Perception of Cantonese tones by Canadian-Cantonese immigrants

Connie Kwok Lai So
Simon Fraser University

The present study examined the perceptual abilities of Cantonese-English immigrants in identifying the six lexical tones in Cantonese. Forty-eight Cantonese words in citation forms were identified by 30 adult listeners in two groups of immigrants (Early and Late) and in a comparison group of native Hong Kong Cantonese speakers. The two immigrant groups differed in age of arrival (young children vs. teenagers) in Canada, while the (adult) native listeners had lived in Canada for less than a year. Listeners' identification scores were analyzed, and confusion matrices for individual groups were constructed and compared. The results revealed that the Late immigrants obtained higher scores in identifying the lexical tones than did the Early immigrants. Moreover, hierarchical patterns in the immigrants' misidentifications were observed.

1 Introduction

Weinreich (1953) has suggested that the two languages of bilingual speakers inevitably influence one another. Studies with respect to language attrition documented that various forms of modifications of one's first language (L1) occur in languages in contact situations. For example, Major (1992, 1997) and Sancier & Fowler (1997) have reported that native speakers were subject to phonetic changes in a second language (L2) environment. In these studies, native speakers of L1 modified and approximated their voice onset time values to the ones of the dominant language in the place where they were living in. However, relatively little research has documented changes/ differences in tonal production and or perception of Cantonese by overseas Cantonese speakers living in an English-speaking country. Light (1977) reported a case study in which a native Cantonese-speaking child experienced disintegration of her tonal system (i.e., tones were mispronounced) after she had moved to an English-speaking environment. So (2000) found that the tonal productions of native (Hong Kong) Cantonese immigrants living in Vancouver, British Columbia, deviated from those of native Cantonese speakers living in Hong Kong. In particular, an apparent reduction in tonal space[1] was observed in their speech utterances; the reduction appears to have greater impact on the tonal space of early immigrants than that of late immigrants.[2]

The purpose of this perception study was to investigate if there are any score differences in identifying Cantonese tones between two groups of Cantonese immigrants

[1] Reductions in tonal space refer to the obtained tone letter values, which were transformed from the physical measurements of fundamental frequency (in Hz), that were far below the values described in literature. For example, the tone letter values for a high level tone [55] was measured as [44] and the one for a high rising tone [25] was measured as [23].

[2] The differences between the Early and the Late immigrant groups will be discussed in Section 2.1.

(Early and Late) and native Cantonese speakers of Hong Kong. If the identification scores of the immigrant groups are compatible with those of the native speaker group, it will imply that they have similar patterns in their tonal systems. If the scores are not compatible, it will imply that their tonal systems deviate from that of the native Cantonese speakers. On the basis of the findings of So (2000), it was predicted that identification scores by the Early and the Late immigrant groups would be lower than those of the native speaker group. In particular, the scores for the Late immigrant group would be higher than those for the Early immigrant group.

1.1 Background

1.1.1 Tonal perception

It is well known that fundamental frequency (Fo) varies with the production of individual speakers, ranging from high Fo for children and women to low Fo for men. In order to perceive lexical tones in speech, normalization of tonal patterns is employed (Ching, 1984; Gandour, 1994; Abramson, 1997). Listeners are assumed to infer the speech source in order to extract similar tonal patterns from the acoustic input (Ching, 1984). In other words, listeners need to infer the speaker's voice range, ignore the absolute Fo values, and extract the invariant Fo patterns of the speech (Abramson, 1976; Leather, 1983). According to Fourcin (1978), normalization depends on an ability to perceive or extract the (tonal) patterns; thus, this perceptual skill is not innate, but acquired. Ching (1984) further suggests that it is a learning process that exists at the basic level of lexical tone labeling.

1.1.2 Cantonese tones

Cantonese, a variety of Yue of the Chinese languages, is spoken across southeast China, including Guangzhou, Hong Kong, and Macao (Arendrup, 1994:521). It is widely spoken in many overseas Chinese communities in Malaysia, Indonesia (Tang & Maidment, 1996), San Francisco, Sydney and Vancouver. It is also one of the Chinese languages taught in many universities in North America, and in weekend Chinese schools set up by local Chinese communities (UCLA Language Profile, 2000).

As in other tone languages, Cantonese makes use of pitch (fundamental frequency-Fo) variations on a syllable in order to represent vary lexicons. It is generally accepted that Cantonese has six contrastive tones in its system (e.g., Hashimoto, 1972; Fok-Chen, 1974; Zee 1995; Bauer & Benedict, 1997; among others). The tonal system contains three level tones (high, mid, low), two rising tones (high and low rising) and a low falling tone. However, not every syllable in Cantonese can carry all six contrastive tones; indeed, only a few syllables can carry the tones. In this present study, the root-word syllables /si/ and /fu/ were used because these two open syllables carry all six lexical tones.

1.1.3 Perception studies in Cantonese

Concerning perception studies of Cantonese tones, Fok-Chen (1974) investigated the performance of native listeners in identifying lexical tone in Cantonese. It was found that the six contrastive tones were not equally vulnerable to perceptual confusion. Tone 1, tone 2, and tone 4 were the most salient among the tones in the perception of Cantonese tones. Confusion was confined to tones with similar patterns. For example, rising tones were confused with other rising tones, and falling and level tones were confused with one another (tone 4 & tone 6).

Vance (1977) attempted to find the range of variation of pitch-time contours for each lexical tone in Cantonese, using synthetic stimuli in order to vary Fo trajectories. A few observations were made: (a) Listeners had more confidence in labeling their responses as tone 1 when high level or high falling contours were given to them; (b) Slope or "gradient" was essential in discriminating between tone 2 and tone 5; (c) Tone 3 was perceived within the entire range; (d) Tone 4 did not receive high identification scores; and (e) Tone 6 was the default tone label when listeners felt unsure in identifying a certain tone.

Gandour (1979, 1981) employed a multidimensional analysis of Fok's confusion data for Cantonese. He found that three underlying dimensions, CONTOUR (i.e., slope), DIRECTION (i.e., rising or falling), and HEIGHT (i.e., average pitch), were important to perception. The first two dimensions proved to be of particular importance for the variance in listener responses. However, in a later perception study (1984) with synthetic tokens, Gandour included Mandarin and Taiwanese listeners, and found that Cantonese listeners attached relatively more importance to the HEIGHT dimension than did Mandarin and Taiwanese listeners.

