

<b>Title</b>	<b>Effect of intonation on Cantonese lexical tones</b>
<b>Author(s)</b>	<b>Ma, JKY; Ciocca, V; Whitehill, TL</b>
<b>Citation</b>	<b>Journal Of The Acoustical Society Of America, 2006, v. 120 n. 6, p. 3978-3987</b>
<b>Issued Date</b>	<b>2006</b>
<b>URL</b>	<b><a href="http://hdl.handle.net/10722/45327">http://hdl.handle.net/10722/45327</a></b>
<b>Rights</b>	<b>Creative Commons: Attribution 3.0 Hong Kong License</b>

# Effect of intonation on Cantonese lexical tones<sup>a)</sup>

Joan K-Y Ma,<sup>b)</sup> Valter Ciocca, and Tara L. Whitehill

*Division of Speech and Hearing Sciences, University of Hong Kong, 5/F, 34 Hospital Road, Hong Kong Special Administrative Region, People's Republic of China*

(Received 8 October 2005; revised 5 September 2006; accepted 20 September 2006)

In tonal languages, there are potential conflicts between the F0-based changes due to the coexistence of intonation and lexical tones. In the present study, the interaction of tone and intonation in Cantonese was examined using acoustic and perceptual analyses. The acoustic patterns of tones at the initial, medial, and final positions of questions and statements were measured. Results showed that intonation affects both the F0 level and contour, while the duration of the six tones varied as a function of positions within intonation contexts. All six tones at the final position of questions showed rising F0 contour, regardless of their canonical form. Listeners were overall more accurate in the identification of tones presented within the original carrier than of the same tones in isolation. However, a large proportion of tones 33, 21, 23, and 22 at the final position of questions were misperceived as tone 25 both within the original carrier and as isolated words. These results suggest that although the intonation context provided cues for correct tone identification, the intonation-induced changes in F0 contour cannot always be perceptually compensated for, resulting in some erroneous perception of the identity of Cantonese tone. © 2006 Acoustical Society of America. [DOI: 10.1121/1.2363927]

PACS number(s): 43.70.Kv, 43.71.Hw [AL]

Pages: 3978–3987

## I. INTRODUCTION

Cantonese is a tone language in which contrasts in tone mark a difference in lexical meaning. Cantonese tones are characterized by tone level (high, mid, and low) and tone contour (rising, falling, and level) (Fok-Chan, 1974). The six basic tones in Cantonese are high level (55), high rising (25), mid level (33), low falling (21), low rising (23), and low level (22). The numerical values in parenthesis describe the level of pitch at the beginning and the end point of the tone (Chao, 1947). Intonation is a universal feature shared by languages of different origins, and it plays an important role in conveying both linguistic and paralinguistic meanings in communication. The intonation of a sentence is mainly carried by fundamental frequency (F0) variations. The use of intonation in a tone language could be potentially confusing as F0 patterns are used to mark both tone at syllabic level and intonation at sentential level. This poses an interesting research question about the influence of intonation on the identity of lexical tones. The identification of lexical tones might be affected when the direction of F0 movement in a tone does not coincide with the F0 changes of intonation, such as when a low-falling tone is placed at the final position of question, which has a rising F0 contour.

The contrast between questions and statements is frequently used in studying the effect of intonation on F0 pat-

terns of tone. The intonation patterns of Cantonese are similar to the reported intonation patterns of other world languages (Bauer and Benedict, 1997; Fok-Chan, 1974). It is commonly agreed that statements are characterized by a falling intonation, while questions have a rising intonation contour (Ohala, 1983). However, researchers disagree on how exactly the difference between questions and statements should be characterized. As summarized by Liu and Xu (2005), several features have been proposed to describe the rising intonation of questions: boundary tones (Pierrehumbert, 1980; Ladd, 1996; Lin, 2004), F0 increase of the entire sentence (Ho, 1977; Yuan, Shih, and Kochanski, 2002) and superposition of a baseline on the phrase curve (Thorsen, 1980; Liu and Xu, 2005). For Cantonese, Gu, Hirose, and Fujisaki (2005) and Ma, Ciocca, and Whitehill (2006b) analyzed the F0 patterns of questions and statements using the Fujisaki model (Fujisaki and Hirose, 1984). They observed that questions were marked by positive tone command values with amplitude larger than usual towards the end of the utterance for all six tones. Ma *et al.* (2006b) also found that there was an increase in the baseline frequency for questions. These two studies suggested that the question and statement contrasts in Cantonese are marked by both global and local F0 changes.

The effect of intonation on the acoustic properties of tone has been a subject of research in various Chinese tone languages such as Chengtu (Chang, 1958), Cantonese (Fok-Chan, 1974; Lee, 2004; Vance, 1976), and Mandarin (Ho, 1977; Rumjancev, cited in Lyovin, 1978; Shen, 1989; Lin, 2004). It is generally agreed that intonation is likely to modify the F0 patterns of tones, but the specific F0 patterns vary across tone languages. Two general patterns have been identified. At sentence final position, intonation-induced F0

<sup>a)</sup>Portions of this work were presented in “The effects of intonation patterns on lexical tone production in Cantonese,” Proceedings of International Symposium on Tonal Aspects of Language: Emphasis on Tone Languages, Beijing, People's Republic of China, March 2004, and “The effect of intonation on perception of Cantonese lexical tones,” Proceedings of International Conference on Spoken Language Processing, Jeju, Korea, October 2004.

<sup>b)</sup>Author to whom correspondence should be addressed; electronic mail: joanma@hkusua.hku.hk

perturbations of lexical tones mostly affect the F0 level, but not the F0 contour (Chang, 1958; Ho, 1977; Lin, 2004; Rumjancev, cited in Lyovin, 1978; Shen, 1989). Specifically, when the direction of F0 movement of tone and intonation are the same (e.g., a rising tone at the end of a rising intonation), the F0 level of the tone will increase further. However, when the direction of F0 movement of tone and intonation do not coincide, the F0 value of the tone would be neutralized such that, for example, the F0 of a falling tone would not fall as low at the end of a rising intonation. The second general pattern is that, in addition to the level of the tone being modified, the tone contour may also deviate from its canonical form due to the effect of sentence intonation (Fok-Chan, 1974; Lee, 2004). That is, while a rising tone at the end of a rising intonation gives rise to a higher F0 level, the contour of a falling/level tone would be changed to rising when it occurs at the end of a rising intonation.

Different results have also been reported on the perceptual effect of intonation on tone. Connell, Hogan, and Rozsypal (1983) claimed that the perturbation in F0 patterns of tones brought about by intonation has little effect on listeners' perception in Mandarin. In contrast, Fok-Chan (1974) reported that listeners' tone perception was poorer when presented with Cantonese stimuli produced in interrogative manner. She suggested that the modification in F0 patterns in tone as a result of intonation was likely to affect listeners' perception.

