The perception of intonation questions and statements in Cantonese[a)]

Joan K.-Y. Ma,[b)] Valter Ciocca,[c)] and Tara L. Whitehill
Division of Speech and Hearing Sciences, University of Hong Kong, 34 Hospital Road, Hong Kong Special Administrative Region, People's Republic of China

(Received 18 June 2009; revised 26 November 2010; accepted 27 November 2010)

In tone languages there are potential conflicts in the perception of lexical tone and intonation, as both depend mainly on the differences in fundamental frequency (F0) patterns. The present study investigated the acoustic cues associated with the perception of sentences as questions or statements in Cantonese, as a function of the lexical tone in sentence final position. Cantonese listeners performed intonation identification tasks involving complete sentences, isolated final syllables, and sentences without the final syllable (carriers). Sensitivity (d' scores) were similar for complete sentences and final syllables but were significantly lower for carriers. Sensitivity was also affected by tone identity. These findings show that the perception of questions and statements relies primarily on the F0 characteristics of the final syllables (local F0 cues). A measure of response bias (c) provided evidence for a general bias toward the perception of statements. Logistic regression analyses showed that utterances were accurately classified as questions or statements by using average F0 and F0 interval. Average F0 of carriers (global F0 cue) was also found to be a reliable secondary cue. These findings suggest that the use of F0 cues for the perception of intonation question in tonal languages is likely to be language-specific. © *2011 Acoustical Society of America*.
[DOI: 10.1121/1.3531840]

PACS number(s): 43.71.Sy, 43.71.Es [RSN] Pages: 1012–1023

I. INTRODUCTION

Intonation is a universal characteristic shared by different languages. It plays an important role in conveying information about various characteristics of human communication, such as the differences between statements and questions, and information about speakers' emotions and attitudes (Kohler, 2007; Ladd, 1996). While the signaling of interrogation in speech communication can be achieved through syntax, semantics or other means of communication (for example, wh-questions, tag questions, and yes/no questions) (van Heuven and Haan, 2000; Kohler, 2004), the present study focused on intonation and, specifically, the "intonation contour," as a cue to the perception of interrogation. The intonation contour is most commonly described by characteristics such as a final rise in fundamental frequency (F0) and a higher overall F0 level (Hirst and Di Cristo, 1998). Questions that are only differentiated from statements by means of differences in intonation contour will be referred to hereafter as "intonation questions." This study investigated the contrast between intonation questions and statements in Cantonese, a tonal language in which differences in syllabic F0 patterns are lexically contrastive (Chao, 1947; Bauer and Benedict, 1997).

The use of a final rise in F0 as an intonation cue to the perception of intonation questions has been shown to be common across languages (Lieberman, 1967; Remijsen and van Heuven, 1999; Gussenhoven and Chen, 2000; House, 2003; Schneider and Lintfert, 2003; Peters and Pfitzinger, 2008). Lieberman (1967) stressed that questions and statements can be reliably distinguished based on the F0 contour during the final 150 to 200 ms within a sentence. A similar conclusion was reached by Gussenhoven and Chen (2000), who investigated the significance of final rise in F0 as an intonation marker for the perception of questions. They asked three groups of monolingual listeners (Chinese, Dutch, and Hungarian) to identify re-synthesized CVCVCV (C = consonant; V = vowel) stimuli of a made-up language with different pitch heights and timings. They found that questions were associated with either a later or a higher F0 peak (i.e., F0 maximum) than statements, regardless of the language background of the listener. These findings were consistent with those obtained from the studies of intonation in various languages, including Swedish (House, 2003), German (Schneider and Lintfert, 2003), and Dutch (Remijsen and van Heuven, 1999). Peters and Pfitzinger (2008) investigated the effect of F0 interval (amount of F0 change between the onset and the offset at the utterance-final syllable) and F0 slope (rate of F0 change over time) of the final increase in F0 on the perception of questions in German. Listeners were presented re-synthesized stimuli with varying F0 contour and voicing duration, embedded at the final position of a short carrier. The results showed that the perception of

[a)]Portions of this work were presented in "Perception of intonation in Cantonese," Proceedings of International Symposium on Tonal Aspects of Language: Emphasis on Tone Languages, La Rochelle, France, April 2006, and in "Acoustic cues for the perception of intonation in Cantonese," Proceedings of International Conference on Spoken Language Processing, Brisbane, Australia, September 2008.

[b)]Author to whom correspondence should be addressed. Present address: Division of Speech and Hearing Sciences, Queen Margaret University, Musselburgh, Edinburgh, EH21 6UU, United Kingdom. Electronic mail: jma@qmu.ac.uk

[c)]Present address: School of Audiology and Speech Sciences, University of British Columbia, 2177 Wesbrook Mall, Vancouver B.C. V6T 1Z3, Canada.

questions was more strongly associated with F0 interval position than with F0 slope when the voicing of the signal was at least 50 ms in duration. In addition to a final rise in F0, the perception of questions has been associated with other supplementary cues such as a larger F0 range in a sentence (Hirschberg and Ward, 1992), a higher overall F0 level, and a pre-focal filled pause (i.e., a pause that occurs prior to the semantically important word in order to signal hesitation) (House, 2003).

A. Potential conflict in the parallel encoding of lexical and intonational meanings in tone languages

The perception of intonation in a tonal language can be potentially confusing, as F0 patterns mark not only intonation at the sentential level but also lexical tone at the syllabic level. For example, a statement ending with a high-rising tone could in principle be wrongly perceived as a question because a final rise in F0 is a cue to the perception of intonation questions. Several studies on the interaction between intonation and lexical tone in Chinese languages have focused on the effect that intonation imposes on the acoustic or perceptual characteristics of lexical tones (Chang, 1958; Fok-Chan, 1974; Ho, 1977; Rumjancev, cited in Lyovin, 1978; Connell *et al*., 1983; Shen, 1989; Lee, 2004; Ma *et al*., 2006b). Two general findings have been reported. In Mandarin, the intonation-induced perturbations of lexical tones mostly affect the *F0 level*, but the *F0 contour* remains similar to that of the canonical form (Chang, 1958; Ho, 1977; Rumjancev, cited in Lyovin, 1978; Shen, 1989). In Cantonese, in addition to the *F0 level* of the tone being modified, the *F0 contour* may also deviate from its canonical form due to the effect of intonation questions (Fok-Chan, 1974; Lee, 2004; Ma *et al*., 2006b). For example, Ma *et al*. (2006b) found that all six tones in Cantonese, regardless of their canonical form, had a rising F0 contour when occurring at the final position of an intonation question. Additionally, the intonation-induced F0 changes affected the lexical identity of tones, as a large portion of the intended falling and level tones at the final position of questions were identified as rising tones (Ma *et al*., 2006b).

The effects of lexical tone identity on the perception of intonation have so far received little attention. The potential confusion resulting from co-existing F0 patterns of tone and intonation has been investigated in Mandarin (Yuan, 2004a,b; Yuan and Shih, 2004). In these studies, questions and statements with a minimal tone contrast at the final syllable were presented. Listeners were asked to identify the sentences as "statements" or "questions." The results showed that the perception of statements was not affected by the identity of the lexical tone at the final position. However, questions ending with rising tones were less accurately identified (i.e., received a higher percentage of "statement" responses) than questions ending with a falling tone. This result was rather unexpected, as a final rise in F0 is one of the cues associated with the perception of question (Lieberman, 1967; Gussenhoven and Chen, 2000; House, 2003). Yuan (2004b) suggested that the intonation contour of the questions ending with a rising tone may not have been distinctive enough from the contour of statements ending with a rising tone, because questions and rising tones share a similar (rising) F0 contour. Therefore, listeners mistakenly identified questions ending with rising tones as statements. This pattern of results would be expected if listeners are biased toward the perception of statements, as suggested by Yuan (2004b) and Peters and Pfitzinger (2008), although no direct evidence of such bias was provided.

