V. (1999). Stroboscopic and acoustic measures of inspiratory phonation. *Journal of Voice, 13,* 389-402.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35,* 104-117.

Leather, J. (1983). Speaker normalization in perception of lexical tone. *Journal of Phonetics, 11,* 373-382.

Matthews, S. & Yip, V. (1994). *Cantonese: A Comprehensive Grammar.* London: Routledge.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese Tones. *Journal of the Acoustical Society of America, 102,* 1864-1877.

Orlikoff, R. F., Baken, R. J., & Kraus, D. H. (1997). Acoustic and physiologic characteristics of inspiratory phonation. *Journal of the Acoustical Society of America, 102,* 1838-1845.

Raphael, L. J., Borden, G. J., & Harris, K. S. (2007). *Speech Science Primer: Physiology,* *Acoustics, and Perception of Speech.* Philadelphia, PA: Lippincott Williams & Wilkins.

Robb, M. P., Chen, Y., Gilbert, H. R., & Lerman, J. W. (2001). Acoustic comparison of vowel articulation in normal and reverse phonation. *Journal of Speech, Language, and Hearing Research*, *44*, 118-127.

So, L. K. H. (1997). Tonal changes in Hong Kong Cantonese. In Wright, S., & Kelly-Holmes, H. (ed.). *One Country, two systems, three languages: a survey of changing language use in Hong Kong.* Philadelphia, PA: Multilingual Matters. *pp. 80- 85.*

So, C. K. L. (2000). *Tonal Production and Perception Patterns of Canadian Raised Cantonese Speakers*. Unpublished master dissertation, Simon Fraser University, Canada.

Timcke, R., von Leden, H., & Moore, P. (1958). Laryngeal vibrations: measurements of glottic wave. Part I – The normal vibratory cycle, *Acta oto-laryngologica***,** *68*¸ 1-19.

Wood, B. Jafek, B., & Chepniak, R. (1986). Laryngeal dysfunction and pulmonary disorder. *Otolaryngology, Head, and Neck Surgery, 94,* 374-378.

Zee, E. (1991). Chinese (Hong Kong Cantonese). *Journal of the International Phonetics Association, 21,* 46-48.

Zemlin, W. R. (1998). *Speech and Hearing Science: Anatomy and Physiology* (4th ed.). Boston, MA: Allyn and Bacon.

# APPENDIX A

Stimuli for Recordings of Speech Samples

| | |
|---|---|
| Tone 1 | 衣 |
| Tone 2 | 椅 |
| Tone 3 | 意 |
| Tone 4 | 移 |
| Tone5 | 耳 |
| Tone 6 | 二 |

APPENDIX B

Consent Form of Participation for Speaker

## 參與研究同意書

以粵語為母語的人士用吸氣發聲法產生粵語聲調

香港大學言語及聽覺科學部現邀請閣下參與由言語及聽覺科學四年級生蔡嘉琪主理的研究調查。這是一項關於粵語(廣東話)聲調的學術研究，旨在探討粵語聲調在不同發聲方法下的聲學表現及發聲機制的關係。

何謂吸氣發聲法?

日常說話，聲音是透過由肺部呼出的空氣震動聲帶而產生。而吸氣發聲法則利用自主地吸入的空氣震動聲帶而發聲。

實驗過程

首先，研究員會教導閣下如何使用吸氣發聲法發聲，然後再練習使用吸氣發聲法讀出六個目標中文單字 (衣，椅，意，移，耳，二)。研究員會要求閣下用兩種不同的發聲方法讀出六個練習過的目標單字。每個單字需要分別用兩種發聲方法各讀三次。

整個過程將會錄音，所有錄音樣本會用作進一步分析。整個錄音過程需時約廿五分鐘並在菲臘牙科醫院內五樓的隔音室內進行。在錄音過程中，可能會覺得喉部有少許不適。在過程中，閣下可以在每一個字的錄音完成後休息和飲用提供的清水以減低可能引起的不適。

是次參與純屬自願性質，閣下可隨時終止參與是次研究，有關決定將不會引致任何不良後果。而所有有關之個人資料將會被銷毀。

閣下有權向研究員要求聽取個人的錄音樣本。所有收集的資料只作研究用途，個人資料將絕對保密。如閣下對是項研究有任何問題，請現在提出。

如日後閣下對是項研究有任何查詢，請與蔡嘉琪聯絡(9186 7140/ iris123@hkusua.hku.hk )。如閣下想知道更多有關研究參與者的權益，請聯絡香港大學非臨床研究操守委員會 (2241-5267)。

如閣下明白以上內容，並願意參與是項研究，請在下方簽署。

參與實驗者姓名：__________________ 研究員姓名：______________________

參與實驗者簽署：__________________ 研究員簽署：______________________

簽署日期：________________

個人資料

姓名:_________________(中文)

______________________________________ (英文)

性別:___________ 學歷: _____________

年齡:___________歲; 出生日期: ______________________

1. 你有沒有聲線的問題?

☐有 ☐沒有

2. 你有沒有呼吸的疾病?

☐有 ☐沒有

3. 你有沒有聽覺的疾病?

☐有 ☐沒有

4. 你在香港居住超過十年嗎?

☐是 ☐不是

5. 你是否以廣東話作為你日常主要的溝通語言?

☐是 ☐不是

6. 你有接受過拼音或語言學的訓練嗎?

☐有 ☐沒有

APPENDIX C

Consent Form of Participation for Listener

# 參與研究同意書

以粵語為母語的人士用吸氣發聲法產生粵語聲調

香港大學言語及聽覺科學部現邀請閣下參與由言語及聽覺科學四年級生蔡嘉琪主理的研究調查。這是一項關於粵語(廣東話)聲調的學術研究，旨在探討粵語聲調在不同發聲方法下的聲學表現及發聲機制的關係。

實驗過程

閣下需要聆聽一些中文單字的錄音聲帶。錄音將由電腦經耳筒直接播放。在研究員播放聲帶時，請閣下辨別出錄音中的是哪個中文字。閣下有 5 至 10 秒時間作答, 答案只需要在預製的答案紙上圈出便可。每段錄音只可重覆播放一次，請細心聆聽。

整個過程需時約一小時。在過程中，閣下可以隨時向研究員示意暫停, 稍作休息，每十五分鐘亦會有一段三分鐘的休息時間。

是次參與純屬自願性質，閣下可隨時終止參與是次研究，有關決定將不會引致任何不良後果。而所有有關之個人資料將會被銷毀。

所有收集的資料只作研究用途，個人資料將絕對保密。如閣下對是項研究有任何問題，請現在提出。

如日後閣下對本研究有任何查詢，請與蔡嘉琪聯絡(9186 7140/ iris123@hkusua.hku.hk )。如閣下想知道更多有關研究參與者的權益，請聯絡香港大學非臨床研究操守委員會 (2241-5267)。

如閣下明白以上內容，並願意參與是項研究，請在下方簽署。

參與實驗者姓名：__________________ 研究員姓名：______________________

參與實驗者簽署：__________________ 研究員簽署：______________________

簽署日期：________________

個人資料

姓名:________________(中文)

____________________________________(英文)

性別:__________　　學歷: ____________

年齡:__________歲; 出生日期: ____________________

你有沒有聽覺的疾病?

☐有　　　　　　　　☐沒有

你在香港居住超過十年嗎?

☐是　　　　　　　　☐不是

你是否以廣東話作為你日常主要的溝通語言?

☐是　　　　　　　　☐不是

你有接受過拼音或語言學的訓練嗎?

☐有　　　　　　　　☐沒有