Although F0 is the primary acoustic correlate for tone, the significance of duration as a perceptual cue to tone identity has also been explored (Blicher, Diehl, and Cohen, 1990; Liu and Samuel, 2004; Tseng, Massaro, and Cohen, 1986). Tseng *et al.* (1986) demonstrated that duration might be used as a secondary cue for the perception of Mandarin tones when the F0 information is ambiguous. Ho (1977) observed that duration was likely to be a function of word position rather than intonation, except at word final position where tones produced in interrogative were found to have a longer duration than those in declarative in Mandarin. Vance (1976) pointed out that vowel duration was not significant for tone perception in Cantonese as similar vowel length was observed for all tones. However, Lee (2004) observed that the duration patterns among the six tones were not maintained at the final position of questions. She reported that tones 33, 21, and 22 produced at the final position of questions were longer than when produced in statements in Cantonese, while tones 55, 25, and 23 showed minimal contrast at the final position of questions and statements. However, she did not report a quantitative analysis of these differences. Therefore, it is not clear whether the duration contrast between the six tones at the final position of questions can provide an additional cue to listeners for tone identification.

The above discussion revealed different findings on the effects of intonation on both F0 patterns and duration of lexical tone in different tone languages. In order to explore the effects of intonation on the production and perception of lexical tones in Cantonese, acoustic and perceptual analyses were performed on the same set of Cantonese tones produced within statements and questions in the current study.

## II. EXPERIMENT 1-ACOUSTIC ANALYSIS

The objective of experiment 1 was to investigate the F0 changes produced by different intonation patterns on Cantonese lexical tones. Comparison was made between tones at initial, medial and final positions within an utterance. Acoustic analysis was performed to measure the F0 and duration patterns of the target words.

### A. Method

#### 1. Speakers

Twenty native Cantonese speakers were recruited (ten males and ten females, aged 19 to 25 years). They were all undergraduates or graduates of Hong Kong universities. Nineteen speakers were born and raised in Hong Kong. The remaining speaker was born in the USA, but he was raised in Hong Kong since he was two months old. Cantonese was the native language, and English the second language, of all speakers. None of the speakers had a reported history of speech problems and all had passed a hearing screening ( $\leq 20$  dBHL at 250, 500, 1000, 2000, and 4000 Hz).

#### 2. Materials

Three sets of target words were derived from the roots /si/, /ji/, and /jvu/. Each set consisted of six words that contrasted only in tone, giving a total of 18 target words. These words were embedded in three different positions within the following sentences: (a) initial position (/X tsi<sub>22</sub> hou<sub>25</sub> lan<sub>21</sub> sɛ<sub>25</sub>/ "X is difficult to write"), (b) medial position (/sɛ<sub>25</sub> kɔ<sub>33</sub> X tsi<sub>22</sub> sin<sub>55</sub>/ "Write the X word first"), and (c) final position (/lei<sub>55</sub> kɔ<sub>33</sub> tsi<sub>22</sub> hɛi<sub>22</sub> X/ "This word is X"). Two types of intonation were studied in the present experiment—question and statement. Each of the sentences could be produced as either a statement or a question by modifying only the intonation marking. Thus, each speaker produced a total of 108 different stimuli (18 target words by three different positions by two intonation patterns).

#### 3. Procedures

Recording was carried out in a sound-attenuated room (IAC single-wall booth), with a Sony TCD-D3 DAT recorder and a Bruel and Kjaer (4003) low-noise unidirectional microphone. A 10 cm mouth-to-microphone distance was maintained for the first 17 speakers. The mouth-to-microphone distance was increased to 15 cm for the last three speakers to prevent clipping of the recordings, as the sponge attached to the microphone was removed due to hygienic reasons during the outbreak of severe acute respiratory syndrome in Hong Kong.

In order to obtain naturally produced speech samples, speakers were engaged in a dialogue in which the first author initiated each exchange and speakers answered with one of the sentences. In each trial, the dialogues were presented visually on the screen of a G4 Apple Macintosh computer running a HyperCard (Apple™) custom program. The sequence of dialog presentation was randomized across sub-

jects. Each production was monitored by the first author, who is a qualified speech and language pathologist, to ensure that the correct tone was produced.

After the recording, each sentence was low-pass filtered at 22 kHz, digitized at sampling rate of 44.1 kHz and stored onto an Apple PowerMacintosh 7100 computer as a separate file, using a DigiDesign Audiomedia II DSP card.

#### 4. Data analysis

Acoustic analysis was performed to measure the F0 values and the duration of each target word, using the Praat software (Version 4.0.46, Boersma and Weenink, 2003). The voiced segment of each word was identified visually from a wideband spectrogram and an amplitude waveform display. F0 was estimated at nine evenly spaced time points from the beginning to the end of the voiced segment of the word using an autocorrelation algorithm. Five time points were chosen for subsequent analysis (0%, 25%, 50%, 75%, and 100% of the total duration). Manual measurements from the amplitude waveform were used when the software produced F0 estimates that were largely different from those of adjacent time points. Out of a total of 2160 speech samples, 128 samples (5.93%) required manual measurement at one or more of the five time points used for analysis. Among these 128 samples, eight samples showed diplophonic features with F0 varying between high and low F0 values in consecutive cycles. For these samples, F0 estimates were obtained by averaging the manually estimated F0 of two consecutive cycles of the waveform. The duration of each target word was measured by calculating the difference in time between the 0% and 100% time points of the voiced segment. The average F0 of each tone at each time point and the average duration of each tone were then calculated for each position and intonation for each speaker. Group averages for male and female speakers were also calculated for later analysis.

Intra- and inter-rater reliabilities were calculated by repeating the analysis for two speakers (10% of the data) by the first author and a second rater, who is experienced in acoustic analysis. Pearson's correlation was used to calculate reliability. Intra-rater reliability was 0.97 ( $p < 0.001$ ) and inter-rater reliability was 0.98 ( $p < 0.001$ ).

### B. Results

#### 1. Fundamental frequency

The mean F0 values for questions and statements for male and female speakers in each tone at all nine time points are displayed in Figs. 1(a)–1(f). The F0 patterns of the six tones were similar for all targets in statements and at the initial and medial positions of questions. All the three level tones (tones 55, 33, and 22) had slightly falling contours, and the relative F0 distance between tones 55 and 33 was greater than between tones 33 and 22. Tones 25 and 23 had a rising tone contour; tone 25 had higher final F0 level than tone 23. Tone 21 started at about the same level as tone 22, and had a falling contour to the lowest F0 value of the six tones. At the final position of questions, both male and female speakers showed rising F0 contour for all six tones. Specifically, tones 55 and 33 showed an increase in F0 from 0% to 100% of the

total duration; while tones 25, 21, 23, and 22 had a slight dip in F0 from the 0% to the 25% time points, followed by an increase in F0 from 25% to 100%. The F0 patterns for tones 25, 21, 23, and 22 were overlapping, with similar F0 levels; tones 55 (highest overall F0) and 33 (slightly lower F0 values) were distinct from the rest.

The mean F0 value for each tone at each intonation and position within each time point was calculated for each speaker. Four-way ANOVAs ( $2 \times 2 \times 3 \times 5$ ) were used to analyze the data for each tone separately. Gender (male vs female) was the between subject factor; intonation (question vs statements), position (initial, medial, and final) and time point (0%, 25%, 50%, 75%, and 100%) were the within subject factors. Overall, females produced significantly higher F0 than male speakers for all six tones [main effect of gender: tone 55,  $F(1, 18) = 189.00$ ,  $p < 0.001$ ; tone 25,  $F(1, 18) = 262.23$ ,  $p < 0.001$ ; tone 33,  $F(1, 18) = 298.16$ ,  $p < 0.001$ ; tone 21,  $F(1, 18) = 250.36$ ,  $p < 0.001$ ; tone 23,  $F(1, 18) = 294.06$ ,  $p < 0.001$ ; tone 22,  $F(1, 18) = 394.22$ ,  $p < 0.001$ ].