There are reasons to expect that the effects of lexical tone identity on the marking of questions and statements in Cantonese differ from those observed in Mandarin. First, the tonal systems of Mandarin and Cantonese differ in the number and in the F0 patterns of tones. Mandarin has four tones with distinctive F0 contours (level, rising, falling–rising, and falling). Cantonese has six contrastive tones: 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 pitch level at the beginning and the end point of the tone; Chao, 1947). While the four Mandarin tones have distinctive F0 patterns, five of the six Cantonese tones have similar F0 contours but different F0 levels (i.e., the three level tones and the two rising tones). The effects of lexical tones on the perception of intonation questions in Cantonese are likely to be based on both F0 level and contour. For example, Ma *et al*. (2006a) found that questions ending with a high-rising tone have a higher F0 peak than questions ending with low-rising or high-level tones. One might predict that intended questions ending with a high-rising tone are more likely to be perceived as questions than intended questions ending with other tones. Second, there is evidence that questions and statements are marked differently in Cantonese and in Mandarin. The command–response model describes the F0 contour of intonation as a linear superposition of a phrase component and an accent component on a base level (Fujisaki and Hirose, 1984). Using this model, the question–statement contrast in Cantonese can be modeled by local F0 changes (that is, the addition of a question boundary tone at the end of the final syllable of questions) (Gu *et al*., 2005, 2006; Ma *et al*., 2006a). In contrast, both Yuan (2004a) and Liu and Xu (2005) argued that the difference between questions and statements in Mandarin could be modeled by global F0 changes, such as an overall increase in F0, because local (tone) structures are largely unaffected by the intonation of questions in Mandarin (Ho, 1977; Rumjancev, cited in Lyovin, 1978; Shen, 1989). It is possible that intended questions impose different changes on the F0 contours of tones at the final position in Cantonese and in Mandarin, and that listeners are sensitive to language-specific cues for the perception of questions and statements. For example, a rising F0 contour of tones at the final position of sentences may be a more reliable cue to the perception of questions in Cantonese than in Mandarin, because all tones at the final position of intended questions have a rising F0 contour in Cantonese but not in Mandarin. By contrast, the increase in overall F0 level may be a stronger cue in Mandarin than in Cantonese because questions appear to be mainly marked by global changes in Mandarin. These predictions are consistent with the hypothesis that listeners are able to make optimal use of the probability distribution of acoustic cues that define a speech category within a language (Clayards *et al*., 2008).

B. Aims of the study

This study included both perceptual and acoustic analyses of intonation questions and statements in Cantonese. The first goal of this study was to investigate the acoustic cues that are used by listeners for the perceptual identification of intonation questions and statements in Cantonese. A final rise in F0 is generally regarded as a significant cue for the identification of questions across languages (Lieberman, 1967; Gussenhoven and Chen, 2000; House, 2003). By contrast, the role of the non-final portion of a sentence in the perception of questions, which is generally assumed to be less important than the cues at the final position, is rarely discussed. The fact that F0 cues at the final position of a sentence are also used for the perception of lexical tones might result in perceptual confusions. It is unknown to what extent a single F0 pattern can provide reliable cues to the perception of intonation questions and statements, as well as to the perception of lexical tones. Therefore, the present study employed several F0-based measurements in order to (i) identify the most likely cues used by listeners to perceive sentences as statements or questions and (ii) determine how these cues might differ from the cues that are used for the identification of lexical tones. Ma *et al*. (2006a,b) found that tones at the final position for questions had higher F0 level overall than the same tones in statements. Therefore, the average F0 of the final syllable was measured as a potential cue to the perception of statements and questions. Measures of F0 interval and F0 slope of the final syllable were included as potential cues to the perception of questions and statements, following Peters and Pfitzinger's (2008) findings. Duration has also been proposed as a cue to the perception of questions in Mandarin. For example, Ho (1977) and Yuan (2004a) reported that questions have a longer final syllable duration than statements. The use of alternative acoustic cues in the perception of questions and statements is possible because listeners have been found to trade the use of one cue (a delayed F0 peak within sentence) for another cue (a higher F0 peak) for the perception of questions (Gussenhoven, 2002). For example, duration could be used to disambiguate conflicting F0 information, such as a rising F0 pattern that could in principle signal either a rising tone within a statement or a level tone within a question. For this reason, voiced segment duration was also analyzed in this study. These acoustic measures were subsequently considered as predictors in a logistic regression analysis, so as to determine (i) which acoustic variables better explain the perception of questions and statements and (ii) whether such variables differ from the acoustic cues that are employed for the perception of lexical tones in Cantonese.

In the perception of intonation questions, the perceptual cues at the final position of an utterance have received more attention than those in non-final position, as shown by the number of studies focusing on the acoustic cues of questions at sentence final position (e.g., Lieberman, 1967; House, 2003). However, the relative significance of the perceptual cues at the final versus non-final portions of a sentence has not been directly compared. This could be particularly important in tonal languages, where a conflict of cues at the final position of a sentence could possibly lead to the use of acoustic cues in the non-final position. Therefore, the second goal of this study was to investigate the perception of intonation of statements and questions in three experimental conditions: complete sentences, carriers (i.e., sentences without the final syllable), and isolated final syllables. Previous studies based on the command–response model showed that the contrast between questions and statements in Cantonese is mainly cued by the final syllable (Gu *et al*., 2005, 2006; Ma *et al*., 2006a). Therefore, it was hypothesized that listeners would be equally accurate in identifying the intonation of the complete sentences and of the isolated final syllables, and that they would perform more poorly for the carriers. In the perception of questions and statements, statements have been assumed to be a default choice when the acoustic cues for intonation identification are ambiguous (Yuan, 2004b; Peters and Pfitzinger, 2008), but no direct evidence has been provided for such conclusion. Therefore, measures of sensitivity and bias based on signal detection theory were used to analyze the perceptual results, and to test empirically whether listeners have a perceptual bias toward the perception of statements. In addition, this study also aimed at further investigating the interaction between lexical tone and intonation. In the current perceptual experiment, listeners identified sentences that included all six contrastive Cantonese tones. In Mandarin, questions ending with falling tones were more accurately identified than questions ending with rising tones (Yuan, 2004a,b; Yuan and Shih, 2004). By contrast, Cantonese listeners face the challenge of distinguishing tonal rising F0 patterns within statements from intonational rising F0 patterns in questions (Ma *et al*., 2006b). The outcomes of the perception experiment were used to determine the extent to which listeners were able to deal with conflicting tonal and intonational cues.

II. METHOD

A. Listeners

Twenty-four females were recruited as listeners. They were all first year undergraduate students in the Division of Speech and Hearing Sciences at the University of Hong Kong. They were between 18 and 19 years old. They were considered naïve listeners as the experiment was carried out within the first 2 months of their enrollment at university, during which they received no phonetic training on tones or intonation. Cantonese was the native language and English was the second language of all listeners. None of the listeners had a reported history of speech or hearing disorders, and all listeners passed a pure-tone hearing screening conducted by the first author [$\leq$20 dB hearing level (HL) at 250, 500, 1000, 2000, and 4000 Hz for both ears].

B. Speech materials

Speech materials were part of the data collected by Ma *et al*. (2006b). A carrier with tonal contrast at the final position, /lei$_{55}$ kɔ$_{33}$ tsi$_{22}$ hɐi$_{22}$ X/ "This word is X," was used in this study. Three sets of target words (X) were obtained by combining the roots /si/, /ji/, and /jɐu/ with each of the six

contrastive Cantonese tones (tone 55, tone 25, tone 33, tone 21, tone 23, and tone 22), giving a total of 18 target words (see Appendix). Eighteen complete sentences were formed by combining the carrier with each of the target words. Each sentence was produced once as a question and once as a statement. Speech data from two native Cantonese speakers, one male (M1) and one female (F1), were used in this study. The utterances from these speakers were selected out of those produced by 20 speakers who participated in the study by Ma *et al.* (2006b), because the F0 patterns of their utterances were closest to the average F0 patterns for speakers of the same gender. With two intonation patterns (questions and statements), and three sets of six tonal contrasts, a total of 36 sentences from each speaker were used as the basic stimuli for the present study.

The utterances were recorded in a sound-attenuated room (IAC single-wall booth), with a Sony TCD-D3 DAT recorder and a Bruel & Kjær (4003) low-noise unidirectional microphone. A 10 cm mouth-to-microphone distance was maintained during the recording. In order to obtain naturally produced speech samples, speakers were engaged in a question–answer or statement–question dialog in which the first author initiated each exchange, and speakers answered with one of the sentences. The question–answer and statement–question dialogs were the same for both speakers to ensure that the same paralinguistic meaning would be encoded in the speakers' response for all questions or statements elicited. For example, in order to elicit a question, the first author would say /lei_{55} $kɔ_{33}$ tsi_{22} $hɐi_{22}$ si_{22}/ (This word is "yes."), and the speaker was asked to respond with /lei_{55} $kɔ_{33}$ tsi_{22} $hɐi_{22}$ si_{22}/ (This word is "yes"?). In each trial, the dialogs were presented visually on the screen of a G4 Apple Macintosh computer running a custom HYPERCARD (Apple™) software. The order of the dialogs was randomized automatically by the software. Each production was monitored by the first author, who is a qualified speech and language pathologist, to ensure that the correct intonations and tones were 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.

