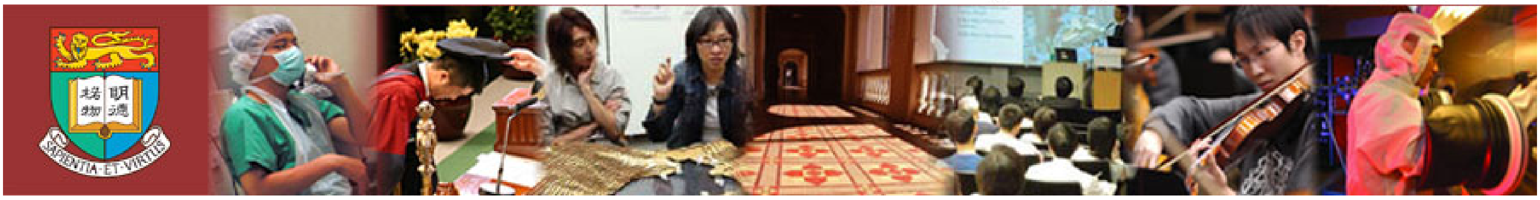


An acoustic analysis of the Cantonese whispered tones

Cheung Ka Yee

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, April 30, 2004.

Abstract

The aim of this project was to conduct an acoustic analysis of the Cantonese whispered tones in order to find out which acoustic cues are likely to be used by listeners to identify lexical tones. Speech samples collected in a perceptual study carried out by Cheung (2003) were used as stimuli. Speakers included 6 native Cantonese speakers. The speech materials were vowels /a/, /i/, /u/ and diphthong /œi/ segmented from Cantonese single words embedded in carrier phrases produced by each speaker. The first (*F1*), the second (*F2*), the third (*F3*) formant frequencies of the selected segments and the intensity at 9 points within the selected interval were estimated by using the Praat version 4.1.17 software (Boersma & Weenink, 2003); the segment duration, the averages and the shifts of intensity and formant frequencies were also calculated. Discriminant analysis revealed that the listeners' perception of whispered tones was cued primarily by average F2 and average F3. The findings were consistent with those of preceding acoustic studies on whispered speech.

Introduction

Cantonese is a tonal language in which pitch variations differentiate words (Matthews & Yip, 1996). Fundamental frequency (*F0*), which is produced by periodic pulsing of vocal folds, is found to be the main relating factor in the perception of related tonal differences (Fok, 1974). Six tones are found to be clearly distinctive in Cantonese (So, 1996). They differ from each other in terms of level—high, mid and low, and direction of movement—rising, falling and level (Bauer & Benedict, 1997). Customarily, the pitch range is divided into five levels and tone contours are characterized in terms of the respective starting and ending pitch levels—high level (55), mid level (33), low level (22), high rising (25), low rising (23) and low falling (21) (Bauer, 1998; Matthews & Yip, 1996). It should be noted here that the level and direction of movement of these tone contours are of relative but not absolute values (Bauer & Benedict, 1997).

Whispering is a naturally occurring speech act in which phonation is absent (Tartter, 1989). Unlike normally phonated speech in which resonances of the vocal tract are excited by periodic vibratory movements of vocal folds (Morris & Clements, 2002), aperiodic sound is generated by rapidly flowing air passing through constrictions in the vocal tract in whispered speech (Solomon, McCall, Trosset & Gray, 1989). Whispering is used for various purposes. It may be included as a part of voice therapy when vocal rest or minimal voice use is indicated, though its efficacy is still controversial (Colton & Casper, 1996; Solomon et al., 1989).

People also whisper when normal speech is inappropriate, whereas aphonic individuals who find hard to phonate may reconstruct normal speech from whispers by using voice prostheses (Morris & Clements, 2002).

There are no periodic vocal fold vibrations and thus no F0 in whispering. Because of the absence of F0, normal cues guiding listeners to identify the pitch of whispered stimuli presented are missing. It is therefore interesting to study how tones can be identified in whispered speech (Abramson, 1972; Higashikawa & Minifie, 1999). Physiological (e.g. Monoson & Zemlin, 1984; Solomon et al., 1989), perceptual (e.g. Abramson, 1972; Fok, 1974) and acoustic approaches (e.g. Kallail & Emanuel, 1984; Tartter, 1989) have been employed in investigating whispered speech. Among those studies, the acoustic studies of whispered speech will be reviewed in detail in the following discussion because other approaches are outside the scope of the primary interest in the current project.

Von Helmholtz (1954), as reported by Thomas (1969), used “listen-and-compare” method to determine resonances of whispered vowels. In his study, the listeners were asked to listen to whispered vowels and then compare the perceived pitches with some standard frequency source. A clear association between the perceived pitches of the whispered vowels and the formant frequencies was indicated— the pitch corresponded to the typical values of the first formant (*F1*) and second formant (*F2*) for the back vowels and the front vowels respectively. Meyer-Eppler (1957; as cited in Thomas, 1969) did not totally agree with this simple one-to-one correspondence, and he claimed that the association between the perceived pitch of whispered vowels and the formant frequencies was likely to be more complex. Specifically, movements of the third and higher formants or shifts in the intensity of the whispered vowels were proposed to contribute to the subjective changes in pitch. Thomas (1969) performed a “listen-and-compare” test after reviewing these two studies. Three listeners were asked to listen to nine whispered English vowels presented by a male and female speaker, and determine pitch by adjusting the output of a pure-tone oscillator to match with the stimulus accordingly. F2 frequency was found to correspond very closely with the pitch of all whispered vowels. The listeners were asked to expect more than one pitch in subsequent tests, additional pitches corresponding to F1 for the vowels /a/ and /ɔ/ were identified by two listeners.