Tones produced in questions had higher F0 levels than the same tones in statements [main effect of intonation: tone 55,  $F(1, 18) = 132.77$ ,  $p < 0.001$ ; tone 25,  $F(1, 18) = 262.23$ ,  $p < 0.001$ ; tone 33,  $F(1, 18) = 148.12$ ,  $p < 0.001$ ; tone 21,  $F(1, 18) = 123.48$ ,  $p < 0.001$ ; tone 23,  $F(1, 18) = 153.54$ ,  $p < 0.001$ ; tone 22,  $F(1, 18) = 197.55$ ,  $p < 0.001$ ]. Significant interaction effects between intonation and position were also found for all six tones [tone 55,  $F(2, 36) = 36.46$ ,  $p < 0.001$ ; tone 25,  $F(2, 36) = 36.51$ ,  $p < 0.001$ ; tone 33,  $F(2, 36) = 120.22$ ,  $p < 0.001$ ; tone 21,  $F(2, 36) = 89.78$ ,  $p < 0.001$ ; tone 23,  $F(2, 36) = 88.73$ ,  $p < 0.001$ ; tone 22,  $F(2, 36) = 125.73$ ,  $p < 0.001$ ]. Post hoc analysis showed that the F0 level of questions was significantly higher than that of statements for (i) all tones (Tukey HSD test,  $p < 0.01$ ) except tone 21 (Tukey HSD test,  $p > 0.05$ ) at the initial position, (ii) tones 55, 25, 33, and 23 at the medial position (Tukey HSD test,  $p < 0.01$  for all), and (iii) all six tones at the final position (Tukey HSD tests,  $p < 0.001$  for all).

The three positions (initial, medial, and final) were found to be significantly different from each other for each of the six tones [main effect of position: tone 55,  $F(2, 36) = 80.99$ ,  $p < 0.001$ ; tone 25,  $F(2, 36) = 125.76$ ,  $p < 0.001$ ; tone 33,  $F(2, 36) = 127.19$ ,  $p < 0.001$ ; tone 21,  $F(2, 36) = 19.00$ ,  $p < 0.001$ ; tone 23,  $F(2, 36) = 80.75$ ,  $p < 0.001$ ; tone 22,  $F(2, 36) = 50.85$ ,  $p < 0.001$ ]. In statements, tones at the initial position had higher F0 level than tones in medial position, which in turn had higher F0 than tones in final position. The difference in F0 level between the initial and the final positions was significant for both genders and all six tones (Tukey HSD test,  $p < 0.001$  for all). For questions, the differences in F0 between the initial and medial positions were significant for all six tones in both genders (Tukey HSD test,  $p < 0.001$  for all), except for tone 22 produced by male speakers (Tukey HSD test,  $p > 0.05$ ). At the final position, owing to the differences in tone contour compared with the other two positions, the differences in overall F0 level between the final and the initial positions and between the final and medial positions were not compared statistically. However, Figs. 1(a)–1(f) show that all tones at the final position of questions began at an F0 level below that of the initial and

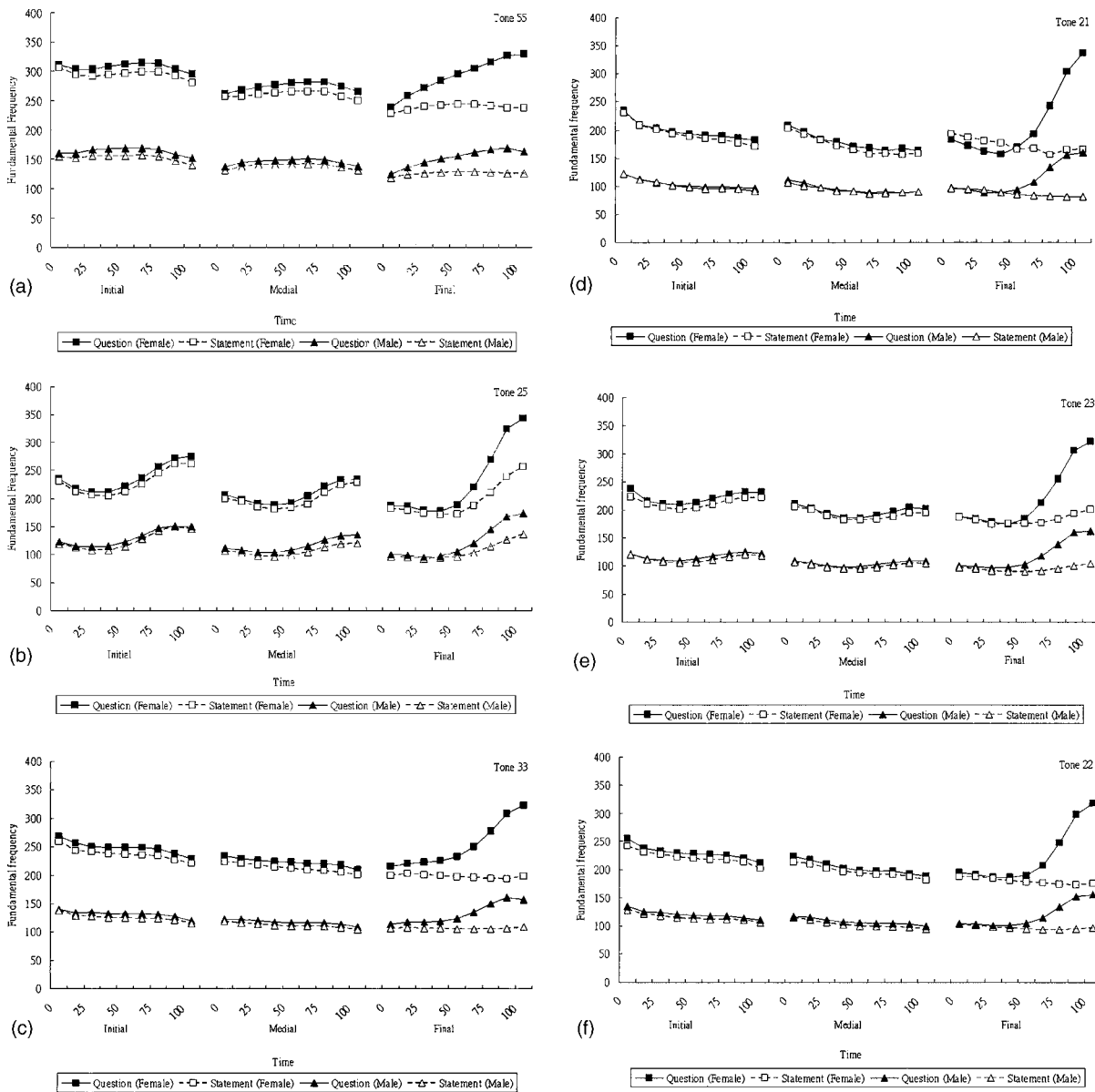


FIG. 1. (a)–(f). Mean F0 values for the six tones are displayed for (a) tone 55, (b) tone 25, (c) tone 33, (d) tone 21, (e) tone 23, and (f) tone 22. Within each figure, the F0 patterns at the initial, medial, and final positions are shown horizontally. For each tone, the mean F0 values at nine time points are displayed. Open symbols represent statements while filled symbols represent questions. Squares represent mean F0 values for female speakers and triangles represent male speakers.

medial positions, and increased to a level above that of the initial position from either 50% or 75% of the total duration onwards.

