The Effect of Intonation on Perception of Cantonese Lexical Tones

Joan K-Y Ma, Valter Ciocca & Tara L Whitehill

Division of Speech and Hearing Sciences
University of Hong Kong
joanma@hkusua.hku.hk

Abstract

The aim of this study was to investigate the perception of intonation-induced changes in tone in Cantonese. Previous studies have shown that the F0 level and contour of tones are likely to be modified by different intonations. As perception of tones depends primarily on the fundamental frequency (F0) pattern, it is likely that these intonation-induced changes in F0 will affect listeners’ perception. In the present study, speech materials of two different intonations (question and statement) were used. Results showed that overall tones were accurately perceived, except at the final position of questions, where tones 33, 21, 23 and 22 were confused with tone 25. The perceptual patterns of individual listeners were also analysed, with differences in tone identification strategies identified across listeners.

1. Introduction

Cantonese is a tone language in which a contrast in tone marks a difference in lexical meaning. Cantonese tones are characterized by tone level (high, mid and low) and tone contour (level, rising and falling) [1]. There are six contrastive tones in Cantonese: high-level (55), high-rising (25), mid-level (33), low-falling (21), low-rising (23) and low-level (22) [2]. Cantonese tones are of relative relations and the F0 levels are relative, not absolute.

The use of intonation in a tone language could be potentially confusing as fundamental frequency (F0) serves two similar but separate functions [3]. The variation of F0 at sentential level marks intonation while the variation within a syllable marks tone. Studies in different languages have disagreed in terms of how tones are modified by intonation (e.g. [3][4][5]), but it is generally agreed that tone is likely to be modified by intonation (e.g. [4][7]).

In Cantonese, there have been a few studies on the effect of intonation on lexical production using acoustic analysis. Fok-Chan [1] studied tones produced in neutral, anger, empathetic and interrogatory manners, and found that all the tones produced in an interrogatory manner have a rising F0 contour. Vance [3] examined the effect of word position within statements and found the existence of sentence-final tone lowering in Cantonese. In investigating the effect of questions and statements on tone production, Ma, Ciocca & Whitehill [8] showed that both tone level and tone contour are modified by sentence intonation. They found that questions raise the F0 level of the tone at all positions, when compared to statements. In addition, at the final position of statements, all six tones showed a rising contour regardless of the inherent contour. These studies showed that the F0 patterns of tones are differentially affected intonation when placed at different positions of the sentence.

F0 is the most important perceptual cue in tone perception [1][3]. Fok-Chan [1] suggested that once the structure of the tone is disturbed, perception would be affected. Fok-Chan [1] found that tones produced in interrogatory manners were likely to be confused with tone 25 and tone 23, and explained the perceptual pattern with the acoustic changes observed in tones produced in interrogatory manner. However, in Fok-Chan’s study [1], the stimuli used were produced in isolation form, which might be different from tones produced within sentences, where the context might provide cues to listeners to compensate for the intonation-induced F0 changes in tone. Fok-Chan [1] also mentioned that tones produced in isolation are difficult to perceive accurately. This is because the context in which the tone is produced provides the listeners’ cues for tone normalization, which are reported to be important for tone perception [9][10]. Therefore, stimuli presented in isolation might underestimate listeners’ ability to perceive intonation-induced changes in a natural context. The purpose of the present study was to examine how intonation-induced changes in F0 pattern of tone are perceived when presented within a natural sentence.

2. Method

2.1. Listeners

Ten females served as listeners (aged 18 to 19 years old). They were first-year undergraduates in the Division of Speech and Hearing Sciences, University of Hong Kong. They were considered naïve listeners, as the experiment was carried out within their first two months at university, during which they received limited training in phonetics.

Cantonese was the first language for all the listeners. All listeners passed a hearing screening (≤ 20 dBHL at 250, 500, 1000, 2000 and 4000 Hz). A tone perception screening was carried out with all listeners, consisting of 24 trials (four trials for each tone). Listeners had to correctly perceive at least three out of four trials per tone in order to pass the screening. They achieved an overall average of 99.58% accuracy.

2.2. Speech Materials

Speech materials were collected from two speakers (one male and one female). Two contexts were designed, with the target word in initial position (/X tsi22 hou25 η /35 ‘X is difficult to write’) or final position (/lei55 k 33 tsi22 h i25 X/ ‘This word is X’).
Three sets of target words were embedded in the above contexts. Each set consisted of six words that differed only in tone. The three sets of words were derived from the syllables /sɪ/, /ʃɪ/, /ʃɪl/. All target sentences could be read in two different intonations – question or statement. With two contexts, eighteen target words and two intonation patterns, there were a total of 72 different stimuli from each speaker.