Three presentation conditions (complete sentences, carriers, and isolated final syllables) were used to investigate the perception of questions and statements, and the relative significance of different acoustic markers. In the complete sentence condition, the stimuli consisted of the original utterances produced by speakers M1 and F1. As all possible perceptual cues for intonation identification were present in complete sentences, it was expected that the stimuli in this presentation condition would be identified most accurately. In the other two presentation conditions, only part of the utterance (the carrier or the final syllable) was presented. The speech stimuli for the carrier condition (that is, the original sentence without the final syllable) and for the isolated final syllable condition were manually extracted from the complete sentences by using the PRAAT software (Boersma and Weenink, 2003). The boundary between the end of the carrier and the beginning of the final syllable was selected by inspecting both the amplitude waveform display and the corresponding wide-band spectrogram; all cutting points were at zero crossings to avoid introducing transients in the edited stimuli. For each presentation condition, the perceived loudness of the stimuli was equalized by the first author by modifying the overall amplitude of each token as necessary using the PRAAT software (Boersma and Weenink, 2003). The intensity level of the final syllables of the equalized stimuli was measured by calculating the mean power (Pa^2/s) between the beginning and the end of each voiced segment using the PRAAT software (Boersma and Weenink, 2003). All the final syllables were within the intensity range of 62.13–73.40 dB sound pressure level (SPL), with similar intensity ranges among the six tones.

C. Procedures

The perceptual experiment was carried out in a sound-attenuated room (IAC single-wall booth). Speech materials were presented through a Sennheiser HD 545 headset, connected to an Apple Macintosh G4 computer, with an Aardvark USB 3 sound card. A custom experimental software, developed using HYPERCARD (Apple™) software, was used to present the stimuli and to collect the responses in each trial. The stimuli were divided into six blocks according to presentation condition (complete sentences, carriers, and isolated final syllables) and speaker. Within each block, there were a total of 108 trials, as each of the 36 stimuli was presented three times. The order of presentation of blocks was counterbalanced by stimulus type and speaker across listeners, and the order of trials within each block was randomized. Task instructions were displayed visually on the screen and repeated verbally by the first author before the task began, explaining the nature of the stimuli for each presentation condition. Listeners were also instructed that the stimuli were produced as questions or statements, differing in their intonation contours. For each trial, two buttons labeled "question" and "statement" were presented on the screen. The listeners were asked to identify the intonation of each stimulus by clicking on one of the buttons. In cases when sentence intonation could not be easily identified, listeners were asked to make their best possible judgment. Within a trial, a stimulus was presented once, after which listeners could opt to listen to the stimulus a second time by clicking on the "repeat sound" button. Each block took about 10 min to finish, and the whole session took about 1 h to complete.

D. Acoustic analysis

Acoustic analyses of the F0, duration and intensity patterns of carriers and final syllables were conducted in order to estimate the perceptual cues that were used by the listeners in identifying intonation. Analysis of the complete sentence stimuli was not included, as previous studies showed that listeners used primarily cues at the final position of a sentence in intonation identification (Lieberman, 1967; Gussenhoven and Chen, 2000; House, 2003). For each carrier, the F0 of each word was measured at nine evenly spaced time points from the beginning to the end of the voiced segment of the word, which was identified visually from a wide-band spectrogram and an amplitude waveform display. The autocorrelation algorithm of the PRAAT software (Boersma and Weenink, 2003)

was used for the F0 estimates, with the pitch range set to 80–200 Hz for speaker M1, and to 150–400 Hz for speaker F1. F0 values at the time points of 0%, 25%, 50%, 75%, and 100% of the voiced segment were used for subsequent analysis, as our previous acoustic study showed that five measurement points were sufficient for an accurate representation of the F0 pattern within a syllable (Ma *et al.*, 2006b). The average F0 value for each carrier was then obtained by calculating the mean of the F0 values of the voiced segments of the four words in the carrier. Additionally, the F0 maxima and minima of each carrier stimuli were also measured using the "get maximum pitch" and "get minimum pitch" functions of the PRAAT software. The F0 range of each carrier was then calculated as the difference between the F0 maximum and minimum. The total duration of the carrier was obtained by calculating the difference in time between the 0% measurement point of the first word and the 100% measurement point of the final word of the carrier. The mean intensity value of each word in the carrier was measured from the beginning to the end of the voiced segment by using the "mean power" function of the PRAAT software (Boersma and Weenink, 2003). The overall intensity level of the stimuli was not compared as the loudness of the carriers was equalized before the experiment. Instead, the intensity variation within each carrier was analyzed. First, it was empirically determined that 99% of the carriers had the peak intensity level at one of the first three syllables, and that 84% of the carriers had the minimum intensity level at the fourth syllable (i.e., the last syllable in the carrier). Therefore, the ratio between the peak intensity level of the carrier and the intensity level of the fourth syllable was calculated for each carrier, in order to represent the amount of intensity variation within the carrier.

For the final syllable, the average F0 value was calculated by averaging the F0 values at the 0%, 25%, 50%, 75%, and 100% time points of the voiced segments as described above. Additionally, as F0 changes within the final syllable are likely to be affected by both the intended intonation and lexical tone (Ma *et al.*, 2006b), F0 interval and F0 slope measures were included in order to represent the degree of F0 change. The F0 interval was calculated as the semitone (ST) difference between the 0% and the 100% time points ($F0_{100\%} - F0_{0\%}$). The F0 slope (ST/s) was estimated using the formula: F0 slope $= 12/\log\ 2 \times [(\log F0_{100\%} - \log F0_{0\%})/(\text{time}_{100\%} - \text{time}_{0\%})]$ (Peters and Pfitzinger, 2008). The duration of the voiced segment was measured by calculating the difference in time between the 0 and 100% time points. The mean intensity of the final syllable was not analyzed as the loudness level of each syllable was equalized before the experiment. Although an intensity ratio could have been calculated for final syllables (for example, the ratio between the peak intensity level of the carrier and the level of the final syllable) such ratios were not analyzed because they were not representative of the stimuli that were presented to the listeners in this condition.

III. RESULTS

A. Identification accuracy

Identification accuracy for each stimulus type was calculated as the percentage of responses identified as the intended intonation, as a function of presentation condition (complete sentence, carrier, and isolated final syllable), intonation (question and statement), and lexical tone (tone 55, tone 25, tone 33, tone 21, tone 23, and tone 22). For example, for statements with tone 25 at the final position presented in complete sentence condition, the intended intonation (statement) was correctly identified in 378 out of 432 trials, resulting in an identification percentage of 87.50%. Average identification accuracy data are displayed in Fig. 1. This figure shows that accuracy was lower for the carrier than for the complete sentence and the final syllable conditions. Additionally, these data also show that accuracy was overall lower for questions than for statements.

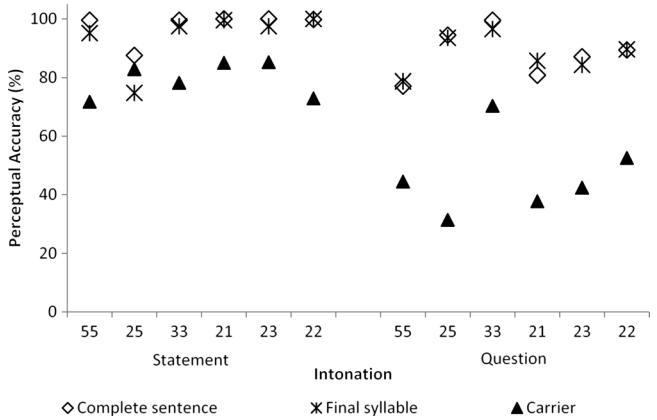

FIG. 1. Mean percentages of correct identification for each presentation condition and each tone.