Ching (1984) used both natural and synthetic tonal tokens in order to assess the ability of Cantonese children to perceive the differences in Fo patterns. The ages of the children ranged from 4 to 10 years. Five types of tokens were used: (a) natural speech, (b) natural speech with vocal tract information essentially absent using a low pass filter at 1 kHz, (c) synthetic speech approximated to natural tokens, (d) synthetic speech with tonal patterns being transposed logarithmically, and (e) synthetic speech with tonal patterns expanded logarithmically by almost an octave. Ching's findings are listed as follows: (i) Recognition ability improves with age; (ii) Children at age six or younger require natural tokens to make confident judgments; (iii) Children older than six years of age are able to make linguistic decisions (identification of tones) based on patterns; (iv) At about age ten, children make confident judgments based on both speech and pattern forms; (v) Children who are good at labeling transposed and expanded tonal patterns, also responded well to natural tokens; (vi) Synthetic tokens with transposed tonal patterns had better responses than the ones with expanded tonal patterns; (vii) The performances of children who make confident judgments in response to the transposed stimuli are compatible with that of the adults; (viii) Tone 4 is best identified in all stimuli types; (ix) Much of the significant confusions in the responses of the children were between tone 3 and tone 6, and between tone 5 and tone 6.

2 Methods

In this study, 30 listeners in three groups (two Cantonese immigrant groups and a native speaker group) identified 48 target words in citation forms produced by four Cantonese language instructors who were all native speakers from Hong Kong. All stimuli had been examined and were selected in a pilot study before they were given to the participants. The stimuli were presented by multiple speakers in order to minimize the familiarity effect of a particular speaker's voice on listeners' responses (Abramson, 1976). Fok's identification paradigm (1974; Yiu & Fok, 1995) was adapted as a method for listeners to assess the target words once they were given a stimulus.

Two aspects of the performances by the two immigrant groups were compared with those of the native Cantonese group. First, listeners' percentages of correct identifications were examined in order to determine which tones are the ones that showed deviations. Second, as in the previous studies (Fok, 1974; Ching 1984; Yiu & Fok, 1995), this study adopted the approach of employing confusion matrices for investigating the tones that were misidentified by the listener groups, as a study dimension. Confusion matrices for individual groups were constructed to illustrate the results of the above comparison. Confusion matrices would show the responses by the listeners in the groups when a target tone was presented.

2.1 Participants

Five Cantonese language instructors (2 males and 3 females), from 36 to 43 years of age, were invited to take part in the experiment (Four of them participated as speakers and one as a listener in a pilot study). They were born and raised in Hong Kong as native Hong Kong Cantonese speakers, and have been teaching Cantonese from 3 to 16 years. These language instructors were teaching Cantonese in Canada at the time of recording. They all reported normal speech and hearing.

There were six listeners who had participated in a pilot study in order to evaluate the selected natural stimuli for the perception test. These listeners were native Hong Kong Cantonese speakers, including one experienced Cantonese instructor, three native Hong Kong Cantonese speakers and two Late immigrant speakers (see below). Their ages ranged from 18 to 42 years, and they had been living in Canada from 2 to 25 years.

With respect to the listeners for the perception test, thirty adult speakers in groups of ten (5 males and 5 females) participated in the experiment. They were the same participants employed in the author's previous study (So, 2000). All of them reported normal hearing. The two Cantonese immigrant groups were separated on the basis of age of arrival (AOA) in Canada. They were labeled as Early Immigrants (EI) and Late Immigrants (LI).

The Early Immigrant (EI) group consisted of seven speakers who were born in Canada, and three additional speakers who had moved to Canada from Hong Kong before the age seven. The participation of the three Hong Kong born speakers may be questioned. Their inclusion, however, can be justified by the fact that their performance on the tonal production was the same as those of the Canadian born Cantonese participants. They were all between 18 and 24 years (M = 19.60 years) at the time of the experiment. The AOAs of the three Hong Kong born speakers were 6, 6.5, 6.5 years, respectively. The mean length of residence was 20.29 years for the seven Canadian born speakers, and 11.17 years for the other three speakers. These speakers had studied Cantonese at typical Chinese schools in Vancouver, for between 6 months and 11 years, except for one speaker, who reported that his Cantonese was learnt only from his parents. As expected, Cantonese was the language used for communication with their parents, while English generally was the primary language used to communicate with their siblings and friends, as well as in their daily activities indicating that they preferred to use English rather than Cantonese in these activities.

The Late Immigrant (LI) group consisted of speakers who were all born in Hong Kong and had moved to Canada during their early adolescence (from 10 to 15 years of age). They had learnt Cantonese at school in Hong Kong, as well as at home. At the time of the experiment, the ages of the speakers ranged from 18 to 22 years (M = 19.60 years). Their mean AOA was 12 years, and their mean length of residence was 7.6 years. They reported that Cantonese was their primary language for communication with their family members and friends, as well as in their daily activities (e.g., watching television). This implies that in general they preferred to use Cantonese rather than English.

The third group was a comparison group consisting of Native Hong Kong Cantonese speakers (NC), who were all born and raised in Hong Kong, and had lived in Canada less than two years at the time of the experiment. Their ages ranged from 18 to 22 years, (M = 19.5 years). Their average AOA was 18.8 years. Their average length of residence was 9.6 months (less than a year). They had a stronger preference to use Cantonese than English.

2.2 Materials

2.2.1 Natural stimuli

Twelve Chinese words from the root-words /si/ and /fu/ were employed in this perception study as target words. The four speakers produced 384 natural two-word phrase speech tokens. The 12 target words were deliberately put in the final words of the phrases (see Appendices 1 & 2). The reason for embedding the target words into phrases was to eliminate the production of an incorrect tone in the target words. The speakers were instructed to read the phrases clearly and at a relatively slow rate, so that the target signals could be segmented from the phrases and so that they would be suitable to be presented as stimuli in the perception test. Half of the recorded signals (192 in number) were randomly chosen, and were digitized into a Macintosh computer. Preliminary examinations of fundamental frequencies and tonal patterns were done using Signalyze software. This screening process was performed in order to ensure that the chosen signals were produced (i) with reasonable Fo heights (level) in relation to the tonal space of the individual speakers, and (ii) with typical Fo contour patterns. It also ascertained that the chosen signals were not produced in an exaggerated way, such as being too long in duration and/ or too high in pitch level.[3] Among the signals, only 96 tokens (48 tokens per root-word set) were selected for the pilot study.

The pilot study was designed to finalize 48 natural stimuli from the above 96 for the perception test. Six listeners[4] were employed to judge the stimuli. Selections were made on the basis of one criterion -- each natural stimulus had to be correctly identified at least 75% of the time by the six listeners, so that the NC listeners would not have too much trouble in recognizing the stimuli in the perception test.

---

[3] The syllable and vowel durational ranges, and the Fo ranges for the natural stimuli from the male and female speakers are listed here: (a) Syllable durations: 417.60 ms - 619.68 ms; (b) Vowel durations: 118.10 - 227.07 ms; (c) Fo range: 83 - 126 Hz (male speakers), and 150-243 Hz (female speakers).

[4] See Section 2.1 for details.