The role of F2 in pitch perception of whispered vowels was investigated by McGlone and Manning (1979). In their study, they had 96 judges rank order whispered and phonated vowels produced by four adult males. The investigators found that there was a high correspondence between the perceived-pitch ranking of vowels and the mean F2 value in the whispered condition and the phonated condition. As a part of their experiment, acoustic energy above F1 was filtered out manually and the judges were asked to rank the presented stimuli in the same manner. The rank ordering changed significantly after the vowels were altered acoustically, so McGlone and Manning (1979) suggested that major clues for pitch detection should be at or above F2.

Further investigation of the role of formant frequency in the perception of pitch in whispering was conducted by Higashikawa, Nakai, Sakakura and Takahashi (1996). In their study, the whispered vowel /a/ was produced by six male and six female speakers at high, normal and low pitch. Five listeners were asked to rank the pitch level of the vowels as high, low or medium. Pitches correctly judged by at least three out of five listeners (i.e. four of the six male speakers and five of the six female speakers) were then analyzed acoustically. For the male vowels, significant differences between F1 in ordinary whispering and that in normal speech and between F1 in high-whispering and that in low-whispering were found. For the female vowels, significant differences between F1 in ordinary whispering and that in normal speech, and between F3 in high-whispering and that in low-whispering were found. The findings proved that speakers could change pitch during whispering and listeners could perceive the pitch difference, suggesting that reliable acoustic cues were available for the perception of pitch change in whispering. Acoustic analysis revealed that the perception of whispered pitch corresponded to formant frequencies, but the most important formant frequency which discriminated pitch in whispering was not verified.

Higashikawa and Minifie (1999) investigated the acoustical-perceptual relationships in the identification of whispered pitch in vowel production. In their experiments, synthetically generated vowel /a/ was prepared. The formant frequencies (i.e. F1, F2, or F1& F2) to be shifted, the magnitude (i.e. 20Hz, 40Hz, 60Hz) and the direction (i.e. up or down) of the formant frequency shifts were varied systematically. Ninety-four paired stimuli were presented to 17 judges who were asked if the second member of each pair sounded higher, lower or the same in pitch than the first member of the pair. The results showed that at least 20Hz formant frequency shifts were required for the perception of pitch changes in whispering. It was also found that more consistent change in the perception of whispered pitch was yielded when F1 and F2 were shifted at the same time, implying that simultaneous shifts in F1 and F2 were the most important cues to the perception of whispered pitch.

Two systematic studies have been conducted on the perception of Cantonese whispered tones. Fok (1974) investigated the perception of whispered tones in her perceptual study of tones in Cantonese. The stimuli were 21 isolated whispered words presented by a male and female speaker. Five listeners were asked to identify the tone by selecting a correct word among the relevant words for each stimulus on response sheets. The rate of identification was 16%, and it was slightly below that by chance. She thus claimed that recognition of tones in Cantonese whispered speech was poor in an absence of F0. Cheung (2003) carried out a perceptual study which aimed at investigating the recognition of lexical tones in Cantonese whispered speech. Six native Cantonese-speaking volunteers were recruited and they read aloud 24 Cantonese words which represented the six contrastive tones with four syllables /ji/, /fan/, /fu/ and /sœi/ embedded in the medial position of a carrier phrase “我會讀 (target) 俾

你聽” (i.e. “I read (target) to you”). Results showed that lexical tones in Cantonese whispered speech could be identified with 22% accuracy, which was relatively low. Results from binomial tests showed that this performance was above chance level. Statistically significant main effects of speaker identity, syllable identity and tone identity on tone identification were found. It was found that the identification accuracy was dependent on the speaker’s whispering style and experience of vocal training. Tones in certain syllables were better identified than the others (i.e. tones in /sœi/ and /fan/ better than that in /ji/ and /fu/). Different identification rates were also resulted across different tones—Tone 25 was significantly better identified than tone 33, tone 21 was significantly better identified than tone 33 and tone 23, while tone 55 was significantly worse identified than all tones except tone 33.

The current study thus aimed at investigating the perception of Cantonese whispered tones from the acoustic perspective. Together with physiological phenomena and perceptual phenomena, acoustic analysis contributes to a unified understanding of speech (Kent & Read, 2002). In respect of the above-mentioned literature, an effective communication by whispering (Kallial & Emanuel, 1985) and the above-chance-level recognition of whispered lexical tones in Cheung’s (2003) perceptual study, it appears that acoustic cues other than fundamental frequency are present in aiding the perception of Cantonese whispered tones. The research question of primary interest in the current project arose: What acoustic cues are available for the perception of Cantonese whispered lexical tones?

Method

Acoustic analysis

In the current study, speech samples employed in Cheung's (2003) perceptual study, that is, a total of 288 stimuli were analyzed acoustically by using the Praat version 4.1.17 software (Boersma & Weenink, 2003). The speech stimuli were vowels /a/, /u/, /i/ and diphthong /œi/ segmented from Cantonese single words embedded in the medial position of a carrier phrase "我會讀 (target) 俾你聽" (i.e. "I read (target) to you") produced by each speaker. In a tonal language, the vowel is the segment where the contrastive tones are usually conveyed, so the vowel of each speech sample was segmented (Ladefoged, 1993) while in the stimuli /sœi/, the diphthong /œi/ was segmented.