Analysis of the interaction of position, intonation, and time point showed the influence of intonation on F0 for different contexts at each time point. At the initial and medial positions, the F0 value of tones in questions was higher than in statements in all time points. Significant differences were noted at most of the time points for tones 55 (Tukey HSD test,  $p < 0.05$  at 25%, 50%, 75%, and 100% of the time) and 33 (Tukey HSD test,  $p < 0.05$  for all time points), but the differences between questions and statements were generally not significant across time points for tones 25, 21, 23, and 22 for both genders (Tukey HSD test,  $p > 0.05$ ). At the final position, tones produced in questions had higher F0 values than those of statements at most time points. The F0 values

of tones produced in questions were significantly higher than in statements at 75% and 100% of the total duration for all six tones in both genders (Tukey HSD test,  $p < 0.001$  for all), suggesting that the difference in F0 between statements and questions grew bigger towards the end of the tones at the final position.

## 2. Duration

The mean duration of each tone at each position of the two intonations is shown in Fig. 2. A four-way repeated ANOVA ( $2 \times 2 \times 3 \times 6$ ) was used to compare the differences in duration of the six tones between the two intonations across the three positions of both genders. Male and female speakers produced similar duration for the targets [ $F(2,36) = 195.18, p > 0.05$ ]. The main effect of position was statisti-

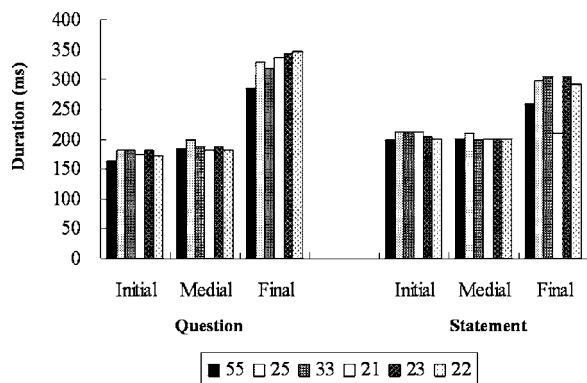


FIG. 2. Mean duration for the six tones across the three positions (initial, medial, and final) in questions and statements are displayed.

cally significant [ $F(10,180)=13.52, p < 0.05$ ]. Specifically, the duration of the targets at the final position was significantly longer than those in the initial and medial positions (Tukey HSD test,  $p < 0.05$  for both), while there was no significant difference in duration between the targets at initial and medial positions (Tukey HSD test,  $p > 0.05$ ). The position by tone interaction [ $F(10,180)=13.52, p > 0.05$ ] showed that this pattern of differences among the three positions was the same for all six tones. Although the main effect for intonation was not significant [ $F(1,18)=0.02, p > 0.05$ ], the intonation by position interaction [ $F(2,36)=149.60, p < 0.05$ ] showed that targets at the initial and medial positions of statements had significantly longer duration than their counterparts produced in questions (Tukey HSD test,  $p < 0.05$  for both); while targets at the final position of questions had a longer duration than those in statements (Tukey HSD test,  $p < 0.05$ ). The main effect for tone was statistically significant [ $F(5,90)=18.71, p < 0.05$ ]. Tones 55 and 21 had significantly shorter durations when compared with the other four tones (Tukey HSD test,  $p < 0.05$  for all), while the differences between tones 25, 33, 23, and 22 were not significant (Tukey HSD test,  $p > 0.05$  for all). The tone by intonation interaction [ $F(5,90)=9.79, p < 0.05$ ] showed that tone 55 had shorter duration than the other five tones when produced in questions and tones 25, 33, 23, and 22 when produced in statements (Tukey HSD test,  $p < 0.05$  for all); while tone 21 had shorter duration than the other five tones only when produced in statements (Tukey HSD test,  $p < 0.05$  for all). The interaction between tones, intonation, and position [ $F(10,180)=15.68, p < 0.05$ ] showed that the abovementioned pattern between tones by intonation was only observed at the final position (Tukey HSD test,  $p < 0.05$  for all); while there was no significant contrast in duration between tones at initial and medial positions (Tukey HSD test,  $p > 0.05$  for all).

### C. Discussion

The results of the above analysis showed that intonation affects the F0 patterns of tones across positions: tones produced within questions have a higher F0 level, the F0 level of tones varied across positions in both questions and statements, and a rising F0 contour was found for all tones at the final position of questions. The average F0 level of tones in

questions was higher than their counterparts produced in statements (except for tone 21 at the initial and medial positions, and tone 23 at the medial position). The difference in F0 between intonation contexts was largest at the final position, which is consistent with the rising F0 contour associated with questions (Ohala, 1983). Liu and Xu (2005) proposed that the difference in F0 level between questions and statements is related to the meaning of surprise or incredulity carried in questions.

In statements, tones at the initial position had the highest F0 level, followed by tones at the medial and the final positions. The changes in F0 level for tones at various positions of statement support the findings of previous studies on the F0 pattern of declarative sentences (Ohala, 1978; Vance, 1976). Declarative sentences (or statements) are characterized by a lowering of F0 in successive words from the beginning to the end of the sentence ("downdrift;" Ohala, 1978), and by a final fall in the contour of the last syllable (Vance, 1976). This decrease in the F0 level of tones from the initial to the medial and final position supports the existence of a declination effect in statements in Cantonese.

F0 contours of the six tones remained consistent across positions and intonations, with the exception of contours of tones in the final positions of questions. The rising intonation contour of questions changed the F0 contour of all six tones from the canonical forms to rising contours at the final positions. This intonation-induced F0 change at the final position of questions was in agreement with the findings of Fok-Chan (1974) and Lee (2004) on the interaction between tone and intonation in Cantonese. In addition, tones 25, 21, 23, and 22 showed overlapping F0 contours at the final position of questions. The F0 patterns of these four tones resembled the original contour of tone 25.

The results of experiment 1 also showed that the duration of the targets varied as a function of positions within intonation contexts. The duration of the six tones remained similar within each condition (statement initial, statement medial, statement final, question initial, question medial, and question final), except for shorter duration for tone 55 at the final position of questions and statements, and tone 21 at the final position of statements. Besides the lengthening of tone 21 at the final position of questions, Lee (2004) also reported that tones 33 and 22 had lengthened durations at the final position of questions when compared with those in statements. Similar findings were not observed in this study. However, Lee (2004) did not perform statistical analyses for evaluating the difference in tone duration between intonations. Lee (2004) also reported variability among the four speakers in her study; some speakers showed no observable difference in duration. The findings of the present study also support the idea that duration effects vary among speakers.

### III. EXPERIMENT 2—PERCEPTUAL ANALYSIS

The findings of experiment 1 showed that intonation affects both F0 level and contour. Intonation had no overall duration effect on tone, although interaction between position and intonation showed that the duration of six tones varied as a function of positions within intonation contexts.