### 2.2. Procedures

The experiment was carried out in a sound-attenuated room (IAC sound-proof booth), with the speech materials presented to the listeners through a Sennheiser HD 545 headset, connected to an Apple Macintosh G4 computer. A HyperCard programme was used to run the experiment. For each trial, six Chinese characters were presented on the screen. The carrier was also presented at the top of the screen.

The stimuli were divided into two blocks, one for the male speaker and one for the female speaker. Within each block, there were a total of 144 trials, as each of the 72 stimuli was repeated once. The order of presentation (male, female) was counterbalanced across the ten listeners.

Before each session, the 18 Chinese characters were read aloud to the listeners by the first author to ensure that they were familiar with all the characters. This was necessary as the same Chinese character may have more than one pronunciation in different contexts in Cantonese. Within each trial, the listener was asked to identify the Chinese character that matched with 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 the “repeat sound” button. Each block took about half an hour to finish, and each session took about one hour.

### 3. Results

#### 3.1. Perceptual Accuracy

Confusion matrices were compiled separately according to position (initial and final) and intonation (question and statement). Group confusion matrices were compiled by summing confusion matrices across the ten listeners, and are displayed in Tables 1 to 4. The number in the cells represents the percentage of responses being realized as that particular tone. For example, tone 25 at the initial position of questions was accurately identified 116 times out of 120 trials, resulting in an identification percentage of 96.67%. The percentages of correct identification were relatively similar in all the contexts, except for the final position of questions, where the percentage of correct identification was low. At the final position of questions, tone 55 was perceived with 100% accuracy in this context. For tones 25 and 33, although the accuracy was lower than in the other contexts, the differences between tones at final position of question and statement.

<table>
<thead>
<tr>
<th>Target</th>
<th>Perceived Tones</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>100</td>
</tr>
<tr>
<td>25</td>
<td>96.67</td>
</tr>
<tr>
<td>33</td>
<td>99.17</td>
</tr>
<tr>
<td>21</td>
<td>2.50</td>
</tr>
<tr>
<td>23</td>
<td>6.67</td>
</tr>
</tbody>
</table>

### Tables 1-4: Confusion matrices for the six tones at different positions (initial and final) and intonations (questions and statements). Target tones are on the horizontal axis, and perceived tones on the horizontal axis. Cell numbers represent the percentage of responses for each target tone.

#### Table 1: Initial position of questions.

<table>
<thead>
<tr>
<th>Target</th>
<th>Perceived Tones</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>100</td>
</tr>
<tr>
<td>25</td>
<td>99.17</td>
</tr>
<tr>
<td>33</td>
<td>95.83</td>
</tr>
<tr>
<td>21</td>
<td>100</td>
</tr>
<tr>
<td>23</td>
<td>0.83</td>
</tr>
<tr>
<td>22</td>
<td>99.17</td>
</tr>
</tbody>
</table>

#### Table 2: Initial position of statements.

<table>
<thead>
<tr>
<th>Target</th>
<th>Perceived Tones</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>100</td>
</tr>
<tr>
<td>25</td>
<td>78.33</td>
</tr>
<tr>
<td>33</td>
<td>30.83</td>
</tr>
<tr>
<td>21</td>
<td>70.83</td>
</tr>
<tr>
<td>23</td>
<td>75.00</td>
</tr>
<tr>
<td>22</td>
<td>60.83</td>
</tr>
</tbody>
</table>

#### Table 3: Final position of questions.

<table>
<thead>
<tr>
<th>Target</th>
<th>Perceived Tones</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>100</td>
</tr>
<tr>
<td>25</td>
<td>97.50</td>
</tr>
<tr>
<td>33</td>
<td>94.17</td>
</tr>
<tr>
<td>21</td>
<td>98.33</td>
</tr>
<tr>
<td>23</td>
<td>2.50</td>
</tr>
<tr>
<td>22</td>
<td>0.83</td>
</tr>
</tbody>
</table>

#### Table 4: Final position of statements.

<table>
<thead>
<tr>
<th>Target</th>
<th>Perceived Tones</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>100</td>
</tr>
<tr>
<td>25</td>
<td>92.92</td>
</tr>
<tr>
<td>33</td>
<td>88.75</td>
</tr>
<tr>
<td>21</td>
<td>81.88</td>
</tr>
<tr>
<td>23</td>
<td>76.25</td>
</tr>
<tr>
<td>22</td>
<td>78.96</td>
</tr>
</tbody>
</table>