B. Sensitivity and response bias in the perception of statements and questions

The data reported in Fig. 1 suggest that statements were perceived with higher identification accuracy than questions for all presentation conditions, and particularly for the carrier condition. However, the percentage of correct responses is not an unbiased measure of the ability of listeners to identify statements and questions. For example, a listener who indiscriminately identifies all stimuli as "statements" would reach ceiling performance for intended statements, and yet such a strategy could hardly be considered as representing highly accurate performance. In fact, it is possible that "statement" is the default choice in intonation perception, and that listeners would only identify a stimulus as "question" when there is strong evidence (such as a final rise in F0) against the presence of a statement (Yuan, 2004b; Peters and Pfitzinger, 2008). For this reason, the responses were analyzed by using the theoretical framework of signal detection theory (see Macmillan and Creelman, 2005, for a review). *Hits* were defined as the proportion of "statement" responses for intended statements; *false alarms* were calculated as the proportion of "statement" responses for intended questions. Measures of sensitivity (d') and response bias (c) were calculated from the proportion of hits and false alarms using formulas for classification experiments with two stimuli and two responses (Macmillan and Creelman, 2005). The mean d' and c values for each tone in each presentation condition are summarized in Figs. 2 and 3, respectively.

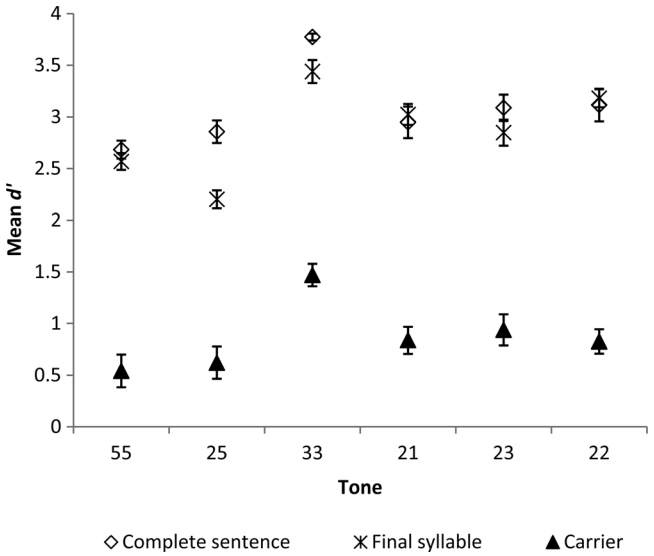

FIG. 2. The mean measure of sensitivity (d') with the standard error bars for each of the six tones in the complete sentence, final syllable, and carrier conditions.

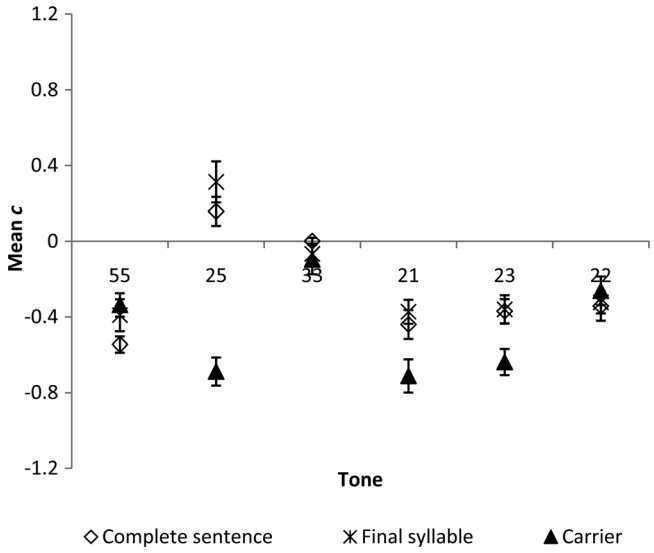

FIG. 3. The mean measure of response bias (c) with the standard error bars for each of the six tones in the complete sentence, final syllable, and carrier conditions.

The measures of d' and c were separately analyzed using a two-way analysis of variance (ANOVA) with repeated measures. For both analyses, the factors were the presentation condition (complete sentence, carrier, and final syllable) and the lexical tone (six contrastive Cantonese tones). Statistical analysis on the measure of sensitivity (d') showed significant differences between the three presentation conditions, $F(2,46) = 327.88$, $p < 0.001$. *Post-hoc* analyses showed that sensitivity was significantly lower for carriers (mean = 0.87) than for both complete sentences (mean = 3.08) and final syllables (mean = 2.88) [Tukey honestly significant difference (HSD) tests, $p < 0.001$ for both comparisons]. Sensitivity was also significantly different among tones [main effect of lexical tone, $F(5,115) = 27.90$, $p < 0.001$]. *Post-hoc* analysis showed that d' was higher for tone 33 (mean = 2.90) than for all other tones (tone 55 mean = 1.93; tone 25 mean = 1.89; tone 21 mean = 2.27; tone 23 mean = 2.29; tone 22 mean = 2.38) (Tukey HSD tests, $p < 0.001$ for all). Additionally, the d' values for tones 55 and 25 were significantly lower than those of the other four tones (Tukey HSD tests, $p < 0.01$ for all); no significant difference was found between tones 55 and 25 (Tukey HSD test, $p > 0.05$). Analysis of the interaction effect between presentation condition and lexical tone showed that the effect of tones differed among presentation conditions, $F(10,230) = 2.97$, $p < 0.001$. Listeners showed similar sensitivity in the perception of complete sentences and isolated final syllables for all tones (Tukey HSD tests, $p > 0.05$ for all) except tone 25. For this tone, sensitivity was higher for complete sentences than for final syllables (Tukey HSD test, $p < 0.001$). Carrier stimuli were perceived with lower sensitivity than complete sentences and final syllables, regardless of tone (Tukey HSD tests, $p < 0.001$ for all). Values of d' for all carrier stimuli, except for those with tone 33, were lower than 1, indicating poor sensitivity in the carrier condition overall.

A two-way ANOVA for the measure of bias c showed a significant main effect for presentation condition, $F(2,46) = 7.60$, $p < 0.001$. *Post-hoc* analysis showed that c was significantly lower for the carrier condition (mean = −0.46) than for complete sentences (mean = −0.26) and final syllables (mean = −0.20) (Tukey HSD tests, $p < 0.005$ for both comparisons). That is, listeners showed a larger response bias toward the choice of statements for carrier stimuli than for the other two presentation conditions. The main effect of lexical tone showed a significant difference among the six tones, $F(5,115) = 32.99$, $p < 0.001$. Listeners showed a significantly higher response bias toward "statement" responses for tones 55 (mean = −0.42), 21 (mean = −0.51), 23 (mean = −0.46), and 22 (mean = −0.31) than for tones 25 (mean = −0.07) and 33 (mean = −0.05) (Tukey HSD tests, $p < 0.001$ for all). Additionally, a significant difference in c was found between tones 21 and 22 (Tukey HSD test, $p < 0.005$). The interaction between presentation condition and lexical tone was significant, $F(10,230) = 24.06$, $p < 0.001$. *Post-hoc* analysis showed a similar bias for each tone between the complete sentence and the final syllable conditions (Tukey HSD tests, $p > 0.05$ for all). While no significant differences were found for tones 55, 33, and 22 between the carrier condition and the other two presentation conditions (Tukey HSD tests, $p > 0.05$ for all), significantly larger negative values of c were found for tones 25, 21, and 23 for carriers than for complete sentences or final syllables (Tukey HSD tests, $p < 0.05$ for all). Overall, the negative values observed in most conditions (exceptions were the complete sentences and the final syllables with tone 25, and all stimuli with tone 33) supports the idea of a bias toward the perception of statements.

C. Relationship between perceptual and acoustic measures of intonation

1. The carrier

Four potential acoustic cues to the perception of questions and statements were evaluated in the carrier stimuli: average F0, F0 range, duration, and intensity ratio. Owing to the

difference in F0 range in male and female speakers, the measurements of average F0 of speakers M1 and F1 were normalized by subtracting the speaker's average F0 from the F0 measurements in ST (Fant, 2004). The differences between questions and statements were compared for each acoustic cue using Wilcoxon matched-pair tests. The average normalized F0 was significantly higher for questions [mean = 0.34 ST, standard deviation (SD) = 0.65] than for statements (mean = −0.34 ST, SD = 0.50) ($T = 79$, $p < 0.001$). No statistically significant difference was observed between the F0 range of questions (mean = 11.23, SD = 1.11) and statements (mean = 10.79, SD = 1.29) ($T = 240$, $p > 0.05$). In terms of duration, the carrier of statements (mean = 667, SD = 101.66) had an average duration significantly longer than the duration of the carriers of questions (mean = 623, SD = 83.95) ($T = 79$, $p < 0.001$). A significantly higher intensity ratio was found for questions (mean = 0.95, SD = 0.02) than for statements (mean = 0.93, SD = 0.02) ($T = 122$, $p < 0.001$).