The effect of intonation on F0 was especially strong at the final position of questions, where all six tones showed rising F0 contours regardless of their canonical forms, and overlapping of F0 patterns between tones 25, 21, 23, and 22 were found. Previous studies showed that F0 is the most important perceptual cue in tone perception (Fok-Chan, 1974; Vance, 1976; Gandour, 1981). Fok-Chan (1974) suggested that once the F0 structure of the tone is disrupted, then perception would be affected. The second experiment was designed to investigate whether intonation-induced changes in the F0 patterns of tones (i.e., F0 level and F0 contour) affect tone identity. Previous perceptual experiments in Mandarin showed that intonation-induced F0 changes do not necessarily change the lexical identity of tones for listeners (Connell, Hogan, and Rozsypal, 1983). Although the acoustic analysis showed different effect of intonation on the F0 contour of tones in Mandarin and Cantonese, the presence of extrinsic context (i.e., the intonation contour of the sentence) may provide listeners cues to perceptually “normalize” for the changes in the F0 pattern of the lexical tones. Therefore, three different experimental contexts (original carrier, isolation, and neutral carrier) were designed to test whether the extrinsic context can help listeners in the identification of tones whose F0 contour has undergone intonation-related changes.

## A. Method

### 1. Listeners

Twelve females, aged 18–19 years old, served as listeners. They were all first-year undergraduates in the Division of Speech and Hearing Sciences, University of Hong Kong. None of them had participated in experiment 1. They were considered naïve listeners, as the experiment was carried out within their first two months at university, during which they received no phonetic training on tones. Cantonese was the first language for all the listeners. All listeners passed a hearing screening ( $\leq 20$  dBHL at 250, 500, 1000, 2000, and 4000 Hz). A tone perception screening, consisting of 24 trials (four trials for each tone), was carried out with all the listeners. Stimuli used were different from those in the experimental session. Listeners had to correctly identify each of the six tones in at least three out of four trials in order to pass the screening. They achieved an overall average accuracy of 99.3%.

### 2. Speech materials

Speech materials from two of the speakers in experiment 1, one male and one female, were used for the present experiment. The two speakers were selected because the F0 patterns of their target tones were closest to the average F0 patterns of their own gender. All 108 utterances of each speaker were used in this experiment (three sets of six tones of target words, three positions, and two intonations). Three presentation conditions (original carrier, isolation, and neutral carrier) were selected to investigate the use of extrinsic context by listeners in the perception of intonation-induced F0 changes in tones. In the original carrier condition, the targets were presented with the original carriers, as recorded

in experiment 1. In the isolation condition, targets were presented as single words in order to determine the effect of a lack of extrinsic context on tone identification. Stimuli were manually extracted, using the Praat software (Boersma and Weenink, 2003), from the initial, medial, and final positions of the original utterances. The beginning and the end of the syllable were selected from the amplitude waveform display. The neutral carrier condition was designed to investigate whether the presence of a carrier sentence that provided cues about the average F0 and the F0 range of the speaker could be exploited to improve the accuracy of the identification of intonation-induced F0 changes. This condition was motivated by the finding that the presence of extrinsic context in the form of a carrier sentence results in sharper identification boundaries for level tones (Francis, Ciocca, and Ng, 2003). In the neutral carrier condition, all the target tones in isolation form were appended at the final position of the same carrier /lei<sub>55</sub> kc<sub>33</sub> tsi<sub>22</sub> h<sub>v</sub>i<sub>22</sub>/ (This word is). The neutral carrier was synthesized from the production of the original carrier with targets in final position (/lei<sub>55</sub> kc<sub>33</sub> tsi<sub>22</sub> h<sub>v</sub>i<sub>22</sub>/) by these two speakers. For each speaker, the average F0 of each syllable of all the 18 productions of /lei<sub>55</sub> kc<sub>33</sub> tsi<sub>22</sub> h<sub>v</sub>i<sub>22</sub>/ was calculated. Among the 18 productions, the utterance with F0 values that were closest to the average F0 in each of the four syllables was chosen. The F0 of each syllable within the selected utterance was then resynthesized using the PSOLA algorithm of Praat software (Boersma and Weenink, 2003) to be within 2 Hz of the average F0 for each syllable at each of the nine time points. The two synthesized carriers (one for the male speaker and the other for the female speaker) were judged to sound natural by the first author and two other native Cantonese speakers (all qualified speech and language pathologists). All the target tones in isolation were then appended at the final position of the neutral carrier. In each presentation condition, the overall loudness of each of the stimuli was then equalized by the first author by modifying the overall amplitude of each token as necessary using the Praat software (Boersma and Weenink, 2003).

### 3. Procedures

The experiment was carried out in a single wall IAC sound-attenuated booth, with the speech materials presented to the listeners through a Sennheiser HD 545 headset, connected to a G4 Apple Macintosh computer, with an Aardvark USB 3 sound card. A HyperCard (Apple™) program was used to run the experiment. For each trial, six Chinese characters were presented on the screen, representing one set of contrastive tones. The carrier was also presented at the top of the screen with the target word represented by “\_” for the original carrier and neutral carrier conditions.

The stimuli were divided into six blocks, according to presentation condition (original carrier, isolation form, and neutral carrier) and speaker. Within each block, there were a total of 216 trials, as each of the 108 stimuli was repeated once. Three experimental sessions were scheduled at least one week apart to minimize learning effects. In each session, two blocks (male and female) of one presentation condition were presented. The order of presentation for both presentation condition and speaker was randomized and counterbal-

anced across listeners. Each block took about 1/2 h to finish, and each session took about 1 h. Listeners were informed that the stimuli consisted of both questions and statements. Before each session, the 18 Chinese characters were read aloud by the first author to ensure that listeners were familiar with all the characters, as the same Chinese character may have more than one pronunciation in different contexts in Cantonese. Within each trial, listeners were asked to identify the Chinese character that matched the word they heard by clicking on the button representing the character. Each stimulus was presented once and the listener could opt to listen to it a second time by clicking on a “repeat sound” button.

## B. Results

### 1. Overall perceptual accuracy

Confusion matrices were compiled separately according to presentation condition (original carrier, isolation, and neutral carrier), position (initial, medial, and final), and intonation context (question and statement) for each listener. Group confusion matrices were compiled by summing confusion matrices across 12 listeners. The overall percentages of correct identification for each of the three presentation conditions were compared using a series of Wilcoxon matched pair tests as ceiling effects were observed for some targets. The overall accuracy for tones presented within the original carrier (mean 90.57, SD=25.40) was significantly higher than for targets presented in isolation (mean 72.22, SD =33.43) ( $T=0, p<0.005$ ) and for targets presented within the neutral carrier (mean 70.39, SD=35.72) ( $T=0, p<0.005$ ); there was no difference between the isolation and the neutral carrier conditions ( $T=29.5, p>0.05$ ). Statistically significant differences in perceptual accuracy were observed for original carrier and the other two conditions in all intonation contexts and positions ( $p<0.05$  for all), except for the final targets of questions where the differences between original carrier, neutral carrier, and isolation were not significant ( $p>0.05$  for all). Targets presented in isolation had significantly higher accuracy than when presented within the neutral carrier at initial and medial positions of questions, and initial position of statements ( $p<0.05$  for all); while targets presented within neutral carrier had perceptual accuracy higher than those presented in isolation at the medial and final positions of statements ( $p<0.05$  for all).