#### Table 5: Mean and standard deviation of the percentage accuracy of the six tones

<table>
<thead>
<tr>
<th>Tone</th>
<th>Mean (%) Accuracy</th>
<th>Standard Deviation (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>92.92</td>
<td>9.78</td>
</tr>
<tr>
<td>33</td>
<td>88.75</td>
<td>15.42</td>
</tr>
<tr>
<td>21</td>
<td>81.88</td>
<td>35.15</td>
</tr>
<tr>
<td>23</td>
<td>76.25</td>
<td>43.62</td>
</tr>
<tr>
<td>22</td>
<td>78.96</td>
<td>37.20</td>
</tr>
</tbody>
</table>

#### 3.2. Intonation and Position

The percentages of correct identification were relatively similar in all the contexts, except for the final position of questions, where the percentage of correct identification was low. At the final position of questions, tone 55 was perceived with 100% accuracy in this context. For tones 25 and 33, although the accuracy was lower than in the other contexts, the differences between tones at final position of question and statement.
other contexts were not significant (p > 0.05 for all), except for the difference between the final position of questions and the initial position of statements for tone 25, where the difference was significant (T = 0, p < 0.005). For the other three tones (tone 21, tone 23 and tone 22), tones at the final position were the least accurately perceived and were significantly different from the other three contexts (T = 0, p < 0.005 for all).

3.3. Error Patterns

Error patterns in identification were also analysed. Each response was classified as either correct or incorrect according to the target tone. Incorrect responses were classified into either (1) tone level error: the stimulus was identified as a tone of the same contour, but an incorrect level of tone was perceived (for example, tone 55 → tone 22), or; (2) tone contour error: the stimulus was perceived as a tone of a different contour (for example, tone 55 → tone 25). The error patterns for each tone were summarized in terms of types of errors and are displayed in Table 6. The numbers in the cells represent the percentage of a given error type. For example, out of the 34 identification errors for tone 25, 9 tone level errors involving tones from the final position of questions were observed. This gives a percentage of 26.47% of errors (9 out of the 34 errors) in this category.

As described in the previous section, stimuli presented at the final position of questions were least accurately perceived. Most of the errors in this condition involved misperceiving the target tone as tone 25. This accounted for 75.93% of the identification errors for tone 33, 97.70% of the identification errors for tone 21, 78.95% of the identification errors for tone 22 and 82.18% of the errors. There was a small difference was significant (T = 0, p < 0.05). For the other three tones (tone 21, tone 23 and tone 22), tones at the final position were the least accurately perceived and were significantly different from the other three contexts (T = 0, p < 0.005 for all).

3.4. Identification Pattern for Individual Listeners

Confusion matrices of individual listeners were examined. The percentages of correct identification of each listener at the final position of questions were summarized in Table 7. Listeners could be classified into four groups according to perceptual patterns:

- **Group A.** Listeners in this group were able to identify tones 55, 33, 21, and 22 at the final position of questions correctly, while their perception of tones 25 and 23 were lowered. Two listeners were this category (L1 and L2).
- **Group B.** Listeners in this group were able to identify tones 55 and 25 correctly, but most of the responses for tones 33, 21, 23, and 22 were tone 25. The identification pattern of four listeners fitted this description (L3, L4, L5, and L6).
- **Group C.** In addition to being able to perceive tones 55 and 25 correctly (as for group B), this group of listeners was also able to perceive tones 23, 25, and 22 correctly. Two listeners were classified into Group C (L7 and L8).
- **Group D.** As in Group C, tones 55, 25, and 33 were perceived with a high level of accuracy. Although demonstrating relatively low accuracy, this group of listeners was also able to perceive tones 21, 23, and 22 accurately in some trials. This group comprised two listeners (L9 and L10).

Table 7: Identification pattern for individual listeners. Numbers in the cells represent the accuracy for that particular tone by the listener on the left.

![Table 7](image)

4. Discussion

4.1. Perceptual Accuracy and Error Patterns

Except at the final position of questions, the six tones showed identification accuracy above 90% accuracy in the different contexts. Tones at the final position of questions had low percentages of correct identification. This is presumably related to the intonation-induced changes in F0 pattern at the final position of questions.