The following criteria were followed in selecting the segments: (1) the maximum formant frequency of each speaker was estimated to avoid wrong formant estimate in the low frequency region, while the window length was adjusted to maximize the resolution. The maximum formant frequency for the male speakers ranged from 4800-5000 Hz and that of the female speakers ranged from 5800-5900 Hz in the current study; (2) a speech sample was played back, and the target word embedded in the medial position of a frame sentence "我會讀 (target) 俾你聽" (i.e. "I read (target) to you") was identified; (3) the potential segment was zoomed in and the selection was determined by visual inspection of spectrograms. As in the case of /fan/ (see Appendix A), the initial fricative /f-/ and the vowel could be identified by a distinctive noise region of fricative while the nasal could be identified by an abrupt change in the spectrum at the transition between the vowel and the nasal consonant (Kent & Read, 2002). Similarly, the fricatives in /sœi/ and /fu/ could be identified with a relative ease. The selection in /ji/ (see Appendix B) was aided by identification of formant transitions at the onset of the vowel, while the formant transition portion was excluded from selection to eliminate its influence on formant estimate of the vowel. For some of the stimuli, the selection could also be aided by choosing the starting point and the ending point such that the selected segment did not contain a perceived quality of the glide consonant when listened to it.

A script was run upon selection of the interval to be analyzed. Parameters including the segment duration, intensity and the first three formant frequencies at nine time points within the selected interval were estimated by the computer program automatically. Averaged values and shifts of intensity and the formant frequencies within the selected interval were also measured manually.

Reliability

Pearson Product Moment Correlation was used to determine reliability of the acoustic measurements (Cleff, 1998). One tenth of the speech samples, that is, a total of 24 speech samples were randomly selected and were analyzed a second time by the same investigator and by another judge to test intra- and inter-judge reliability respectively. All of the correlations were significant at $p<0.05$ and they ranged from fairly high to very high. Hence the measurements were sufficiently reliable. Correlations for the acoustic measurements are shown in Table 1.

Table 1.

Correlations for the acoustic measurements

	Average					Shift			
	Duration (s)	Intensity (dB)	F1 (Hz)	F2 (Hz)	F3 (Hz)	Intensity (dB)	F1 (Hz)	F2 (Hz)	F3 (Hz)
Intra-judge	0.89	0.97	0.99	0.99	0.99	0.83	0.62	0.86	0.67
Inter-judge	0.62	0.97	0.94	0.99	0.99	0.80	0.54	0.52	0.53

Results

Acoustic measurements

The acoustic measurements, including the segment duration, averages of intensity and the first three formants, and shifts of intensity and the first three formants were measured. Appendix C to Appendix F present the means of the acoustic measurements of different segments for both male and female speakers.

Though there is no normative data of formant frequencies of the Cantonese whispered tones for reference, by referring to the averages of the formant frequencies across different segments, averages of formant frequencies, especially average F1 were found to be quite high.

In addition to the averages, the intensity shift and formant frequencies shifts were obtained. The magnitude of intensity shift ranged from –5.67dB to 3.07dB and –8.29dB to 5.47dB for males and females respectively, and the intensity reduced in most of the segments, except that in /i/, in which increases were measured in tone 55, 25, 33 and 23 in males and tone 25, 33, 21, 23 and 22 in females respectively. Positive formant frequencies shifts in raising tones, negative formant frequencies shifts in falling tones and relatively steady values in level tones were expected when looking at formant frequencies shifts, but these expected shifts were not indicated in most of the segments. These trends were only indicated in tone 25, 21 of /a/, tone 23 of /u/ and tone 25 of /œi/ presented by males, and tone 23 of /a/, tone 25 of /u/ presented by females. Great shifts were obtained in certain segments. For instance, for male speakers, all formant frequencies shifts exceeded –300 Hz in tone 22 of /i/, while for female speakers, shifts more than 400Hz were obtained in tone 22 of /i/, tone 55, 33 and 21 of /œi/.

Statistical analysis

In order to determine which acoustic parameters guide the listeners discriminate between tones, discriminant analysis was performed. All observations, that is, a total of 6912 independent judgments from the study by Cheung (2003) were included in the analysis. All measured acoustic parameters were regarded as independent variables while the listeners' responses were regarded as the dependent variables. There were totally six possible responses (i.e. six tones), so there were six groups to be discriminated. The application of the disciminant analysis would permit us determining which acoustic parameters (i.e. independent variables) guided the listeners perceive tone (i.e. making certain response). Because multiple group discriminant analysis was employed in the current study, canonical functions were evaluated. The canonical functions which were found to be statistically significant (i.e. $p<0.05$) were further examined. Contribution of the respective variables to the discrimination between groups was evaluated by comparing standardized coefficients for canonical variables. The first two variables with greatest absolute magnitude were considered.

Since different identification rates across whispered tones were resulted in Cheung's (2003) perceptual study, discriminant analysis across different tones was conducted to verify if factors associated with the perception of whispered tones differ across tones. Different contributions of formant frequencies on perception of whispered tones were expected across words because of varying vowel contexts. Discriminant analysis across different words was also conducted to look for other factors associated with the perception of whispered tones across words. The findings follow.

Discriminant analysis across different tones

Tone 55 Three canonical functions were statistically significant ($p<0.05$). The three roots accounted for 60%, 19% and 13% of the explained variance respectively. For the first root, the standard coefficients were 0.79 and 0.75 for average F2 and average F3 respectively, so both of them were reflected. In order, the standardized coefficients for the second root were –1.01 and 0.61 for average F2 and average F3, indicating that the second root reflected primarily the former variable. The standardized coefficients for the third root were 0.80 and –0.42 for duration and F1 shift respectively, showing that the third root reflected primarily duration.

Tone 25 Three canonical functions were statistically significant (p<0.05). The three roots accounted for 68%, 16% and 12% of the explained variance respectively. The standardized coefficients for the first root were –1.33 and 1.01 for average F3 and average F2 respectively, suggesting that average F3 was reflected primarily. In order, the standardized coefficients for the second root were –1.18 and 1.06 for average F2 and average F1, both of them were also reflected. The standardized coefficients for the third root were 0.60 and –0.50 for duration and average intensity, likewise, both of them were reflected.