The percentages of correct identification of targets presented with the original carrier in each intonation context and position were compared individually for each tone. Performance was similar in all the conditions, except for the final targets of questions for tones 21, 23, and 22, for which accuracy was significantly lower than in other conditions ( $T=0, p<0.005$  for all).

The confusion matrix for the perception of the six tones as final targets of questions in the original carrier condition is shown in Table I. The numbers in the cells represent the percentage of responses realized as that particular tone. For example, tone 25 at the final position of questions was correctly identified 81.9% of the time (118 out of 144 trials). Correct identifications are shown by the numbers appearing on the diagonal of the matrix (boldtype). The mean accuracy

TABLE I. Confusion matrix for perceptual accuracy at the final position of questions within the original carrier condition.

Target	Perceived tone					
	55	25	33	21	23	22
55	<b>100.0</b>					
25		<b>81.9</b>		1.4	6.3	10.4
33		25.7	<b>71.5</b>		2.8	
21		71.5		<b>27.1</b>	1.4	
23		78.5	1.4		<b>9.7</b>	10.4
22		62.5	6.3		9.7	<b>21.5</b>

Note. Target tones are on the vertical axis, and perceived tones on the horizontal axis. Cell numbers represent the percentage of responses for each target tone.

of the six tones in this context was compared using Wilcoxon matched-pair tests. The percentage of correct identification for tone 55 (mean=100%) was significantly higher than for the other five tones ( $T=0, p<0.05$  for all). The accuracy for tones 25 (mean=82%) and 33 (mean=72%) was significantly higher than for tones 21, 23, and 22 ( $p<0.05$  for all), but the difference between tones 25 and 33 was not statistically significant ( $p>0.05$ ). The mean accuracy for tone 21 (mean=27%) was higher than for tones 23 (mean=10%) and 22 (mean=22%) ( $T=0, p<0.05$  for both); no significant difference was found between tones 23 and 22 ( $T=3, p>0.05$ ).

### 2. Error patterns

Error patterns were analyzed by calculating the percentage of misidentifications of each tone out of the total number of identification errors observed for that tone within each presentation condition. As described in the previous section, in the original carrier condition, stimuli presented at the final position of questions were the least accurately perceived. Except for tones 55 and 25, most of the errors in this context involved misperceiving the target tone as tone 25. This error accounted for about 66% of the identification errors for tone 33, 93% of the identification errors for tone 21, 76% of the identification errors for tone 23, and 73% of the identification errors for tone 22. Two other common perceptual errors were noted in the original carrier condition. Tone 25 was misperceived as tone 23 (69% of all perceptual errors involving tone 25). The other common, but less frequent, error resulted from the mutual confusion between tones 33 and 22.

For targets presented with the neutral carrier and in isolation, the error pattern for the final targets of questions was similar to that for the original carrier condition. Tones 33, 21, 23, and 22 were misperceived as tone 25. Also, a large number of tone level errors were observed in these two presentation conditions in all six contexts, accounting for about 65%–96% of all the perception errors for each tone in each presentation condition. For the three level tones, tone 55 was perceived as tone 33, and tones 33 and 22 were perceived as tones 55, 33, or 22; the two rising tones (tones 25 and 23) were confused with each other.



TABLE II. Performance of individual listeners for the final position of questions within the original carrier condition.

Listeners	Tone					
	55	25	33	21	23	22
L1	100	33.3	100	91.7	33.3	83.3
L2	100	16.7	100	91.7	8.3	91.7
L3	100	100	0	0	0	0
L4	100	100	8.3	0	0	0
L5	100	100	0	0	0	0
L6	100	100	50	0	0	0
L7	100	83.3	100	8.3	0	0
L8	100	100	100	0	0	0
L9	100	100	100	8.3	8.3	8.3
L10	100	100	100	25	0	16.7
L11	100	75	100	50	50	25
L12	100	66.7	100	50	16.7	25

Note. Numbers in the cells represent the accuracy (percentage correct) for that particular tone by each listener.

### 3. Individual error patterns

The confusion matrices of all individual listeners were examined. The percentages of correct identification of each listener at the final position of questions are shown in Table II. Listeners could be classified into four groups according to their identification patterns:

a. Listeners 1 and 2 were able to correctly identify tones 55, 33, 21, and 22 for final targets of questions. Their perception of tones 25 and 23 was poor (tone 25 being misperceived as tone 23, 21, or 22, and tone 23 misperceived as tone 25 or 22).

b. Listeners 3, 4, 5, and 6 were able to correctly identify tones 55 and 25, but frequently misidentified the other tones as tone 25.

c. In addition to being able to perceive tones 55 and 25 correctly (as for listeners in the previous group), listeners 7, 8, 9, and 10 were able to perceive tone 33 with 100% accuracy, while tones 21, 23, and 22 were likely to be confused with tone 25.

d. Listeners 11 and 12 perceived tones 55 and 33 with a high level of accuracy. These listeners were also able to perceive tones 21, 23, and 22 more accurately than listeners in groups 2 and 3, but they were less accurate in identifying tone 25.

### C. Discussion

For the targets presented with the original carriers, identification accuracy for all six tones was close to 100% (ranged from 91% to 100%) in all conditions except for the final targets in questions, whose F0 contours deviated from their canonical form. For these target tones, the identification accuracy ranged from 10% (tone 23) to 100% (tone 55). The results showed that a large proportion of tones 21, 23, and 22 were misperceived as tone 25 at the final position of questions, likely due to the overlapping rising F0 contours of these tones 21, 23, and 22 with tone 25, as shown in experiment 1. Therefore, listeners relied heavily on the intrinsic F0 patterns of these tones, as suggested in previous studies

(Fok-Chan, 1974; Gandour, 1981; Vance, 1976). Although all three level tones (tones 55, 33, and 22) had a rising contour at the final position of questions, listeners showed better perceptual accuracy for tones 55 and 33. The distinctiveness in the F0 level of tones 55 and 33, especially at the initial portion of the tone, is likely to have cued an accurate identification of these tones, given that there are no other rising tones at the same F0 level during the first half of the segments. Tone 55 also showed a rise in F0 throughout the segment, while the other five tones had a level F0 pattern in the first half of the segment. The perceptual accuracy for targets at the final position of questions in the original carrier condition was lower than in the other five contexts with the original carrier, and similar to that of the isolation conditions. This finding suggests that F0 information provided by the extrinsic context, such as the intonation type and the position of a word within a sentence, may not be used effectively when intonation-induced changes result in F0 patterns that differ considerably from the canonical pattern.

The results of experiment 2 showed that while in the original carrier condition listeners were able to employ the F0 cues provided by the extrinsic context in order to compensate for the intonation-induced F0 changes, and when the extrinsic context was removed (isolation condition), identification accuracy decreased. It is possible that subjects may not have made the maximal use of the F0 information regarding intonation patterns with isolated tones, as listeners were not told explicitly the intonation of the original carrier from which each isolated tone was extracted. However, Ma, Ciocca, and Whitehill (2006a) found that listeners had no difficulty in identifying the intonation pattern when they heard isolated tones produced in sentence-final position.