The potential use of these acoustic variables in differentiating questions from statements was explored using logistic regression analysis. All four variables (average F0, F0 range, intensity, and duration) were included as the predictors of a forward-stepwise logistic regression; the intended production was the dependent variable. Table I summarizes the results of the logistic regression analysis ($\chi^2 = 30.40$, $df = 2$, $N = 72$, $p < 0.001$). The prediction of the regression model showed that questions and statements could be classified by the combination of average F0 ($\chi^2 = 12.56$, $p < 0.001$) and intensity ($\chi^2 = 6.62$, $p < 0.05$). The contribution of F0 range and duration to the classification of questions and statements was not significant in this model ($p > 0.05$). The classification matrix of questions and statements based on this model showed an overall classification accuracy of 76%. The model accurately classified 72% of the intended questions and 81% of the intended statements. The predicted accuracy of this model was higher than the identification accuracy of the listeners (48% for questions and 79% for statements), suggesting that the listeners were not able to make optimal use of average F0 and intensity in the identification task.

In order to determine the role of the acoustic cues on the perception of questions and statements, a second logistic regression was conducted with the four acoustic variables as the predictor and the "perceived identity" (question, statement, or ambiguous) of each stimulus as the dependent variable. The perceived identity of each token was determined by using a binomial test ($N = 72$, $p = 1/2$, $\alpha = 0.05$). The results of a multinomial logistic regression analysis ($\chi^2 = 81.56$, $df = 2$, $N = 72$, $p < 0.001$) showed that average F0 was a significant predictor ($\chi^2 = 8.93$, $p < 0.05$). F0 range, intensity ratio and duration were not significant predictors in this model ($p > 0.05$). This model provided an accurate estimate of the perceived identity of the stimuli, as it correctly classified 89% of the stimuli perceived as questions, and 98% of the stimuli perceived as statements. However, this model correctly classified only 33% of the ambiguous stimuli.

As only average F0 was the significant predictor of the perceptual data, an additional logistic model with the intended intonation as dependent variable was calculated with average F0 as the sole predictor. The logistic regression model was statistically significant ($\chi^2 = 22.88$, $df = 1$, $N = 72$, $p < 0.001$), and showed that average F0 was a significant predictor of the intended intonation ($\chi^2 = 13.81$, $p < 0.001$). This model had an overall classification accuracy of 74% (69% for questions and 78% for statements). The predicted accuracy of this model was similar to the identification accuracy of statements (79%) but was still higher than the identification accuracy of questions (48%).

2. The final syllable

The perception of statements and questions for final syllables was explored by examining four intonation cues: average F0, F0 slope, F0 interval, and duration. Figures 4(a)–4(d) displays the F0 contour of the six Cantonese tones produced by speakers M1 and F1 at the final position of intended statements and questions. A rising F0 contour was found for all tones at the final position of questions, while the canonical forms were maintained at the final position of statements. A Wilcoxon matched-pair test was used to compare each of the intonation cues between questions and statements. A significantly higher normalized average F0 was observed for final syllables of questions (mean = 1.79 ST, SD = 2.29) than for those occurring at the end of statements (mean = −1.79 ST, SD = 2.73) ($T = 0$, $p < 0.001$). For F0 interval, questions (mean = 6.58 ST, SD = 3.14) showed larger positive interval values (i.e., an increase in F0 from the beginning to the end of the syllable) than statements (mean = 0.61 ST, SD = 4.45) for all tones ($T = 9$, $p < 0.001$). Of note, the F0 interval for tones 25 (6.73 ST for speaker M1 and 6.90 ST for speaker F1) and 23 (1.47 ST for speaker M1 and 3.50 ST for speaker F1) produced in statements was within the range of the F0 interval of questions (ranging from 0.36 to 12.85 ST). A significantly larger positive F0 slope was found for questions (mean = 18.95 ST/s, SD = 8.89) than for statements (mean = 0.39 ST/s, SD = 16.59) for all tones ($T = 16$, $p < 0.001$). The slope of the two rising tones in statements (tone 25 = 22.87 ST/s for speaker M1 and 20.50 ST/s for speaker F1; tone 23 = 4.29 for speaker M1 and 11.36 ST/s for speaker F1) was within the range of the F0 slope of the final syllable in questions for all six tones (ranging from 1.48 to 38.55 ST/s). These results suggest that the perception of

TABLE I. The results of the logistic regression analysis in the carrier for questions and statements in Cantonese.

Predictor	β	SE β	Wald's χ^2	df	p	e^{β}
Constant	−35.05	13.65	6.59	1	0.000	0.00
Average F0	2.23	0.63	12.56	1	0.010	9.26
Intensity	37.31	14.50	6.62	1	0.010	1.590E16

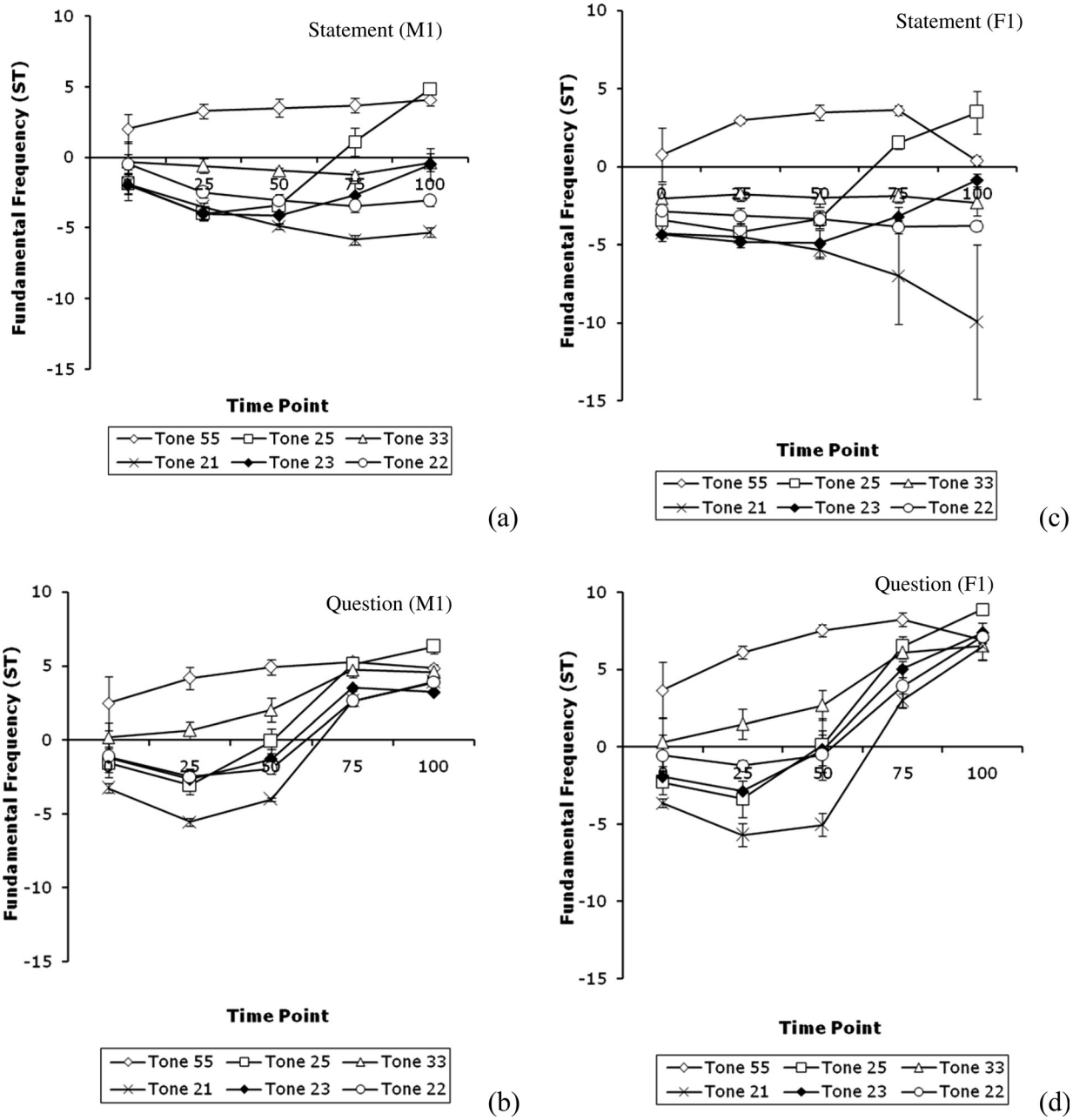

FIG. 4. Normalized F0 contour (in STs) for each lexical tone with standard error bars are displayed for (a) statements by speaker M1, (b) questions by speaker M1, (c) statements by speaker F1, and (d) questions by speaker F1.

questions and statements cannot be explained by reference to a simple perceptual boundary along the F0 interval or F0 slope continua, independent of lexical tone.