Tone 33 Three canonical functions were statistically significant (p<0.05). The three roots accounted for 50%, 26% and 16% of the explained variance respectively. The standardized coefficients for the first root were –0.70 and 0.59 for F2 shift and average F3 respectively, indicating that F2 shift was reflected primarily. In order, the standardized coefficients for the second root were –1.80 and 0.87 for average F2 and average F3, suggesting that the former variable was reflected primarily. The standardized coefficients for the third root were –0.98 and 0.83 for average F2 and average F3, indicating that both of them were reflected primarily.

Tone 21 Two canonical functions were statistically significant ($p<0.05$). The two roots accounted for 38% and 30% of the explained variance respectively. In order, the standardized coefficients for the first root were –0.84 and 0.66 for average F2 and duration, indicating that average F2 was reflected primarily. The standardized coefficients for the second root were –1.32 and 1.1 for average F2 and average F3 respectively, likewise, average F2 was reflected primarily.

Tone 23 Four canonical functions were statistically significant ($p<0.05$). The four roots accounted for 64%, 16%, 10% and 7% of the explained variance respectively. The standardized coefficients for the first root were 0.65 and –0.50 for duration and F1 shift respectively, indicating that duration was reflected primarily. The standardized coefficients for the second root were 1.03 and –0.91 for average F3 and intensity shift, revealing that both variables were reflected. In order, the standardized coefficients for the third root were –1.17 and 0.57 for average F2 and duration, showing that the former one was reflected primarily. The standardized coefficients for the forth root were 1.19 and –0.79 for average F3 and average F1 respectively, indicating that average F3 was reflected primarily.

Tone 22 Three canonical functions were statistically significant ($p<0.05$). The three roots accounted for 39%, 29% and 18% of the explained variance respectively. The standardized coefficients for the first root were –0.95 and 0.93 for average F1 and average F2 respectively, revealing that both variables were reflected. The standardized coefficients for the second root were –0.80 and 0.73 for average F3 and F2 shift respectively, so both of them were also reflected. In order, the standardized coefficients for the third root were –0.68 and –0.44 for average F2 and duration, indicating that average F2 was reflected primarily.

Discriminant analysis across different words

/fan/ Two canonical functions were statistically significant ($p<0.05$). The two roots accounted for 62% and 22% of the explained variance respectively. Standardized coefficients of the first root were –1.02 and 0.72 for average F3 and average intensity respectively, indicating that the first root reflected primarily average F3. Standardized coefficients of the second root were 1.02 and 0.35 for intensity shift and F2 shift respectively, indicating that the second root reflected primarily intensity shift.

/fu/ Three canonical functions were statistically significant ($p<0.05$). The three roots accounted for 69%, 16% and 8% of the explained variance respectively. Standardized coefficients of the first root were 0.79 and 0.74 for F3 shift and F1 shift respectively, suggesting that the first root reflected both variables. Standardized coefficients of the second root were 1.48 and –0.87 for the average F1 and the average F2 respectively, showing that the second root reflected primarily the former one. Standardized coefficients of the third root were 1.81 and –0.63 for average F2 and average F3 respectively, indicating that the third root reflected primarily average F2.

/ji/ One canonical function was statistically significant (p<0.05). The root accounted for 78% of the explained variance. In order, the standardized coefficients were –1.00 and 0.97 for average F3 and average F2, thus the root reflected both variables.

/sœi/ Three canonical functions were statistically significant (p<0.05). The three roots accounted for 53%, 23% and 12% of the explained variance respectively. The standardized coefficients for the first root were 0.77 and 0.48 for duration and intensity shift respectively, the duration was thus reflected primarily. In order, the standardized coefficients for the second root were –1.78 and 0.84 for average F2 and average F3, so this root reflected primarily the former one. The standardized coefficients for the third root were 1.12 and –0.75 for average F1 and average F3 respectively, so average F1 was reflected primarily.

Table 2 and Table 3 summarize which acoustic parameters guided the listeners discriminate between tones across different tones and different words respectively. The order of canonical functions (roots) was numbered. In general, the higher order of the root, the lesser extent of its contribution.

Table 2.

Discriminant analysis across different tones

	Shift					Average			
Tone	Duration	Intensity	F1	F2	F3	Intensity	F1	F2	F3
	(s)	(dB)	(Hz)	(Hz)	(Hz)	(dB)	(Hz)	(Hz)	(Hz)
55	✓[3]							✓[1][2]	✓[1]
25	✓[3]					✓[3]	✓[2]	✓[2]	✓[1]
33				✓[1]				✓[2][3]	
21								✓[1][2]	
23	✓[1]	✓[2]						✓[3]	✓[2][4]
22				✓[2]			✓[1]	✓[1][3]	✓[2]

Note. Factor contributing to perception of whispered tone is marked with an asterisk. The root number is indicated in a bracket.

Results of discriminant analysis across different tones revealed that average F2 was associated with the perception of whispered tones in all tones. Average F2 was followed by average F3, which was associated with the perception of whispered tones in four out of six tones. Duration was also found to be associated with the perception of whispered tones in half of the tones. Average intensity, intensity shift and formant frequencies shifts seemed to contribute less comparatively.

Table 3.

Discriminant analysis across different words

		Shift				Average			
Word	Duration	Intensity	F1	F2	F3	Intensity	F1	F2	F3
	(s)	(dB)	(Hz)	(Hz)	(Hz)	(dB)	(Hz)	(Hz)	(Hz)
/fan/		✓ [2]							✓ [1]
/fu/			✓ [1]		✓ [1]		✓ [2]	✓ [3]	
/ji/								✓ [1]	✓ [1]
/sœi/	✓ [1]						✓ [3]	✓ [2]	

Note. Factor contributing to perception of whispered tone is marked with an asterisk. The root number is indicated in a bracket.