The perceptual patterns of the targets presented with the neutral carriers showed how extrinsic context cued tone perception. Perceptual errors were most common for tones 33, 21, and 22 in initial position and tone 22 in medial position, and most of the errors involved confusions with tones of the same contour but of higher F0 level. As shown in experiment 1, the F0 level of tones varied across position, with F0 level being the highest at the initial position, followed by that of medial and final positions. When all targets were placed at the final position of the neutral carrier, targets of the initial and medial positions were of higher F0 level than typical targets at the final position. These results are consistent with the "recency strategy" (Wong and Diehl, 2003), according to which the immediate preceding context provides cues for listeners to infer the F0 range of the talker's voice when processing an utterance. The extracted F0 range (or average F0) is then used to identify tones by carrying out a tone normalization process (Francis *et al.*, 2006; Leather, 1983).

### IV. GENERAL DISCUSSION

The results of the present study showed that the F0 patterns of tones in all positions of statements, and in the initial and medial positions of questions, were comparable with the canonical forms. By contrast, the F0 contours of all tones at the final position of questions were modified to rising by the final rise of the question. These results are in agreement with

the findings from other studies on the interaction between tone and intonation in Cantonese (Fok-Chan, 1974; Vance, 1976; Lee, 2004). These conclusions contrast with those from studies on Mandarin Chinese, which reported that the F0 contour of tones was not affected by sentence intonation (Ho, 1977; Rumjancev, cited in Lyovin, 1978; Shen, 1989). This difference might be related to how question intonation is produced by speakers of the two languages. Studies on intonation in Cantonese suggested that the difference between questions and statements is modeled by both global F0 increase and local F0 changes at the end of the final syllable of questions (Gu, Hirose, and Fujisaki, 2005; Ma, Ciocca, and Whitehill, 2006b). Gu *et al.* (2005) also suggested that the increase in amplitude of the positive tone command is larger than usual at the end of questions. The local F0 changes observed at the final position of questions provided a possible explanation of the modification of F0 contour at the final position of questions in Cantonese. Lin (2004) also used the concept of "boundary tones," the F0 changes at the final one or two syllables of the utterance which convey information about sentence intonation (e.g., question vs statement), to explain the difference between questions and statements in Mandarin. He found that boundary tones reflect the F0 pattern of a question by changing the F0 level at the beginning of the boundary tone and/or the slope of the rising F0. By contrast, Yuan (2004) suggested that boundary tones are not necessary for modeling of questions in Mandarin. He proposed that the difference between the two intonations could be explained by an overall higher phrase curve and higher strength value towards the end of questions. Liu and Xu (2005) found that the F0 difference between the two intonations is mostly related to the superposition of an exponential or double-exponential F0 baseline on the phrase curve of the sentence, which attributed the final F0 rise in questions to part of the global intonation changes of questions. This suggests that questions are mainly marked by global F0 changes in Mandarin and, therefore, the tone structures are largely unaffected.

The current results also showed that the intonation-induced F0 changes observed at the final position of questions affect listeners' perception. A large proportion of tones 21, 23, and 22 were perceived as tone 25 at the final position of questions. These results agreed with previous reports that tones produced within questions were more difficult to recognize than those in declarative sentences (Fok-Chan, 1974). By contrast, a previous perceptual study of Mandarin found that the perturbation in F0 pattern caused by intonation had minimal effect on perception for native listeners (Connell, Hogan and Rozsypal, 1983). The higher tendency of misperception in Cantonese when compared with Mandarin is probably related to the difference in the effect of intonation on tone as mentioned above, as well as to the features of the tones in each language. In Cantonese, there are three level tones and two rising tones. Therefore, a slight modification in F0 level or contour may result in erroneous identification. For example, an increase in F0 range of tone 23 may lead to the perception of a tone 25; similarly, a slight increase in F0 level could cause a tone 22 to be perceived as tone 33. By contrast, Mandarin has four tones with distinctive F0 con-

tours (level, rising, falling-rising and falling) and levels. Therefore, a small rise in the contour of the level tone (tone 1) caused by interrogative intonation, as shown in Ho's experiment (1977), might not affect the listeners' perception, as the resulting contour would still not resemble other tones in Mandarin. Furthermore, Tseng, Massaro, and Cohen (1986) found that duration and amplitude could become secondary cues in tone perception in Mandarin when F0-related cues are ambiguous. Several studies have explored the possibility that perceptual cues such as duration (Blicher, Diehl, and Cohen, 1990; Liu and Samuel, 2004; Tseng *et al.*, 1986) and amplitude envelope (Fu and Zeng, 2000; Whalen and Xu, 1992) are used for Mandarin tone identification. Unlike Mandarin tones, the present study showed that duration of tones varied as a function of position within intonations in Cantonese. Similar duration was observed for the six Cantonese tones of the same position and intonation, except for tones 55 and 21 at the final position of statements and tone 55 at the final position of questions. This suggests that the duration difference is unlikely to be a reliable cue in tone identification. Vance (1976) also found that the difference in vowel length among Cantonese tones was too small to be used for tone identification. It is possible that amplitude-related cues, not included in the present study, could be used as secondary cues to tone perception by Cantonese listeners.

The acoustic analysis in experiment 1 showed that the F0 contours of tones 21, 23, and 22 at the final position of questions were modified by the final-rise in questions to a surface F0 assimilating that of tone 25. Vance (1976) suggested that the tonal variations owing to intonation context could be classified as a form of tone sandhi. Wang and Li (1967) proposed that if a tone undergoes tone sandhi, it should be treated as a homophone of the resulting tone. In the current study, although most of the tones 21, 23, and 22 at the final position of questions were perceived as rising tones in experiment 2, the large amount of variability in the individual data (as shown in Table II) indicate that these tones were not homophonous to tone 25. This suggests that the tonal variations observed at the final position of question in Cantonese should be considered as the result of contextual modification by intonation rather than as a form of tone sandhi.

The large perceptual variability for tones at the final position of questions in the original carrier demonstrated that, although the canonical F0 patterns of tones were distorted by intonation context, some listeners were still able to recover the intended tones while some listeners took the distorted F0 pattern at "face value." One of the possible explanations is that different F0 cues were used by different listeners. Gandour (1981) proposed that Cantonese tones are perceived by features (like F0 level and F0 contour) rather than as whole tone units. The present results provided evidence that some listeners (listeners 1 and 2) placed more emphasis on tone level than contour; therefore, they were able to distinguish the closely spaced tones 21 and 22 accurately by the differences in F0 level, but failed to use the rising contour as a cue in perceiving both tones 25 and 23. By contrast, another group of listeners (listeners 3, 4, 5, and 6) weighed tone contour as the main perceptual cue, and

misperceived tones 33, 21, 23, and 22 as tone 25, while the contrast in F0 level between tone 33 and tone 25 was not perceived. For some listeners (listeners 7–12), both F0 contour and level were utilized as perceptual cues in tone perception. Therefore, they were able to perceive tones 55 and 33 accurately by exploiting the F0 level cue, while tones 21, 23, and 22 were confused with tone 25 due to their rising contour.