For the perception test, the selected 48 randomized natural stimuli (24 per root-word set) were dubbed through a JVC Double Cassette deck (TD W709) from the computer onto a SONY normal tape with two versions: one was prepared with the /si/ root-word set being presented first and then the /fu/ root-word set, the other version was prepared with the /fu/ root-word set being presented first and then the /si/ set. As in the sub-test condition of previous studies (Fok-Chen, 1974; Wong, 1998), stimuli were presented with multiple-speakers, in order to prevent responses triggered by familiarity with a particular speaker's voice (Abramson, 1976: 2). Before each stimulus was presented, the listener would hear a stimulus number immediately followed by a 0.5 second pure tone[5] indicating the beginning of the stimulus presentation. Next, the stimulus was presented twice, a sequence of successive tokens of the given stimulus separated by a 1.5 second interval. This was followed by 5 seconds of silence, in order to give time to individual listeners to make a decision, after which a second pure tone indicated the end of testing that stimulus. The way that each stimulus was presented is illustrated below:

No. 1,	0.5 sec pure tone, **A**, 1.5 sec. silent time, **A**, 5 sec. silent decision time, 1 sec. pure tone again.
No. 2,	0.5 sec pure tone, **B**, 1.5 sec. silent time, **B**, 5 sec. silent decision time, 1 sec. pure tone again.
	" " " " " "
No. ,	0.5 sec pure tone, **N**, 1.5 sec. silent time, **N**, 5 sec. silent decision time, 1 sec. pure tone again.

Note: the letters, A, B, ... N, representing the stimuli.

2.2.2 Identification paradigm

The present study adapted Fok's identification paradigm (1974, Yiu and Fok 1995) which was designed to access the target words for the listeners after they have perceived a given token. In assessing the subjects' perceptual ability, the pictures and their corresponding Chinese characters were provided simultaneously to the listeners. The advantage of this method is that it provides at least two routes at the semantic level to access the target words (Yiu & Fok, 1995: 84). One is a phoneme-ideograph route (auditory word - written word matching) and the other is a phoneme-semantic route (auditory word - picture matching).

The modified version of the identification paradigm in the present study deliberately provided one more route at the semantic level for listeners to access the target words. This extra route provided the corresponding English glosses (auditory word - visual English glosses matching) simultaneously with the other two routes -- pictures and Chinese characters -- to the listeners. This was important for the EI listeners, because they were either unable to read Chinese words or could only recognize a limited number of Chinese characters. Therefore, listeners were given answer choices with target words presented simultaneously in three routes: (i) the traditional Chinese characters, (ii) the semantic pictures, (iii) and the corresponding English glosses. Since the two root-words, /si/ and /fu/ were presented in one set at a time (with 24 randomized stimuli), listeners were given six target words each time as possible choices. Accordingly, two papers

[5] The pure tone was created at 160 Hz with amplitude of 40 dB by the SoundEdit program.

containing 6 picture-choices corresponding to all six possible target words were prepared for the listeners to identify (see Appendices 1 & 2).

2.3 Procedures

Identification test was conducted on an individual basis in a sound-treated room in the Phonetics Laboratory at Simon Fraser University. The participants were instructed to listen to the given stimuli of two root-word sets, /si/ and /fu/. The two sets of stimuli were played on a National stereo recorder (RX-FM23) through stereo headphones (a Panasonic and a Sony, MDR-CD170) to both the listener and the experimenter (the author). Listeners were given two papers, each of which contained a set of six picture-choices corresponding to all six possible target words (i.e., the identification paradigm; see Appendices 1 & 2). They identified the target words by pointing to the appropriate picture, character, or English gloss once the presentation of a stimulus had been completely presented. The experimenter recorded their responses on an answer sheet.

2.4 Analyses

To explore the response differences among the listener groups systematically, statistical analyses were used focusing on two dimensions: (i) the correct percentages (%) on the identification task among the three groups, and (ii) confusion matrices for the individual groups. Comparisons among the correct percentages of the three listener groups were made in order to establish (i) whether there were differences among the groups in the identification test, and (ii) if differences did exist, which tones were the ones that showed the differences (in correct percentages) among the groups. Confusion matrices, which have been used by Fok (1974), Ching (1984), and Yiu & Fok (1995) as a method to investigate the listeners' tonal confusions, were also constructed to justify the comparison in correct percentages of identifications. The patterns obtained from the confusion matrices would reveal how the listener groups responded to the given target tones (i.e., words). Specifically, it would indicate which pairs of tones showed significant confusions in listener groups.

For this study, (i) the structure of each confusion matrix and (ii) the method to determining whether the observed frequency for tonal misidentifications (i.e., tonal confusions) were statistically significant followed those used in the study of Yiu & Fok (1995). In each confusion matrix, the correct numbers of responses with their converted percentages were indicated in the cells along the diagonal (from the left upper corner to the right lower corner) of the matrix (see Table 3, for example). These cells are shaded and the numerical data are in bold face. For each confusion matrix, it was assumed that if the listeners' responses were entirely random, the obtained frequency in each cell would be equal (Yiu & Fok, 1995). On the other hand, individual cells (i.e., tonal misidentifications) in the confusion matrix would be considered to be statistically significant if each obtained frequency (for the incorrect responses) in the cell was larger than the expected frequency (ibid.). The expected frequency in each cell was calculated as the total number of possible responses divided by 36 cells (6 target tones X 6 response choices), and this corresponds to a 3 % (1/36) chance level (ibid.). According to the approach of Yiu & Fok (1995), for this study, the expected frequency of each cell for the listener groups (Tables 3 - 5) would be 13.33 (480/ 36 = 13.33). If any obtained frequency for a cell was greater than 13.33, confusion between the two tones (one is the target tone and the other is the responded tone) was considered to be statistically significant. The cell is highlighted by presenting the obtained frequency in bold fonts with an asterisk.

3 Results & Discussion

The mean for the correct scores for the identification task of the three groups in each lexical tone are tabulated in Table 1. As seen in the table, no prominent difference in the correct scores between the female and the male speakers from the same group can be observed. Therefore, the statistical analyses presented below only compared the correct percentages (converted from the scores in the table) among the three groups without further subdividing the correct percentages according to their gender.

Table 1. The mean correct scores for the listener groups in the perception test.