Results of discriminant analysis across different words revealed that average F2 was associated with the perception of whispered tones in most of the presenting words (i.e. Three out of four), while average F1 and average F3 were factors associated with the perception of whispered tones in half of the presenting words. Duration, intensity shift and formant frequencies shifts played minor roles comparatively in most of the words.

Discussion

The purpose of the current project was to analyze Cantonese whispered tones acoustically and to find out which acoustic cues help listeners discriminate between different whispered tones. As a general assumption, the pitch of vowels in normally phonated speech is related to the number of times the vocal folds vibrate in a second, formant frequencies are then believed to be contributors of the perception of whispered speech because of an absence of F0 (McGlone & Manning, 1979). Formant frequencies were thus of the main interest in this acoustic study. In addition to the formant frequencies, segment duration and intensity were also measured, so the investigator could check if any of them also contributed to the perception of whispered tone. Both averaged values and shifts of the formant frequencies and intensity were obtained and analyzed to find out the best estimation of their relationship with the perception of whispered tone. Discriminant analyses across different words and across different tones were performed separately to investigate for the presence of different factors associated with the perception of whispered tones across tones and words.

The findings obtained in the current project were in general agreement with most of the past results about the acoustic correlates of whispered pitch or tone, average F2 and average F3 were found to be the primary acoustic cues guiding listeners discriminate between different whispered tones. The significant roles played by the formant frequencies in perceiving whispered tones will be discussed. Possible explanations of the factors associated with the perception of whispered tones across different tones and words will also be proposed in the followings.

Significant roles played by the formant frequencies in perceiving whispered tones

Formant frequencies, as described by Kent and Read (2002), are resonances of the vocal tract and are excited by source of energy. In whispering, voiceless noise energy is generated by a partial abduction of the vocal folds, so formant frequencies are also observed. Formant frequencies change with the shape and length of the vocal tract, which are determined by the position of the lips, advancement of the tongue and lifting or lowering of the larynx. They vary among different vowels because they have different vocal tract configurations (Higashikawa et al., 1996).

In general, data presented in the current study indicated that formant frequencies, especially average F2 and average F3, were influential for the perception of whispered tones. As discussed by Kuhn (1975; as cited in McGlone & Manning, 1979; Kallail & Emanuel, 1984), F2 and F3 are associated with the resonance of the front cavity resonance. Specifically speaking, in the present study, vocal resonances in the front cavity are believed to play the most significant roles in determining if listeners are able to discriminate between different whispered tones.

As mentioned in the result section, the formant frequencies, especially F1, seemed to be very high in the present study. Higashikawa et al. (1996) also observed higher shifts in formant frequencies, especially in F1 in whispering than in normal speech. Transformation of the vocal tract configuration and change of the manner of speech in whispering were proposed. These explanations are also applicable in the current project, in which formant frequencies measured were higher than the expected ones. Great F1 difference between whispered and phonated vowels was indicated in Kallail and Emanuel (1984)’ s study as well. “The first formant is always dependent on the shape of the entire vocal tract” (Fant, 1960; as cited in Kallail & Emanuel, 1984, pp. 184), a reflection of great glottal difference manifested by F1 in whispering is thus hypothesized. But it is worth noting that the formant differences were only judged by visual inspection of the data in the current project, whether formant frequencies of the Cantonese whispered tones measured in the present study are really significantly higher than the voiced counterparts can only be verified when acoustic measurements of voiced counterparts produced by the same group of speakers are available.

Factors associated with the perception of whispered tones across different tones

Average F2 was found to be the primary contributors of tone discrimination and it was responsible for the discrimination in all tones. Followed by F2, average F3 also contributed much in tone discrimination, as discrimination of four out of six tones could be guided by it. The findings were also parallel with relevant studies, which claimed major cues for pitch detection should be at or above F2 (McGlone & Manning, 1979).

By referring to the acoustic cues guiding listeners discriminate low falling tone, which was best identified among all tones in Cheung’s (2003) study, it was found that only average F2 contributed to the discrimination, while at least two different acoustic cues were indicated in other tones. This uniqueness of low falling tone might explain its comparatively high identification rate in Cheung’s (2003) study. Except the low falling tone, the present data could not explain observations of different identification rates of the remaining tones.

Factors associated with the perception of whispered tones across different words

By referring to the results obtained from the discriminant analysis across different words, different acoustic cues were found to be responsible for the identification of whispered tones, aiding in explaining different identification rates of different words indicated in Cheung’s (2003) perceptual study. Generally speaking, averaged values of formant frequencies, especially F2 was found to be the primary contributors of discrimination, as it contributed to the tone discrimination in three out of four syllable types. Followed by average F2 were average F1 and average F3. Additionally, when average F1 was found to be one of the contributors, as in /fu/ and /sœi/, average F2 also played a role. This finding was also consistent with previous studies, in which simultaneous contribution of the first two formants were indicated (Higashikawa & Minifie, 1999). Duration was believed to be responsible for an overall better identification of /fan/ and /sœi/ in Cheung’s (2003) perceptual study. From the present data, duration was also found to be the major contributor for discrimination in /sœi/ but it was insignificant in discriminating tones in the stimuli /fan/. The findings could be explained by the criteria adopted in the segment selection. Because only vowel or diphthong segment, where the contrastive tones are usually marked over

(Ladefoged, 1993) was selected in the current project, an exclusion of the final consonant in /fan/ might eliminate the contribution of duration in perceiving tones in /fan/. Apart from duration, presence of other explanation for the better identification of /fan/ was thus suspected. According to the current data, shift in intensity was found to be a contributor to the tone discrimination in the stimuli /fan/, and it might serve as a possible explanation.