It is not clear how listeners weighed different F0 cues in tone identification. One possibility is that some listeners take into account the sentence intonation when they identify tones. When tones were presented with the original context, as in experiment 2, listeners tried to compensate for the final F0 rise in questions while trying to identify the tones of the target words. Therefore, they might have focused on the F0 level cues since all tones had rising F0 contours as a result of the final rise in questions. Alternatively, some listeners may have chosen to ignore the sentence intonation during the perception task, as they were asked to focus only on identifying the tones. In this case, listeners would perceive the tones on the basis of their surface F0 patterns and, therefore, fail to differentiate tones of similar F0 levels (tones 25, 21, 23, and 22). Further research on the perception of tones and intonation in Cantonese would be necessary to determine how different perceptual cues are utilized by individual listeners.

## ACKNOWLEDGMENTS

The authors wish to thank Anders Löfqvist and two anonymous reviewers for their helpful comments on an earlier draft of the manuscript.

- Bauer, R. S., and Benedict, P. K. (1997). *Modern Cantonese Phonology* (Mouton de Gruyter, Berlin).
- Blicher, D. L., Diehl, R. L., and Cohen, L. B. (1990). "Effects of syllable duration on perception of Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement," *J. Phonetics* **18**, 37–49.
- Boersma, P., and Weenink, D. (2003). Praat 4.0.46: A system for doing phonetics by computer (computer software) (University of Amsterdam, Amsterdam, The Netherlands). Available online: <http://www.fon.hum.uva.nl/praat/> Last accessed 10/30/2006.
- Chang, C. T. (1958). "Tones and intonation in Chengtu dialect," *Phonetica* **2**, 59–85.
- Chao, Y. R. (1947). *Cantonese primer* (Greenwood Press, New York).
- Connell, B. A., Hogan, J. T., and Rozsypal, A. J. (1983). "Experimental evidence of interaction between tone and intonation in Mandarin Chinese," *J. Phonetics* **11**, 337–351.
- Fok-Chan, Y. Y. (1974). *A Perceptual Study of Tones in Cantonese* (University of Hong Kong Press, Hong Kong).
- Francis, A., Ciocca, V., and Ng, B. K. C. (2003). "On the (non)categorical perception of Cantonese lexical tones," *Percept. Psychophys.* **65**, 1029–1044.
- Francis, A., Ciocca, V., Wong, N. K. Y., Leung, W. H. Y., and Chu, P. C. Y. (2006). "Extrinsic context affects perceptual normalization of lexical tone," *J. Acoust. Soc. Am.* **119**, 1712–1726.
- Fu, Q.-J., and Zeng, F.-G. (2000). "Identification of temporal envelop cues in Chinese tone recognition," *Asia Pac. J. Speech, Language Hearing* **5**, 45–57.
- Fujisaki, H., and Hirose, K. (1984). "Analysis of voice fundamental frequency contours for declarative sentences of Japanese," *J. Acoust. Soc. Jpn.* **5**, 233–242.
- Gandour, J. (1981). "Perceptual dimensions of tone: Evidence from Cantonese," *J. Chin. Linguist.* **9**, 20–36.
- Gu, W.-T., Hirose, K., and Fujisaki, H. (2005). "Analysis of the effects of word emphasis and echo question on F0 contours of Cantonese utterances," *Proceedings of the Interspeech-2005*, Lisbon, Portugal, pp. 1825–2828.
- Ho, A. T. (1977). "Intonation variation in a Mandarin sentence for three expressions: Interrogative, exclamatory and declarative," *Phonetica* **34**, 446–457.
- Ladd, D. R. (1996). *Intonational Phonology* (Cambridge University Press, Cambridge).
- Leather, J. (1983). "Speaker normalization in perception of lexical tone," *J. Phonetics* **11**, 373–382.
- Lee, W.-S. (2004). "The effect of intonation on the citation tones in Cantonese," *Proceedings of the First International Symposium on the Tonal Aspects of Languages*, Beijing, China, pp. 107–110.
- Lin, M. (2004). "Boundary tone of Chinese intonation and its pitch (F0) pattern," in *From Traditional Phonology to Modern Speech Processing*, edited by G. Fant, H. Fujisaki, J. Cao, and Y. Xu (Foreign Language Teaching and Research Press, Beijing), pp. 309–327.
- Liu, F., and Xu, Y. (2005). "Parallel encoding of focus and interrogative meaning in Mandarin intonation," *Phonetica* **62**, 70–87.
- Liu, S.-Y., and Samuel, A. G. (2004). "Perception of Mandarin Lexical tones when F0 information is neutralized," *Lang Speech* **47**, 109–128.
- Lyovin, A. V. (1978). "Review of tone and intonation in Modern Chinese by M. K. Rumjancev," *J. Chin. Linguist.* **6**, 120–168.
- Ma, J. K.-Y., Ciocca, V., and Whitehill, T. L. (2006a). "Perception of intonation in Cantonese," *Proceedings of the Second International Symposium of Tonal Aspects of Languages*, La Rochelle, France, pp. 93–96.
- Ma, J. K.-Y., Ciocca, V., Whitehill, T. L. (2006b). "Quantitative analysis of intonation patterns in statements and questions in Cantonese," *Proceedings of the Third International Conference on Speech Prosody*, Dresden, Germany, pp. 277–280.
- Ohala, J. J. (1978). "Tone Rules," in *Tone: A Linguistic Approach*, edited by V. A. Fromkin (Academic, New York), pp. 5–39.
- Ohala, J. J. (1983). "Cross-language use of pitch: An ethological view," *Phonetica* **40**, 1–18.
- Pierrehumbert, J. (1980). "The phonology and phonetics of English intonation," unpublished Ph.D. thesis, MIT.
- Rumjancev, M. K. (1972). *Ton i intonacija v sovemennom kitajskom jazyke (Tone and intonation in Modern Chinese)*. (Izdatel'stvo Moskovskogo Universiteta, Moscow).
- Shen, X. N. S. (1989). *The Prosody of Mandarin Chinese* (University of California Press, Berkeley, California).
- Thorsen, N. G. (1980). "A study of the perception of sentence intonation—evidence from Danish," *J. Acoust. Soc. Am.* **67**, 1014–1030.
- Tseng, C.-Y., Massaro, D. M. and Cohen, M. M. (1986). "Lexical tone perception in Mandarin Chinese: Evaluation and integration of acoustic features," in *Linguistics Psychology and the Chinese People*, edited by H. S. R. Kao and R. Hoosain (University of Hong Kong, Hong Kong), pp. 91–104.
- Vance, T. J. (1976). "An experimental investigation of tone and intonation in Cantonese," *Phonetica* **33**, 368–392.
- Wang, S.-Y. W., and Li, K.-P. (1967). "Tone 3 in Pekinese," *J. Speech Hear. Res.* **10**, 629–636.
- Whalen, D. H., and Xu, Y. (1992). "Information for Mandarin tones in the amplitude contour and in brief segments," *Phonetica* **49**, 25–47.
- Wong, P. C. M., and Diehl, R. L. (2003). "Perceptual normalization for inter- and intratalker variation in Cantonese level tones," *J. Speech Lang. Hear. Res.* **46**, 413–421.
- Yuan, J.-H. (2004). "Intonation in Mandarin Chinese: Acoustics, Perception, and Computational Modeling," unpublished Ph.D. thesis, Cornell University.
- Yuan, J.-H., Shih, C.-L., and Kochanski, G. P. (2002). "Comparison of declarative and interrogative intonation in Chinese," *Proceedings of the Speech Prosody-2002*, Aix-en-Provence, France, pp. 711–714.