An analysis of the average duration data showed that final syllables produced in questions (mean = 340 ms, SD = 50) had a longer duration than those produced in statements (mean = 299 ms, SD = 65) ($T = 91.5$, $p < 0.001$). All six tones produced by speaker M1, and four of the six tones (tones 25, 33, 23, and 22) produced by speaker F1 had longer syllable duration in questions (ranging from 6 to 86 ms) than in statements, but speaker F1 produced longer syllable duration in statements for tones 55 (1 ms) and 21 (15 ms). Both Lehiste (1970) and Bochner *et al.* (1988) reported a difference limen of about 40 ms for a standard signal of about 300 ms. A comparison of the duration of the final syllable within the 36 statement–question pairs (i.e., the same utterance produced once as a statement and once as a question by the same speaker) showed that 61% of these pairs differed by less than the 40-ms difference limen. The final syllable of questions was 46–263 ms longer than its counterpart in statements in the remaining 39% statement–question pairs.

Logistic regression analysis was used to determine the relationship between the acoustic variables of the final syllable and the production of questions and statements in Cantonese. Average F0, F0 interval, F0 slope, and duration were included as the predictors in a forward-stepwise logistic regression. Because 96% of the final syllables were accurately identified as questions or statements (binomial test; $N = 72$, $p = 1/2$, $\alpha = 0.05$), the logistic regression model was calculated by using the intended productions as the dependent variable, with

statement coded as 0 and question as 1. The results of the logistic regression analysis are summarized in Table II. The regression model was statistically significant ($\chi^2 = 61.49$, $df = 2$, $N = 72$, $p < 0.001$), and it predicted the classification of questions and statements by the combination of F0 interval ($\chi^2 = 11.69$, $p < 0.005$) and average F0 ($\chi^2 = 10.61$, $p < 0.005$). F0 slope and duration did not make a statistically significant contribution to the classification of questions and statements in this model ($p > 0.05$). The exponential regression coefficients show that the odds of classifying an utterance as a question increased by 2.36 times for each 1 ST increase in average F0, and that the odds of "question" responses increased by 1.98 times for each 1 ST increase in F0 interval. The classification matrix of questions and statements based on this model shows that the model's overall classification accuracy was about 86%, and that 89% of intended questions and 83% of intended statements were classified correctly.

IV. DISCUSSION

A. Acoustic cues for the identification of intonation questions and statements in final syllables

Overall, the results showed that listeners had equally high sensitivity in the identification of statements and questions when presented with complete sentences or with the isolated final syllable of an utterance. However, sensitivity was significantly reduced for carrier stimuli. These findings are in agreement with those of Lieberman (1967) and Gussenhoven and Chen (2000), who showed that listeners mostly rely on the perception of the final syllable within a sentence for identification of intonation questions. These results also support the hypothesis that the perception of the intonation of questions in Cantonese is mainly based on the perception of local F0 changes.

The contribution of four acoustic variables (average F0, F0 interval, F0 slope, and duration) at the final syllable to the identification of statements and questions in Cantonese was determined using logistic regression analysis. The resulting model correctly classified about 86% of the stimuli, which was close to the accuracy achieved by the listeners in the identification task (91%). Nonetheless, this discrepancy indicates that there is a remaining amount of variance in the listener's responses that cannot be accounted for by the variables measured in the present study. The results of the logistic regression showed that the production of questions and statements could be classified by the combination of F0 interval and average F0 of the final syllable, but not by F0 slope and duration. The increased likelihood of question identification associated with an increase in average F0 was not unexpected, as House (2003) found that differences in F0 level could be used as a reliable cue for the identification of questions. The relative significance of the F0 interval as a perceptual cue for questions has also been previously discussed. Peters and Pfitzinger (2008) suggested that a constant change in F0 interval was important for correct identification of questions. They suggested that an F0 interval larger than 2 ST could be used as a reliable cue to the perception of intonation questions. Our findings are consistent with this proposal, as we observed F0 intervals larger than 2 ST at the final syllable for 94% of the intonation questions.

F0 slope was not a significant predictor for the identification of questions in this study. This result is consistent with the findings of Peters and Pfitzinger (2008), who concluded that F0 slope is not likely to be a reliable cue because its effect is dependent on the duration of the voiced segments. The use of F0 slope to describe the rate of F0 changes at the final syllable in the present study was also complicated by the fact that the F0 contour of Cantonese tones is not always linear. For example, as shown in Figs. 4(b) and 4(d), the F0 contour of the low-falling tones begins with a slight dip followed by rising F0 toward the end. An investigation of the F0 shape of the tone and the classification of questions and statements might be interesting, as recent studies suggest that intonation contrast is marked not only by a rising–falling F0 contour and by overall F0 level but also by the curvature of the rising and falling F0 contours of the final syllable. In investigating the expression of turn-yielding and turn-holding communicative functions in German, Dombrowski and Niebuhr (2005) found that a convex F0 rise occurs at the phrase-final position for turn-yielding questions, and that a concave rise was associated with the communicative function of turn-holding. Similar findings were observed in Estonian (Asu, 2006). Although the current study did not directly explore the shape of the F0 contour at the final position of questions and statement, a visual inspection of F0 patterns in Figs. 4(b) and 4(d) indicates that most of the final rises in questions were associated with a concave shape. The lack of a clear association between F0 shape and intonation in the data collected in the present study might be related to the interaction between the canonical F0 patterns of the Cantonese tones and intonation. A more systematic investigation of this association in lexical tone languages such as Cantonese might prove informative about additional F0 cues that might enable speakers to accurately perceive the intonation of questions.

Duration was not a significant predictor for the classification of questions and statements in this study. This outcome is consistent with the observation that the duration of the final syllables of questions and statements was smaller than the difference limen in 61% of the statement–question pairs, although the statistical analysis showed that the duration of the final syllable of questions was significantly longer than that of statements. In a study of dysarthric speech in English, Patel (2003) observed that, when the F0 contour of a sentence was preserved, listeners were able to correctly identify

TABLE II. The results of the logistic regression analysis in the final syllable for questions and statements in Cantonese.

Predictor	β	SE β	Wald's χ^2	df	p	e^{β}
Constant	−3.46	1.16	8.86	1	0.003	0.03
F0 interval	0.68	0.20	11.67	1	0.001	1.98
Average F0	0.86	0.26	10.61	1	0.001	2.36

sentence intonation even though the duration of the final syllable in questions and statements was equalized. Her findings support the idea that duration of the final syllable does not play a major role in the identification of questions and statements. However, it is possible that duration could act as secondary cue when the primary F0-related cues are not available. For example, by comparing the perception of questions in phonated speech and whispered speech, Heeren and van Heuven (2009) found that duration was a more important cue to question perception in whispered speech than in phonated speech.