Limitations

The reliability of the measurements of segment duration and shifts in intensity and formant frequencies was not as high as other acoustic measurements, possibly due to the dramatic difference of the obtained values brought by different selection of the starting point and the ending point of the segment. Unless these two points were exactly the same, a discrepancy of few milliseconds could result in a great difference in the shifts. More intensive training is highly recommended for investigators who are interested in replicating the project so as to obtain the best estimation.

In the current project, stimuli presented by male and female speakers were not analyzed separately, and the sensitivity of statistical analysis to reveal factors associated with the perception of whispered tones might be reduced. Separate analysis can be performed in further studies to display more precise relationship between acoustic parameters and the perception of the whispered tones presented by males and females respectively.

Acknowledgements

I would like to take this chance to thank many people whose support made possible completion of this project. First, I would like to express my sincere gratitude to my supervisor, Dr. Valter Ciocca, for his continuous guidance on this project. I would also like to thank Dr. Eddy Lam, Ricky Chan and Fan Yiu, for their assistance in the statistical analysis. I must thank my family, Po, Pastor Ip and Miss So, for the love and concern they always give me freely. My classmates May Lam, Cindy Chan, Jess Hon and Olivia Yeung, who support me continuously, are also gratefully acknowledged.

References

Abramson, A. S. (1972). Tonal experiments with whispered Thai. In A. Valdman. (Ed.). *Papers in Linguistics and Phonetics* (pp. 31-44). The Netherlands: Mouton & Co.N.V.

Boersma, P., & Weenink, D. (2003). *Praat (Version 4. 1. 17).* University of Amsterdam, Institute of Phonetics Sciences. Available at: http://www.praat.org.

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

Bauer, R. S. (1998). Hong Kong Cantonese tone contours. In S. Matthews (Ed.). *Studies in Cantonese Linguistics* (pp. 1-33). Hong Kong: Linguistic Society of Hong Kong.

Cheung, Y. M. (2003). *Recognition of lexical tones in Cantonese whispered speech* (Unpublished dissertation). Hong Kong: The University of Hong Kong.

Clegg, F. (1998). *Simple statistics: A course book for the social sciences (15th ed.).* Cambridge: Cambridge University Press.

Colton, R. H., & Casper, J. K. (1996). *Understanding voice problems: A physiological perspective for diagnosis and treatment (2nd ed.).* Baltimore, Maryland: Williams & Wilkins.

Fok, C. Y. Y. (1974). *A perceptual study of tone in Cantonese.* Hong Kong: University of Hong Kong Press.

Higashikawa, M., Nakai, K., Sakakura,A., & Takahashi, H. (1996). Perceived pitch of whisphered vowels--relationship with formant frequencies: A preliminary study. *Journal of Voice, 10* (2), 155-158.

Higashikawa, M. & Minifie, F.D. (1999). Acoustical-perceptual correlates of “whisper pitch” in synthetically generated vowels. *Journal of Speech, Language, and Hearing Research, 42,* 583-591.

Kallail, K. J., & Emanuel, F. W. (1984). An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects. *Journal of Phonetics, 12,* 175-186.

Kallail, K. J., & Emanuel, F. W. (1985). The identifiability of isolated whispered and phonated vowel samples. *Journal of Phonetics, 13,* 11-17.

Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech (2nd ed.).*. San Diego: Singular Publishing Group, Inc.

Ladefoged, P. (1993). *A course in phonetics (3rd ed.).* Orlando: Harcourt Brace College Publishers.

Matthews, S., & Yip, V. (1996). *Cantonese: A comprehensive grammar.* New York: Routledge.

McGlone, R. E., & Manning, W. H. (1979). Role of the second formant in pitch perception of whispered and voiced vowels. *Folia Phoniatria, 31* (9), 9-14.

Monoson, P., & Zemlin, W. R. (1984). Quantitative study of whisper. *Folia Phoniatrica, 36,* 53-65.

Morris, R.W., & Clements, M. A. (2002). Reconstruction of speech from whispers. *Medical Engineering and Physics, 24,* 515-520.

So, L. K. H. (1996). Tonal changes in Hong Kong Cantonese. *Current Issues in Language and Society, 3* (2), 186-189.