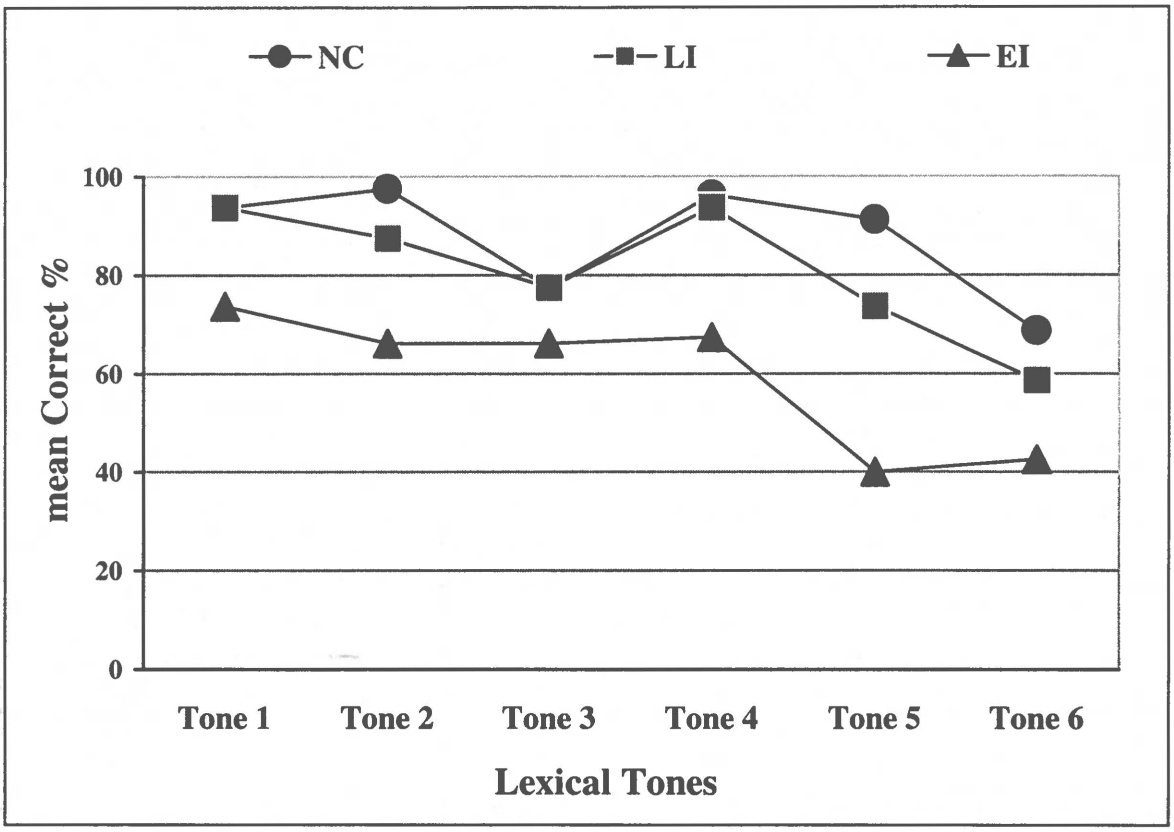
Groups		Tone 1	Tone 2	Tone 3	Tone 4	Tone 5	Tone 6	Total Correct
NC	F	38	40	32	39	37	31	**217 / 240**
	M	37	38	30	38	36	24	**203 / 240**
	GP	**75**	**78**	**62**	**77**	**73**	**55**	**420 / 480**
LI	F	37	36	28	38	26	25	**190 / 240**
	M	38	34	34	37	32	22	**197 / 240**
	GP	**75**	**70**	**62**	**75**	**58**	**47**	**387 / 480**
EI	F	29	28	26	25	13	18	**139 / 240**
	M	30	25	27	29	19	16	**146 / 240**
	GP	**59**	**53**	**53**	**54**	**32**	**34**	**285 / 480**

Notes: the number indicated in the individual cell corresponding to the sex groups (F = female and M = male) was out of 40 stimuli, and those indicated in the individual cell corresponding to the listener groups (GP=whole group) was out of 80 stimuli.

The overall mean for the correct percentages, in relation to the six tones of the three groups, are illustrated in Figure 1. As expected, the grand mean correct percentage[6] of the NC group was the highest among the three groups (M = 87.50%), while that of the EI group was the lowest (M = 59.38%). The grand mean for the LI group (M = 80.83%) was between the two groups. This pattern appears to be valid for almost every lexical tone in Figure 1, with two exceptions: The mean correct percentages of the NC and the LI groups for tone 1 and tone 3, were the same. In order to explore the response differences and the patterns of tonal confusions among the three groups systematically, the mean scores for the six tones of the groups were submitted to comparisons in (i) the identification test and (ii) confusion matrices.

[6] This grand mean was obtained by averaging the mean correct percentages for the six lexical tones.

Figure 1. The correct percentages of the three speaker groups in the perception test.



3.1 Identification test

The differences in the perception test among the three listener groups were evaluated as follows. The mean scores (in correct %) of the six tones for the groups were submitted to a two-way mixed-design ANOVA, with LISTENER GROUPS (3 levels: NC, LI, and EI groups) as a between-subjects factor and TONES as a within-subjects factor. Results revealed that the effects of LISTENER GROUPS and TONES were significant, $F\ (2, 27) = 21.529, p < 0.0001$, and $F\ (5, 135) = 17.918, p < 0.0001$, respectively. The effect of the interaction between LISTENER GROUPS and TONES was significant, $F\ (10, 135) = 2.016, p < 0.05$. The results suggest that the mean scores of the three listener groups were greatly different with regards to certain lexical tones.

Six individual 1-way ANOVAs were carried out in order to compare the group performance for each tone. Results showed that the effect of LISTENER GROUPS was significant for tone 1, tone 2, tone 4, and tone 5, $Fs = (2, 27) = 7.223$, 20.281, 24.130, and 8.270, $ps < 0.01$. Post hoc Tukey tests further revealed that the differences in the mean scores only between the pairs, NC vs. EI and LI vs. EI, were significant for the above four tones ($ps < 0.05$), and the difference between the two NC vs. LI groups for tone 5 was also significant ($p < 0.05$). The findings from the comparison of the mean correct percentages of tonal identifications for the three listener groups are summarized in Table 2.

Table 2. Summary of the results of the identification test

Listener Groups	Lexical tones					
	Tone 1	Tone 2	Tone 3	Tone 4	Tone 5	Tone 6
LI vs. NC					S	
EI vs. NC	S*	S*		S*	S*	

S indicates significant difference in the identification scores from the NC group in the investigated tones.
* indicates significant difference in the identification scores from the LI group at the investigated tones.

When the mean percentages for the tones of the two immigrant groups were compared with those of the NC group, it was found that the mean identification scores of the LI listeners, as expected, were higher than those of the EI listeners. For the LI listeners, only the mean score for tone 5 was significantly lower than that of the NC listeners (73.75% vs. 91.25%). However, for the EI listeners, the mean scores for tone 1 (73.75% vs. 93.75%), tone 2 (66.25% vs. 97.50%), tone 4 (67.50% vs. 96.25%), and tone 5 (40.00% vs. 91.25%) were all significantly lower than those of the NC listeners. In addition, the mean scores for these four tones for the EI listeners were also found to be significantly lower than those of the LI listeners: tone 1 (73.75% vs. 93.75%), tone 2 (66.25% vs. 87.50%), tone 4 (67.50% vs. 93.75%), and tone 5 (40.00% vs. 72.50%). These results suggested that the LI group performed better on the tonal identification test than did the EI group. Moreover, Table 2 indicates that those tones with significantly lower identification scores for the LI and the EI groups (i.e., tone 5 for the LI group, and tone 1, 2, 4, 5 for the EI group) coincided with the tones, which were considered to be reduced in the productions of the two immigrant groups in the previous study (So, 2000).