The results of the present study showed that, in the identification of intonation questions, listeners extract salient F0 features from the F0 pattern of the final syllable, such as the average F0 and the F0 interval. The possibility that different F0 cues could be extracted from a single F0 pattern provides an explanation for how the listeners could cope with the co-existence of intonation and tone in tonal languages. Gandour and Harshman (1978) and Gandour (1981) found that the canonical F0 patterns of lexical tones in Cantonese are perceived according to a number of perceptual dimensions (e.g., average pitch, direction of F0 movement, length, F0 value at the end point, and F0 slope). Gandour (1981) found that the F0 slope and direction accounted for 74% of the correct tone identification in Cantonese. Tone height also made a small but significant contribution to tone identification. Using a discriminant analysis, Khouw and Ciocca (2006) investigated the relative contribution of five different cues (average F0, F0 change at 25%, 50%, 75%, and 100% of the vocalic segment) to the perception of lexical tones. Their results generally agreed with those of Gandour (1981) and highlighted the importance of the perceptual dimensions "direction of F0 changes" and "magnitude of F0 changes" in Cantonese tone perception. They found that the F0 change at the second half of the syllable accounted for about 89% of the variance in distinguishing the high-rising tone (tone 25), the low-rising tone (tone 23), the low-falling tone (tone 21), and the three level tones (tones 55, 33, and 22). The remaining 10% of the variance could be explained by the average F0, which contrasted the six Cantonese tones as four groups (tone 55; tones 25 and 33; tones 23 and 22; and tone 21). If F0 features are considered the basic units in the perception process, it is possible that different communicative functions, such as intonation and lexical tone, can be simultaneously perceived from the multi-layer information carried by a single F0 pattern. To summarize, the results of previous studies on Cantonese lexical tones indicate that F0 change and F0 direction are likely to be the most important cues to the perception of lexical tones. The present study suggests that F0 interval might be the most reliable cue to the perception of intonation questions. Average F0 cues appear to be involved in the perception of both lexical tones and intonation questions. Further studies on the simultaneous identification of question–statement contrast and lexical tone are needed in order to increase our understanding of the concurrent transmission of communicative functions through F0-based cues.

B. Acoustic cues for the identification of intonation questions and statements in carriers

Sensitivity was poor overall for carriers, compared with complete sentences or with the isolated final syllables. Values of the bias measure c showed that listeners were biased toward the perception of statements in most of the conditions of the present study. Additionally, this response bias was higher in the carrier condition than in the complete sentence and the isolated final syllable conditions. The larger bias in the perception of carriers suggests that, in the absence of strong evidence for the perception of interrogation (which is present in the final syllable), listeners are biased toward the perception of statements. This conclusion is consistent with previous proposals which suggested that the perception of statement is the default choice in intonation perception (Yuan, 2004b; Peters and Pfitzinger, 2008), and that listeners only perceive a question in the presence of clear evidence, such as a final rise in F0 (Yuan and Shih, 2004; Peters and Pfitzinger, 2008).

Although the present results showed that F0 cues in final syllable were most important for the identification of questions and statements, the question carrier with tone 33 was perceived with moderately good sensitivity (d' of about 1.5). This finding suggests that listeners can use of intonation cues in carriers for perceiving intonation questions, even though such cues are not as important as those present in final syllables. Four parameters (average F0, F0 range, duration, and intensity ratio) were used to explore the acoustic cues in the carrier portion of question–statement identification. While the predicted accuracy of statements (81%) was similar to the identification accuracy in the carrier-only condition (79%), the predicted accuracy for questions (72%) was higher than the listeners' identification accuracy (48%). The discrepancy between the predicted accuracy and the identification accuracy suggested that some of the statistically significant prediction might not have been perceptually salient to the listeners. Using only average F0 as the predictor of the intended intonation, a logistic regression model correctly predicted intended statements with 78% accuracy and intended questions with 69% accuracy. This model outperformed the listeners', indicating that listeners were not able to make optimal use of average F0 information to identify questions and statements.

It was not surprising that carriers with a higher F0 level were associated increase the likelihood of question responses. This global F0 level cue to intonation marking has been previously reported for Mandarin by several investigators (Ho, 1977; Yuan, 2004a; Liu and Xu, 2005;). Both Ho (1977) and Yuan (2004a) reported a higher intensity level for questions than for statements. The contrast in overall intensity level was not compared across carriers in the current study as the loudness level of the carriers was equalized before the experiment. The ratio of peak intensity level to the pre-final syllable intensity level was used instead. A consistently higher intensity ratio was observed for questions than for statements; although the difference was small (0.02), this acoustic variable made a significant contribution to the prediction of the intended intonation. However, the regression model based on perceived intonation showed that intensity ratio was not likely to be employed by listeners in distinguishing the question–statement contrast. Both F0 range and duration of the carriers were not significant predictors of the identification of questions. This may be related to the fact that the most prominent F0 changes in the marking of intonation occur at the final position. For duration, there was no conclusive evidence

of durational contrast in the carriers of questions and statements. In a previous study comparing the duration of the corresponding syllables in eight-syllable statements and questions, Yuan (2004a) found that syllable duration differs between the carrier portion of questions and statements in Mandarin sentences. By contrast, by measuring the duration of the duration of each syllable nuclear of the five-syllable stimuli, Ho (1977) argued that the most prominent duration contrast between Mandarin questions and statements was at the final portion of a sentence but not in the carrier.

C. The perception of intonation questions and statements as a function of lexical tone

In the present study, listeners were most accurate in distinguishing statements and questions for sentences containing tone 33 at the final position; sentences containing tones 55 and 25 were least accurately distinguished. The ability of the listeners to identify statements and intonation questions can be explained by examining the F0 patterns of tones in intended questions and statements. First, all tones at the final syllable of questions have a rising F0 contour regardless of their canonical form—see Figs. 4(b) and 4(d). A rising F0 contour provides a perceptual cue for listeners in differentiating statements from questions, as questions are associated with a high F0 peak (Remijsen and van Heuven, 1999; Gussenhoven and Chen, 2000; House, 2003; Schneider and Lintfert, 2003; Kohler, 2004), or with a difference of about 2 ST or more in the F0 interval between the onset and the offset of the final syllable (Peters and Pfitzinger, 2008; see also present results). However, the resemblance of the F0 contours of all tones at the final syllable of questions to the canonical F0 contour of the two rising tones (tones 25 and 23) could lead to perceptual confusions. There is evidence that such a confusion occurred in the present study, since tone 25 had overall lower sensitivity than all other tones, except tone 55. The low sensitivity for tone 25 can be explained by the fact that, by examining the identification accuracy in the complete sentence condition, 12.5% of the intended statements with tone 25 as the final syllable were perceived as questions. That is, the listeners confused the rising F0 contour of tone 25 with the final rise in questions. Interestingly, the rising F0 contour of tone 23 at the final position of intended statements was not confused with the rising intonation contour of questions. Instead, tone 23 generated similar d' values to those of other tones of a similar onset F0 level (tones 21 and 22). As the main difference between the two rising tones is the extent of the F0 rise, these results suggest that listeners do not perceive all sentences ending with a rising F0 as questions. Instead, listeners are able to exploit seemingly subtle differences in F0 characteristics in order to distinguish between intonation questions and statements. The reason for the relatively low sensitivity of stimuli containing tone 55 is likely due to the small difference in F0 contour between intended questions and statements for this tone, especially for stimuli produced by speaker M1 [Figs. 4(a) and (b)]. Specifically, two questions containing tone 55 produced by M1 in the complete sentence condition were identified with only 31 and 58% accuracy, while the third question containing tone 55 produced by speaker M1 was identified with 79% accuracy. In contrast, all questions containing tone 55 produced by speaker F1 were perceived with 97–100% accuracy. It is unclear why these tone 55 utterances were perceived with such low accuracy, given that the first author (a native speaker with extensive training in phonetic transcription) did not find these stimuli ambiguous. It is possible that the elicitation procedure employed in this study encouraged some speakers to simply repeat the elicitation sentence produced by the first author during the simulated dialogs. However, the presence of two atypical utterances is unlikely to affect the validity of the present findings, since all other utterances used in this study were accurately identified as the intended sentence type.

Previous reports on the effect of lexical tone on the perception of intonation in Mandarin suggested that the perception of statements was not affected by the identity of the tone at the final position (Yuan, 2004a,b; Yuan and Shih, 2004). The evidence from the current study shows that, in Cantonese, measures of sensitivity (d') and response bias (c) were affected by the identity of lexical tones. Similarly, Yuan (2004a,b) and Yuan and Shih (2004) found that Mandarin questions ending with the falling tone were easier to identify than those ending with the rising tone. Yuan (2004a) hypothesized that this difference is due to the fact that the F0 contour of the falling tone at the final position is flattened by a tone-dependent mechanism, making the F0 contour contrastive for intonation identification. In Cantonese, a rising F0 contour was observed for all six tones at the final position of questions. Listeners might have simply perceived all rising F0 contours in questions as rising tones within a statement, but they did not. Rather, Cantonese listeners must have been able to pay attention to F0 characteristics such as F0 interval and average F0 in order to perceive intonation questions and statements across the various tonal contexts.