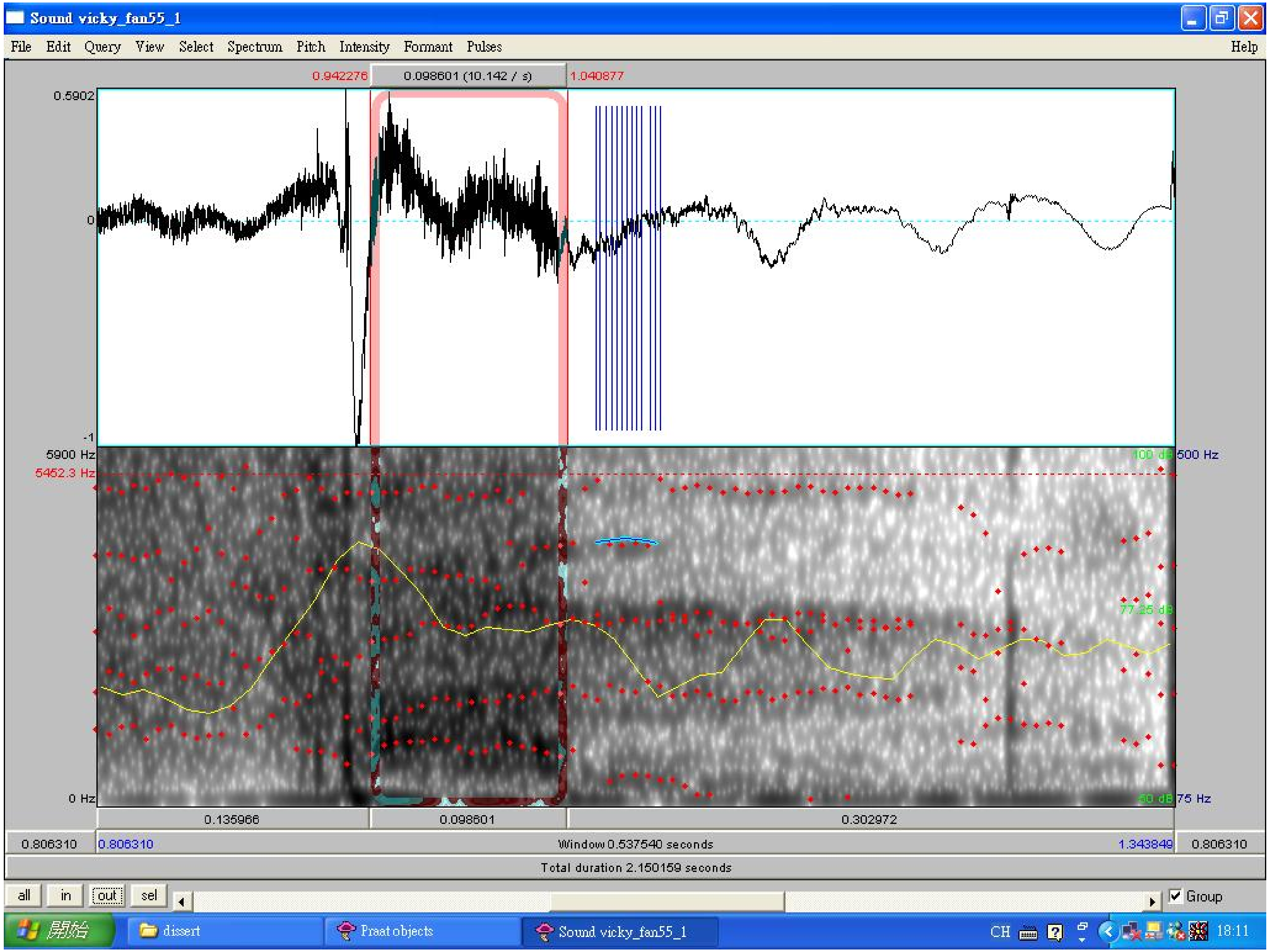
Solomon, N. P., McCall, G. N., Trosset, M. W. & Gray, W. C. (1989). Laryngeal configuration and constriction during two types of whispering. *Journal of Speech and Hearing Research, 32,* 161-174.

Tartter, V. C. (1989). What's in whisper? *Journal of the Acoustical Society of America, 86* (5), 1678-1683.

Thomas, I. B. (1969). Perceived pitch of whispered vowels. *Journal of the Acoustical Society of America, 46,* 468-470.