The reasons for the poor performance of the EI listeners are not fully understood. Possibly, they might have deviant tonal patterns in their mind as suggested by So (2000). They might also have a relatively lower sensitivity to tonal contrast in Cantonese. Subsequently, the EI immigrants have not received proper and sufficient linguistic input for developing their tonal system throughout the years of living in an English-speaking environment. This can be inferred from the fact that their usage of Cantonese in daily activities is apparently less often than the LI and the NC groups.

Besides the above, the data in Table 2 also suggest that several hierarchical patterns can be identified among the tones that were poorly identified by the two groups of immigrant listeners (i.e., tone 1, tone 2, tone 4, and tone 5). From the table, it is clear that the EI listeners have difficulty in identifying all four tones, whereas the LI listeners had problems identifying the low rising tone (tone 5). It can be interpreted as the situation that the immigrant listeners had more difficulty in identifying the contour tones (tone 2, tone 4, and tone 5) than the level tone (tone 1). Among the three contour tones, tone 5 appeared to be the most difficult tone for the listeners in the two immigrant groups, because the mean scores for both the LI and EI groups were significantly lower than that of the NC group. Tone 4, on the other hand, appeared to be the least difficult tone for the immigrant listeners to identify, because only the EI listeners showed confusion between tone 4 and tone 6 (see Table 5). Tone 2 might be put in the middle of the two tones, since the EI listeners showed confusion with this tone, whereas the LI listeners only showed a tendency towards confusion between tone 2 and tone 5 (see Table 4). Finally, among the level tones, tone 1 appeared to be the level tone that the EI listeners had difficulty identifying.

3.2 Confusion matrices

Three separate confusion matrices for the NC, the LI, and the EI groups were constructed (Tables 3 – 5 respectively) in order to find out which tones were confused in each group. Table 3 below is the confusion matrix for the NC group. As seen in the table, although none of the lexical tones received a 100%, relatively high mean correct percentages can be observed for tone 1, tone 2, tone 4, and tone 5, which were all over 90% (see the cells along the diagonal of the matrix). Only tone 3 and tone 6 had relatively low percentages (77.50% and 68.75%, respectively). In examining the incorrect responses, it was found that the tones that were misidentified and the target tones mainly belonged to the same type of tones (i.e., a level tone was misidentified as another level tone). For example, tone 1 (high level) was misidentified as tone 3 (mid level), and tone 2 (high rising) was misidentified as tone 5 (low rising). There were, however, some exceptional cases, such as tone 4 (low falling) and tone 6 (low level): tone 6 was mainly misidentified as tone 4.[7] The matrix also indicates that the NC listeners misidentified tone 3 as tone 6 (obtained frequency = 16), and tone 6 as tone 4 (obtained frequency = 18) at a significant level. In other words, confusions between tone 3 and tone 6, and between tone 6 and tone 4 were found.

Table 3. Confusion matrix for the NC group

Target Tone	Response Tone						Row Total
	Tone 1	Tone 2	Tone 3	Tone 4	Tone 5	Tone 6	
Tone 1	75 93.75%		5				80
Tone 2		78 97.50%			2		80
Tone 3	1		62 77.50%	1		16*	80
Tone 4				77 96.25%		3	80
Tone 5		7			73 91.25%		80
Tone 6			7	18*		55 68.75%	80
Column Total	76	85	75	96	74	74	480

The expected frequency in each cell is 480/36 = 13.33. The shaded cells indicate correct identification with the corresponding percentages.
* indicates that the obtained frequency for the incorrect responses was considered to be statistically significant, because it was greater than the expected frequency.

[7] A discussion of this issue is given in Section 3.2.2 of this paper.

Table 4 presents the confusion matrix for the LI group. The mean correct percentages for tone 1, tone 3 and tone 4 (93.75%, 77.50%, and 93.75%) were similar to those of the NC group (93.75%, 77.50%, and 96.25%, respectively). The mean correct percentages for the other tones, tone 2, tone 5, and tone 6 (87.50%, 73.75%, and 58.75%) were lower than the correct % for the corresponding tones of the NC group (97.50%, 91.25%, and 68.75%, respectively). There was a slightly higher frequency of misidentifications and greater error variations among the tones that were misidentified by listeners in the LI group than by those in the NC group. For example, in addition to the incorrect responses for tone 5 (low rising) being identified as tone 2 (high rising), tone 5 was also identified as tone 3 (mid level). For tone 6 (low level), it was identified as tone 1 (high level), tone 3 (mid level), tone 4 (low falling), and tone 5 (low rising). The matrix reveals that the listeners in the LI group significantly misidentified tone 3 as tone 6 (obtained frequency = 15), and tone 5 as tone 2 (obtained frequency = 20). Moreover, these LI listeners also tended to frequently identify tone 2 as tone 5 (obtained frequency = 10), and tone 6 as tone 3 (obtained frequency = 12). Although the obtained frequencies for these two cases failed to reach significance (expected frequency = 13.33), the relatively high frequencies of misidentifications have to be acknowledged. As a result, on the basis of the above observations, the LI listeners exhibited confusions between tone 3 and tone 6, and between tone 5 and tone 2.

Table 4. Confusion matrix for the LI group

Target Tone	Response Tone						Row Total
	Tone 1	Tone 2	Tone 3	Tone 4	Tone 5	Tone 6	
Tone 1	**75 93.75%**		4			1	**80**
Tone 2		**70 87.50%**			10		**80**
Tone 3	1		**62 77.50%**		2	**15***	**80**
Tone 4			1	**75 93.75%**		4	**80**
Tone 5		**20***	2		**58 72.50%**		**80**
Tone 6	8		12	8	5	**47 58.75%**	**80**
Column Total	**84**	**90**	**81**	**83**	**75**	**67**	**480**

The expected frequency in each cell is 480/36 = 13.33. The shaded cells indicate correct identification with the corresponding percentages.

* indicates that the obtained frequency for the incorrect responses was considered to be statistically significant, because it was greater than the expected frequency.