VI. CONCLUSION

In summary, this study explored the question–statement contrast in a tonal language, Cantonese, using both acoustic and perceptual analyses. The results showed that the F0 patterns at the final syllable played the most significant role in the identification of questions, and the distinction between intonation and lexical tone could be associated with different perceptual dimensions. While the average F0 appears to be involved in both the perception of intonation questions and tone identification, the F0 interval of the final syllable was found to be a significant predictor for question identification. Previous studies found that the perception of tones relies primarily on the F0 changes that occur in the second half of the syllable. The outcomes of the perception experiment showed that the perception of intonation questions is affected by lexical tones, and that listeners have a general bias toward the perception of statements. This study contributes to the ongoing debate regarding potential conflicts between tone and intonation in lexical tone languages, by identifying differential perceptual cues that may be used by listeners in disambiguating overlapping F0 patterns. The present findings also demonstrated that the cues that are employed for the perception of intonation questions may differ among different lexical tone

languages, due to the differences in the number and type of F0 patterns that characterize the specific lexical tone systems.

ACKNOWLEDGMENT

The equipment used in this study was substantially supported by a grant from the Hong Kong Research Grants Council (Grant No.: 7224/03H). We would like to thank two anonymous reviewers for the helpful comments on an earlier draft of this manuscript.

APPENDIX: LIST OF TARGET WORDS

Target	Translation	Target	Translation	Target	Translation
/si$_{55}$/	Poem	/ji$_{55}$/	Clothes	/jɐu$_{55}$/	Rest
/si$_{25}$/	History	/ji$_{25}$/	Chair	/jɐu$_{25}$/	Pomelo
/si$_{33}$/	Try	/ji$_{33}$/	Meaning	/jɐu$_{33}$/	Thin
/si$_{21}$/	Time	/ji$_{21}$/	Child	/jɐu$_{21}$/	Swim
/si$_{23}$/	Market	/ji$_{23}$/	Ear	/jɐu$_{23}$/	Have
/si$_{22}$/	Yes	/ji$_{22}$/	Two	/jɐu$_{22}$/	Again

Asu, E. L. (**2006**). "Rising intonation in Estonian: An analysis of map task dialogues and spontaneous conversations," in *Proceedings of Fonetiikan Päivät 2006*, Helsinki, Finland, pp. 1–9.

Bauer, R. S., and Benedict, P. K. (**1997**). *Modern Cantonese Phonology* (Mouton de Gruyter, Berlin), Chap. 2, pp. 109–276.

Bochner, J. H., Snell, K. B., and MacKenzie, D. J. (**1988**). "Duration discrimination of speech and tonal complex stimuli by normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **84**, 493–500.

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, http://www.fon.hum.uva.nl/praat/ (Last viewed October 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), 242 p.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (**2008**). "Perception of speech reflects optimal use of probabilistic speech cues," Cognition **108**, 804–809.

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.

Dombrowski, E., and Niebuhr, O. (**2005**). "Acoustic patterns and communicative functions of phrase-final rises in German: Activating and restricting contours," Phonetica **62**, 176–195.

Fant, G. (**2004**). *Speech Acoustics and Phonetics* (Kluwer Academic Publishers, Netherlands), Chap. 6, pp. 221–300.

Fok-Chan, Y. Y. (**1974**). *A Perceptual Study of Tones in Cantonese* (University of Hong Kong Press, Hong Kong), 191 p.

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.

Gandour, J., and Harshman, R. (**1978**). "Cross language differences in tone perception: A multidimensional scaling investigation," Lang. Speech, **21**, 1–33.

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," in *Proceedings of Interspeech-2005*, Lisbon, Portugal, pp. 1825–1828.

Gu, W.-T., Hirose, K., and Fujisaki, H. (**2006**). "A comparative study between intonation question and particle question in Cantonese on their realization of F0 contours," in *Proceedings of the 2nd International Symposium on Tonal Aspects of Languages*, La Rochelle, France, pp. 63–66.

Gussenhoven, C. (**2002**). "Intonation and interpretation: Phonetics and phonology," in *Proceedings of the Speech Prosody-2002*, Aix-en-Provence, France, pp. 47–57.

Gussenhoven, C., and Chen, A.-J. (**2000**). "Universal and language-specific effects in the perception of question intonation," in *Proceedings of the International Conference on Spoken Language Processing-2000*, Beijing, China, pp. 91–94.

Heeren, W., and van Heuven, V. J. (**2009**). "Perception and production of boundary tones in whispered Dutch," in *Proceedings of the Interspeech 2009*, Brighton, UK, pp. 2411–2414.

Hirschberg, J., and Ward, G. (**1992**). "The influence of pitch range, duration, amplitude and spectral features on the interpretation of rise-fall-rise intonation contour in English," J. Phonetics **20**, 241–251.

Hirst, D., and Di Cristo, A. (**1998**). "A survey of intonation systems," in *Intonation Systems: A Survey of Twenty Languages*, edited by D. Hirst and A. Di Cristo (Cambridge University Press, New York), pp. 1–44.

Ho, A. T. (**1977**). "Intonation variation in a Mandarin sentence for three expressions: Interrogative, exclamatory and declarative," Phonetica **34**, 446–457.

House, D. (**2003**). "Perceiving question intonation: The role of pre-focal pause and delayed focal peak," in *Proceedings of the International Congress of Phonetic Sciences*, Barcelona, Spain, pp. 755–758.

Khouw, E., and Ciocca, V. (**2006**). "Acoustic and perceptual study of Cantonese tones produced by profoundly hearing-impaired adolescents," Ear Hear. **27**, 243–255.

Kohler, K. (**2004**). "Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions," 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. 201–215.

Kohler, K. (**2007**). "Beyond laboratory phonology—The phonetics of speech communication," in *Experimental Approaches to Phonology*, edited by M. J. Sole, P. S. Beddor, and M. Ohala (Oxford University Press, Oxford), pp. 41–53.

Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge University Press, Cambridge), Chap. 1, pp. 6–41.

Lee, W.-S. (**2004**). "The effect of intonation on the citation tones in Cantonese," in *Proceedings of the 1st International Symposium on the Tonal Aspects of Languages*, Beijing, China, pp. 107–110.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT Press, Cambridge), 194 p.

Lieberman, P. (**1967**). *Intonation, Perception and Language* (MIT Press, Cambridge), 240 p.

Liu, F., and Xu, Y. (**2005**). "Parallel encoding of focus and interrogative meaning in Mandarin intonation," Phonetica **62**, 70–87.

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**). "Quantitative analysis of intonation patterns in statements and questions in Cantonese," in *Proceedings of Speech Prosody-2006*, Dresden, Germany, pp. 277–280.

Ma, J. K.-Y., Ciocca, V., and Whitehill, T. (**2006b**). "Effect of intonation on Cantonese lexical tones," J. Acoust. Soc. Am. **120**, 3978–3987.

Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide* (Lawrence Earlbaum Associates, New York), 492 p.

Patel, R. (**2003**). "Acoustic characteristics of the question-statement contrast in severe dysarthria due to cerebral palsy," J. Speech Lang. Hear. Res. **46**, 1401–1415.

Peters, B., and Pfitzinger, H. (**2008**). "Duration and F0 interval of utterance-final intonation contours in the perception of German sentence modality," in *Proceedings of the Interspeech*, Brisbane, Australia, pp. 65–68.

Remijsen, B., and van Heuven, V. J. (**1999**). "Gradient and categorical pitch dimensions in Dutch: Diagnostic test," in *Proceedings of the International Congress of Phonetics Sciences*, San Francisco, CA, pp. 1865–1868.

Schneider, K., and Lintfert, B. (**2003**). "Categorical perception of boundary tones in German," in *Proceedings of the International Congress of Phonetics Sciences*, Barcelona, Spain, pp. 631–634.

Shen, X.-N. S. (**1989**). *The Prosody of Mandarin Chinese* (University of California Press, Berkeley), 95 p.

van Heuven, V. J., and Haan, J. (**2000**). "Phonetic correlates of statement versus question intonation in Dutch," in *Intonation: Analysis, Modelling and Technology*, edited by A. Botinis (Kluwer Academic Publishers, Dordrecht), pp. 119–144.

Yuan, J.-H. (**2004a**). "Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling," Ph.D. thesis, Cornell University, 416 p.

Yuan, J.-H. (**2004b**). "Perception of Mandarin intonation," in *Proceedings of the International Symposium on Chinese Spoken Language Processing-2004*, Hong Kong SAR, China, pp. 45–48.

Yuan, J.-H., and Shih, C.-L. (**2004**). "Confusability of Chinese intonation," in *Proceedings of the Speech Prosody-2004*, Nara, Japan, pp. 131–134.