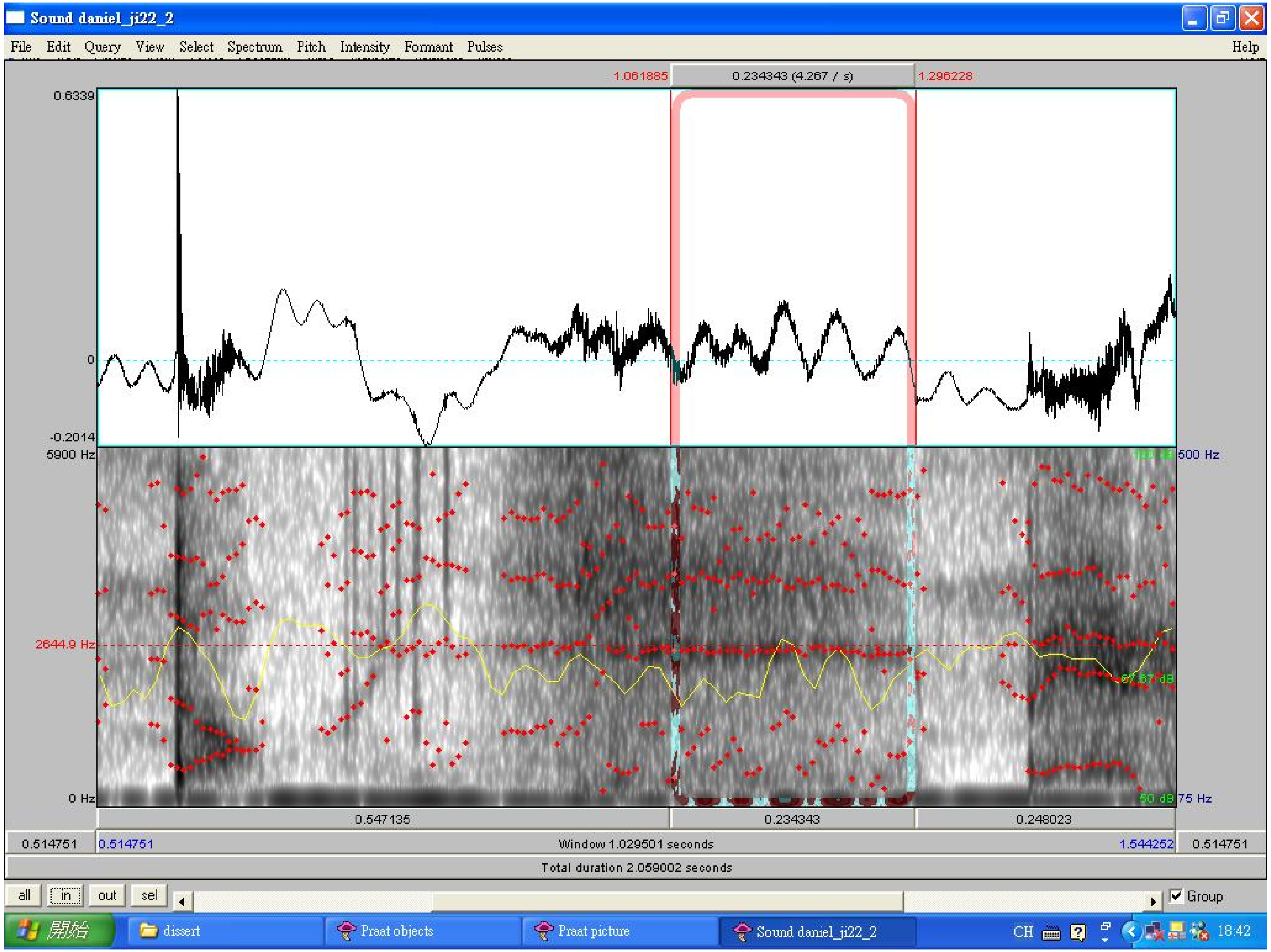
Appendix A

Segment selection of /fan/



Appendix B

Segment selection of /ji/



Appendix C

Means of acoustic measurements for /a/

			Shift				Average			
Gender	Tone	Duration	Intensity	F1	F2	F3	Intensity	F1	F2	F3
		(s)	(dB)	(Hz)	(Hz)	(Hz)	(dB)	(Hz)	(Hz)	(Hz)
Male	55	0.06	-0.03	21	-54	18	75	887	1581	2484
	25	0.07	-1.10	180	73	88	74	892	1514	2439
	33	0.07	-4.44	63	65	11	75	866	1516	2549
	21	0.09	-3.69	-68	-63	-33	74	928	1569	2519
	23	0.08	-1.16	-20	96	5	75	926	1505	2501
	22	0.09	-2.04	-118	-93	-182	74	872	1549	2504
Female	55	0.06	-4.94	76	80	-61	76	1020	1741	2849
	25	0.07	-1.32	71	-107	-43	77	994	1778	2893
	33	0.07	-8.29	185	83	-61	73	982	1768	2779
	21	0.10	-4.35	144	278	246	77	992	1765	2793
	23	0.11	-6.25	196	220	93	77	927	1727	2772
	22	0.10	-4.10	51	168	112	77	1025	1779	2813

Appendix D

Means of acoustic measurements for /i/

			Shift				Average			
Gender	Tone	Duration	Intensity	F1	F2	F3	Intensity	F1	F2	F3
		(s)	(dB)	(Hz)	(Hz)	(Hz)	(dB)	(Hz)	(Hz)	(Hz)
Male	55	0.20	3.07	53	117	-52	73	894	2251	3153
	25	0.14	1.47	-26	247	84	72	911	2238	3136
	33	0.16	2.66	203	-16	-126	70	877	2159	2989
	21	0.15	-0.34	140	-109	-316	73	772	2254	3099
	23	0.18	4.61	-118	-21	-343	73	910	2346	3197
	22	0.15	-1.71	-579	-339	-565	73	1008	2257	3141
Female	55	0.16	-0.13	252	366	-81	72	1125	2521	3633
	25	0.15	5.00	-64	-16	-194	73	1134	2501	3541
	33	0.14	3.07	206	67	-48	72	1193	2542	3593
	21	0.16	2.47	70	68	-229	71	1174	2406	3553
	23	0.17	1.88	51	-45	-299	72	1138	2409	3576
	22	0.11	5.47	461	274	-59	72	1204	2556	3604

Appendix E

Means of acoustic measurements for /u/

			Shift				Average			
Gender	Tone	Duration	Intensity	F1	F2	F3	Intensity	F1	F2	F3
		(s)	(dB)	(Hz)	(Hz)	(Hz)	(dB)	(Hz)	(Hz)	(Hz)
Male	55	0.22	-1.84	0	158	268	72	867	1570	2603
	25	0.23	0.65	-4	46	15	72	827	1581	2755
	33	0.21	-5.67	-144	-231	-92	71	844	1582	2563
	21	0.19	-2.94	133	24	171	72	767	1437	2553
	23	0.19	4.41	37	203	166	73	861	1553	2552
	22	0.20	-4.50	-42	20	98	71	804	1511	2677
Female	55	0.15	-2.38	-93	10	240	72	920	1739	2927
	25	0.20	3.65	82	194	193	74	889	1769	2930
	33	0.21	1.48	40	4	270	73	1003	1862	2960
	21	0.18	-6.75	29	165	253	71	913	1784	2939
	23	0.20	-0.29	-172	-135	-245	72	851	1662	2882
	22	0.22	-2.22	142	147	183	74	947	1830	2970

Appendix F

Means of acoustic measurements for /œi/

			Shift				Average			
Gender	Tone	Duration	Intensity	F1	F2	F3	Intensity	F1	F2	F3
		(s)	(dB)	(Hz)	(Hz)	(Hz)	(dB)	(Hz)	(Hz)	(Hz)
Male	55	0.27	-5.08	66	197	-220	72	751	1740	2595
	25	0.26	-1.02	192	339	118	73	785	1734	2578
	33	0.24	-4.31	-32	268	-25	71	739	1710	2576
	21	0.23	-0.71	-202	225	-146	72	664	1692	2514
	23	0.24	-2.50	-207	352	-190	73	685	1697	2533
	22	0.23	-1.08	-293	126	-233	73	735	1718	2569
Female	55	0.23	-2.91	67	456	1	75	813	2021	3004
	25	0.25	-2.46	-153	270	-293	73	804	1981	2892
	33	0.25	2.14	-298	212	-404	74	834	2045	2976
	21	0.23	-3.24	-313	128	-444	72	796	1975	2931
	23	0.25	-2.18	16	316	-355	75	798	2002	2922
	22	0.25	-1.75	323	431	-63	74	857	2008	2944