Lastly, Table 5 below shows the confusion matrix for the EI group. The mean correct percentages of the six lexical tones for the EI group was lower than those of the NC and the LI groups (see Tables 3 and 4 for comparison). Unlike the NC listeners, who exhibited less confusion with regards to the different tones, the responses for the EI listeners demonstrated misidentifications for all tones. Among the six tones, tone 1 was best identified with 73.75% accuracy. The matrix also shows that the target tones, tone 2, tone 4, tone 5, and tone 6 were significantly misidentified as tone 5, tone 6, tone 2, and tone 4, respectively (their obtained frequencies = 17, 16, 34, and 17, respectively). For tone 1, no specific confusion between tone 1 and any other tone was observed to be statistically significant because those incorrect response-choices for tone 1 were spread over the other five tones. This kind of response-pattern was different from the pattern of the poorly identified tones (i.e., tone 2, tone 4, tone 5, and tone 6), in which confusion between the target tones and particular response tones were observed to be significant. It is also notable that the EI listeners frequently identified tone 3 as tone 6 (obtained frequency = 12), and tone 6 as tone 3 (obtained frequency =11), although their obtained frequencies failed to reach the significant level. The results imply they experience confusions with all tones.

Table 5. Confusion matrix for the EI group

Target Tone	Response Tone						Row Total
	Tone 1	Tone 2	Tone 3	Tone 4	Tone 5	Tone 6	
Tone 1	59 73.75%	2	4	4	7	4	80
Tone 2	3	53 66.25%	4	1	17*	2	80
Tone 3	5	1	53 66.25%	4	5	12	80
Tone 4	2	3	1	54 67.50%	4	16*	80
Tone 5	2	34*	5	5	32 40.00%	2	80
Tone 6	6	4	11	17*	8	34 42.50%	80
Column Total	77	97	78	85	73	70	480

The expected frequency in each cell is 480/36 = 13.33. The shaded cells indicate correct identification with the corresponding percentages.
* indicates that the obtained frequency for the incorrect responses was considered to be statistically significant, because it was greater than the expected frequency.

3.2.1 Group differences

The tonal confusion patterns of the three listener groups are summarized in Table 6. The table only lists (i) confusions found to be statistically significant in the group, and (ii) confusing tone-pairs with prominently high frequencies of misidentifications, i.e., obtained frequency ≥ 10 (marked with **H**).

As seen in Table 6 (below), response-choices of the groups conform to the observation by Fok-Chen (1974) and Ching (1984), according to which confusion was confined to tones with similar patterns. For example, a high rising tone (tone 2) was replaced by a low rising tone (tone 5); a low level tone (tone 6) was misidentified as a low falling tone (tone 4). Thus, regardless of the number of confusion tone-pairs showing for the two immigrant groups, the choices for the incorrect responses were confined to tones with similar patterns. This implies that the selection of choices by listeners of the two immigrant groups was not by chance, that is, they appeared to have some knowledge of the lexical tones, such as, contour patterns and relative height.

Table 6. Summary of the tonal confusions of the three listener groups

Target Tones	Listener Groups		
	NC	LI	EI
Tone 1			
Tone 2		H: Tone 2 --> Tone 5	Tone 2 --> Tone 5
Tone 3	Tone 3 --> Tone 6	Tone 3 --> Tone 6	H: Tone 3 --> Tone 6
Tone 4			Tone 4 --> Tone 6
Tone 5		Tone 5 --> Tone 2	Tone 5 --> Tone 2
Tone 6	Tone 6 --> Tone 4		Tone 6 --> Tone 4
" "		H: Tone 6 --> Tone 3	H: Tone 6 --> Tone 3

In addition, the confusion patterns in Table 6 show that not all three listener groups supported the observations by Fok-Chen (1974), which stated that (i) the six lexical tones are not equally vulnerable to perception confusion, and that (ii) tone 1, tone 2, and tone 4 are the most salient in the perception of Cantonese tones. The results of the present study also show that, for the NC group, not only tone 1, tone 2, and tone 4, but also tone 5 appear to be equally salient among the tones: the percentages for the correct identifications were all over 90% (see Table 3), whereas tone 3 and tone 6 caused considerable confusions for the participants.[8] For the LI group, tone 1 and tone 4 appear to be the most salient tones, the correct scores being over 90% (see Table 4). In addition to confusing the two level tones (tone 3 and tone 6), results for the LI group indicate that the LI group experienced confusions with the two rising tones, especially tone 5. Results for the EI group showed that a variety of confusions arose among the six lexical tones. It should be mentioned though, that responses for tone 1 were somewhat better than the other five tones (see Table 5). Consequently, no single tone can be identified as being salient for the EI group. Since the confusions of the NC group match with those of the native speakers in the study of Fok-Chen (1974), this implies that the confusions observed in the NC group are typical, and that the confusions of the EI and the LI groups should be considered as patterns that deviate from the NC group in various degrees.

3.2.2 Level tone confusions

3.2.2.1 Tone 3 and tone 6

In examining the results shown in Table 6, it is noteworthy that listeners in all three groups have obvious confusions in the two non-high level tones, tone 3 and tone 6. NC listeners significantly misidentify tone 3 as tone 6, and tone 6 as tone 4. The LI and EI listeners also show similar confusions in the target tones, tone 3 and tone 6. In fact, the

[8] Tone 3 and tone 6 for the three listener groups will be discussed in Section 3.2.2.1.

tendency to confuse the level tones has been reported in previous literature. Fok-Chen (1974) reported that her listeners confused the three level tones (tone 1, tone 3 and tone 6) in her study. Vance (1977) found that tone 3 was being perceived within the entire (Fo) range of the synthetic speech. Confusion between the two tones (tone 3 and tone 6) was not confined to adult native speakers of Cantonese: Ching (1984) reported that significant confusion between tone 3 and tone 6 also occurs in the responses of young Cantonese children. Wong (1998) observed that the level tones were poorly identified. Abramson's study (1978) on Thai tones indicated that "a considerable overlap" exists in the perception of the level tones. As in the above studies, the present study also found that listener groups confused the three level tones, such as misidentifications of tone 1 as tone 3, tone 3 as tone 6, and tone 6 as tone 3 or tone 4.

According to Gandour (1994: 3120), Fo height is the dominant cue for distinguishing between tones with similar contours. In this case, since level tones in general do not involve much Fo movement, (except the acceptable falling pattern mentioned in literature, such as Abramson (1997) and Bauer & Benedict (1997)), the relative Fo height becomes a crucial factor. Therefore, the observed tendency to confuse tone 3 and tone 6, but not tone 1 in this study might be attributed to the following three factors.

First, the fact that confusion was found particularly with tone 3 and tone 6, but not with tone 1 may relate to the fact that these two tones are closer to an individual's normal speaking voice. Eitel (1947: cited in Fok-Chen, 1974: 18) has suggested that individual speaking voice is somewhere near tone 3. Similarly, Chao (1968) has pointed out that it is normally centered near the lowest part of the speaker's voice range. Thus, because these two level tones (tone 3 and tone 6) are closer to the speaking voice of the speakers, listeners may find the two tones more confusing than tone 1. Moreover, association with higher pitch may cause tone 1 to become more salient than the other two tones (Fok-Chen, 1974; Vance, 1977; Ching 1984). Therefore, it is less likely to be confused with any other tone.

Second, confusions might also relate to the relative Fo (pitch) distance (Nooteboom, 1997) among the three level tones of Cantonese. If the Fo distance between two tones is shorter, then more confusion is expected in identifying the two level tones (tone 3 and tone 6). Jones and Woo (1912: xiv-xv)[9] have suggested that in normal speech situations tone 1 is approximately three semitones above tone 3, which is in turn two semitones higher than tone 6. Thus, it is conceivable that the Fo distance between tone 3 and tone 6 is shorter than the distances between tone 1 and tone 3, and between tone 1 and tone 6. Accordingly, since the Fo distance between tone 3 and tone 6 is shorter, it is expected that more confusions are involved in identifying these two tones. Although the size of the scales may vary from person to person, relative relationship among the tones will remain (ibid.). This might be the reason why the correct identifications for tone 1 in the NC group, as well as in the LI group, were remarkably high (93.75% in both groups), whereas the correct identifications for tone 3 and tone 6 were considerably lower. Accordingly, if the distance between the two tones is greater, the more salient their difference becomes in perception.

Finally, using multiple speakers to present the isolated stimuli may increase the difficulty for listeners to identify tone 3 and tone 6. The method of using multiple

[9] Chao (1947:25) also adopted the musical notes in his book.

speakers to present stimuli has been used as a condition of the studies by Fok-Chen (1974) and Wong (1998). Fok-Chen (1974) found that native Cantonese listeners in general, showed more misidentifications when multiple speakers were employed, than if a single speaker was employed (either a female or a male). Moreover, incorrect responses were more frequently found in level tones than in contour tones in the condition employing multiple speakers. Similarly, in his study, Wong (1998) examined the perception of the three level tones in Cantonese (tone 1, tone 3, and tone 6) by native listeners in varying conditions. One of the conditions was employing multiple speakers to read stimuli in citation forms. The results revealed that the native listeners had great difficulty in identifying the three level tones in this condition. The mean correct percentage for identifying the three tones was 49.83%. Wong suspected that his subjects identified the tones almost by chance. For this study, using multiple speakers to present the isolated stimuli appeared to have more influence on tone 3 and tone 6, because native listeners (in the NC group) were able to identify the lexical tones with a remarkably high correct percentage of identification. Most of the lexical tones were correctly identified more than 90% of the time, except for tone 3 (77.50%) and tone 6 (68.75%).[10] The overall results appear to suggest that the confusing effects of using multiple speakers to present the isolated stimuli had more impact on tone 3 and tone 6. These patterns were quite similar to the results of Fok-Chen (1974).

The remaining question is why tone 3 and tone 6 are more readily confused in the perception test employing multiple speakers. Under a normal situation, listeners will use the perceptual skill, normalization, to infer the information of the tonal patterns from given stimuli (Fourcin, 1978; Ching, 1984). However, in this case, when a set of stimuli from different speakers were presented successively, normalizing the stimuli becomes difficult. Every speaker has his or her unique pitch range (tonal space); it is possible that a high tone for one speaker has the same Fo values as a low tone for another speaker. Conversely, contour tones can be perceived through other perceptual cues, such as the contour (shape), slope, and direction (rising of falling) (Gandour, 1979, 1981, 1984 in Cantonese) -- facilitating the identification of the target tones.

3.2.2.2 Tone 6 and tone 4

In addition to the points discussed above, this study also observed that tone 6 was frequently misidentified as tone 4 by both the NC and the EI groups (see Table 6). However, re-examining individual listeners' responses, it was found that 16 out of 18 of the errors in the NC group (see Table 3) were actually made by three listeners (2 males and 1 female). For the two male listeners, one made 6 errors out of the given 8 stimuli for tone 6, and the other male listener made 5 errors out of 8. In fact, their scores constitute the total number of errors in the group of NC male listeners. The other five errors were made by one of the NC female listeners. This suggests that some of the native listeners are unable to distinguish between tone 6 and tone 4. Unlike the case of the NC group, the

[10] Although the correct % of tone 3 and tone 6 were relatively lower than the other tones, the correct identifications of the tones (in %) in this study were still considerably higher than in Wong's findings (1998). The better performance in this study might be attributed to the fact that the stimuli had undergone screening processes -- signal selections and the pilot study -- in the initial stages of the experiment, before the perception test was conducted.

errors made by the listeners in the EI group were distributed evenly among all members of the group.

Confusion between tone 4 and tone 6 might be attributed to two factors. First, tone 4 and tone 6 are similar to each other, because they resemble each other in pitch level, quality of voice, and are relative shorter than the others in duration (Eitel, 1947: cited in Fok-Chen, 1974: 18). Second, because the tonal pattern for tone 6 appears to be associated with a light falling pattern (So, 2000), this may facilitate the two tones being more like each other in the perception test. Consequently, tone 6 and tone 4 are frequently indistinguishable (Eitel, 1947: cited in Fok-Chen, 1974: 18; Hashimoto, 1972; Bauer & Benedict, 1997).

4 Conclusion

The purpose of the present study was to find out whether perceptual differences in identifying Cantonese lexical tones exist between two immigrant groups (Early and Late) and a comparison group of native Cantonese speakers. As expected, the results revealed that the identification scores of the LI group were comparable to the NC group (with the exception of the rising tones), whereas the perception scores for the EI groups showed more deviations from those of the NC group (at tone 1, tone 2, tone 4, tone 5). This implies that the tonal patterns of the EI listeners deviate from those of the NC listeners to a greater degree than the ones of the LI listeners. Moreover, the patterns in the confusion matrices confirmed that the choices of tones, which were confused in the selection process, were confined to tones with similar patterns. This implied that even listeners in the EI group who made errors with all tones still had some knowledge about the six lexical tones (e.g., tonal patterns). Finally, the results also revealed that hierarchical patterns could be identified from those tones that were poorly identified in the test. In general, immigrant listeners made more errors in the contour tones (i.e., tone 2, tone 4, and tone 5) than in the level tone (i.e., tone 1). They appeared to have much more difficulty in identifying the two rising tones (tone 2 and tone 5) than the falling tone (tone 4), and tone 5 was the one that is highly misidentified by the immigrant listeners. Of the three level tones, tone 1 appeared to be the most difficulty tone to identify. The findings of the present study (including the hierarchical patterns for tonal misidentifications) coincided with the reduced tonal patterns of the immigrant groups observed in So (2000), suggesting that the tonal system of EI speakers deviates from the one of the NC speakers to a greater degree than that of the LI speakers.

References

Abramson, A. (1976). Thai tones as a reference system. In T. Gething, H. J. Kullavanijaya (Eds.), *Tai Linguistics in Honor of Fang Kuei-Li*. Bangkok: Chulalongkorn University Press.

Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech. 21* (4), 319-325.

Abramson, A. (1997). The Thai tonal space. In A. S. Abramson (Ed.), *Southeast Asian Linguistic Studies in Honour of Vichin Panupong*. Bangkok: Chulalongkorn University Press. 1-10.

Arendrup, B. (1994). Chinese. In R. E. Asher, & J. M. Y. Simpson (Eds.), *The Encyclopedia of Language and Linguistics*. England: Pergamon Press. 516-524.

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

Chao, Y. (1947). *Cantonese Primer*. New York: Greenwood Press.

Chao, Y. (1968). *A Grammar of Spoken Chinese*. University of California: Berkeley & Los Angeles.

Ching, Y. (1984). Lexical tone pattern learning in Cantonese children. *Language Learning and Communication. 3* (3), 243-414.

Eitel, E. (1947). *A Chinese-English Dictionary in the Cantonese Dialect.*

Fok-Chen, Y. (1974). *A Perceptual Study of Tones in Cantonese*. Occasional Papers and Monographs, 18, (Center of the Asian Studies, University of Hong Kong.

Fourcin, A. (1978). Acoustic patterns and speech acquisition. In N. Waterson & C. Snow (Eds.), *The Development of Communication*. London: John Wiley.

Gandour, J. (1979). Perceptual dimensions of Cantonese tones: a multidimensional scaling reanalysis of Fok's tone confusion data. *Southeast Asian Linguistic Studies (4),* Canberra: Department of Linguistics, Research School of Pacific Studies, Australia National University. 415-429.

Gandour, J. (1981). Perceptual dimensions of tone: evidence from Cantonese. *Journal of Chinese Linguistics, 9* (1), 20-36.

Gandour, J. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics, 12* (2), 235-260c.

Gandour, J. (1994). The phonetics of tone. In R. E. Asher & J. M. Y. Simpson (Eds.), *The Encyclopedia of Language and Linguistics*. Vol. 6, Oxford: Pergamon, 3116-3123.

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

Jones, D. & Woo, K. (1912). *A Cantonese Phonetic Reader*, University of London Press.

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

Light, T. (1977). Clairetalk: A Cantonese-speaking child's confrontation with bilingualism. *Journal of Chinese Linguistics, 5*, 261-275.

Major, R. (1992). Losing English as a first language. *The Modern Language Journal, 76*, 190-208.

Major, R. (1997). L2 acquisition, L1 loss, and the critical period hypothesis. In James, A. and Leather, J. (eds.). *Studies-Language Speech: Structure and Process.* Germany, Berlin: Mouton de Gruyter, 147-159.

Nooteboom, S. (1997). The prosody of speech melody and rhythm. In W. J. Hardcastle and J. Laver (Eds.), *The Handbook of Phonetic Sciences*. Cambridge: Blackwell Publishers Ltd., 645.

Sancier, M. & Fowler, C. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*, 421-436.

So, C. (2000). *The Reduction of Tonal Space in Adult Canadian Cantonese Speakers.* The Twelfth North American Conference on Chinese Linguistics (NACCL-12), San Diego, CA.

Tang, J. & Maidment, J. (1996). Prosodic aspects of child-directed speech in Cantonese. In V. Hazan, S. Rosen, & M. Holland (Eds.), *Speech Hearing and Language* (an online version), Department of Phonetics and Linguistics, University College London. Retrieved February 6, 1999 from the World Wide Web: http:// ww.phon.ucl.ac.uk/home/shl9/tang/tangjm

Vance, T. (1977). An experimental investigation of tone and intonation. *Phonetica, 33,* .368-392.

UCLA Language Materials: *Cantonese Language Profile.* (2000, January 11). California, LA: UCLA. Retrieved January 11, 2000 from World Wide Web: http://www.Imp.ucla.edu/profiles/profc01.htm

Weinreich, U. (1953). *Language in Contact.* New York: Linguistic Circle of New York.

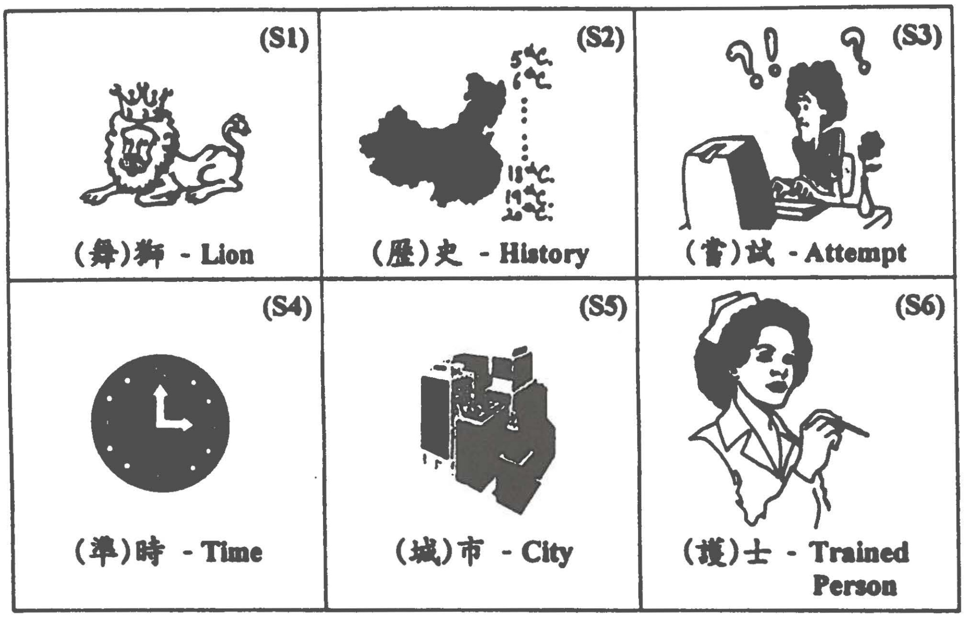
Wong, P. (1998). *Speaker Normalization in the Perception of Cantonese Level Tones: Effect of Context-Target Pitch Distance.* MA Thesis. (an online version). Retrieved January 3, 2000 from World Wide Web: <http://128.83.207.190/patrick/thesis.doc>.

Yiu, E. & Fok, A. (1995). Lexical tone disruption in Cantonese aphasic speakers. *Clinical Linguistics & Phonetics, 9*, 79-92.

Zee, E. (1995). Temporal organization of syllable production in Cantonese Chinese. *Proceedings of the XIIIth International Congress of Phonetic Sciences.* Stockholm, Sweden, 250-3.

Appendix 1

/si/ Identification Paradigm



Appendix 2

/fu/ Identification Paradigm