At both positions (initial and final) of statements and the initial position of questions, the error patterns observed were consistent with those reported in Fok-Chan [1]. Confusion between tones 33 and 22 was noted. Fok-Chan [1], who reported a similar finding, suggested that this confusion is
because both tone 33 and tone 22 are of slightly falling contour, with difference only in F0 level. Tone 55 was confused with neither tone 33 nor tone 22, even though tone 55 was also of a slightly falling contour, owing to the distinctiveness in F0 level. Tone 23 was also misperceived as tone 22 occasionally. Fok-Chan [1] argued that this is because of the gentleness of the slope for tone 23; listeners would have to perceive an actual increase in F0 contour before deciding it is a rising tone. A different finding was noted for tone 25 at the final position of questions, where some trials were perceived as tone 22. When the acoustic pattern was explored in detail, tone 25 in this context began at an F0 level lower than that of tone 22. It is postulated that some listeners might focus on tone level, while the changes in tone contour were compensated for by the F0 changes expected in this intonation pattern (questions).

At the final position of questions, perceptual errors were noted especially for tones 33, 21, 23 and 22. Ma et al. [8] showed that all six tones in this context have rising contours, regardless of the inherent F0 contour. Tones 25, 21, 23 and 22 were found to be very close in both tone level and tone contour, while the level of tone 55 and 33 were distinctive from the rest. Fok-Chan [1] claimed that once the structure of the tone is disturbed, the identity of the tone will be lost and perception will be difficult. This accounted for the relatively well-preserved identification accuracy for tones 55, 25 and 33 in this context, but not for tones 21, 23 and 22. Although the contours for tones 55 and 33 clearly differentiate from canonical F0 patterns, listeners were able to use the distinction in tone levels for these two tones as perceptual cues. For tone 25, the F0 pattern at the final position of questions coincided with the rising contour of its original contour; therefore, the perception in this context was not particularly affected. For tones 33, 21, 23 and 22, most of errors at the final position of questions involved misperceiving the target tone as tone 25. With the similarity in F0 pattern, tones 21, 23 and 22 were confused with tone 25. Although tone 33 in this context had a tone level slightly above that of tone 25, the rising contour of tone 33 in this context confused the listeners, as tone 25 is the only rising tone with this F0 level in Cantonese. On the other hand, a small portion of tone 25 targets was perceived as tone 23 in this context. This is probably related the closeness of tone level between tones 25 and 23 at the final position.

4.2. Differences between Listeners

Gandour [11], who investigated the perceptual dimensions of Cantonese tones, suggested that Cantonese tones were perceived by features rather than by tone unit as a whole, as the six tones showed differences in confusion patterns. The results of his study suggested that the direction of F0 movement was more influential for listener perception than tone level. Although the results from the present study showed more tone level errors than tone contour errors, except at the final position of questions, the differences in identification pattern among listeners at the final position of questions suggested that tone contour was not necessarily the most salient cue in tone perception.

For the listeners in Group A, tone level was more influential than contour in perception. Although all six tones at the final position of questions showed a rising F0 contour, the two rising tones (25 and 23) were the least accurately perceived among the six tones; tones 55, 33, 21 and 22 were perceived more accurately. As listeners’ perception was not affected by the intonation-induced changes in F0 contour, this suggests that the initial frequency of the tone and the tone levels were more salient perceptual features for this group of listeners. The context in which the tone was produced in might provide listeners cues to compensate for the intonation-induced F0 changes in perception.

By contrast, listeners in Group B placed more emphasis on tone contour than level. Most of the trials for tones 33, 21, 23 and 22 were perceived as tone 25. This showed that the rising contour have more saliency for this group of listeners, rather than the contrast in tone level.

For listeners in Group C and D, both tone level and tone contour impacted on perception. In addition to tones 55 and 25, tone 33 was perceived with 100% accuracy in these two groups. Although the contour of tone 33 was modified to a rising contour by intonation, the contrast in tone level was a significant cue for the listeners to distinguish tone 33 from the rest. However, these listeners showed poor ability in distinguishing the overlapping contours of tones 21, 23 and 22 from tones 25 by the level cues (as for Group A). Tone contour appeared to be a more salient cue for listeners in this group, resulting in confusion of tones 21, 23 and 22 as tone 25. Listeners in Group D showed a slightly better ability in identifying tones 21, 23 and 22 than listeners in Group C, with less than 50% identification accuracy of these three tones at the final position of questions.

5. Conclusions

This study showed that intonation-induced F0 changes listener’s perception of lexical tones. Tones at the final position of questions were less accurately perceived than tones in other contexts. Errors were especially noted for tones 33, 21, 23 and 22, where confusion with tone 25 was common. The results also showed that different features (tone level and tone contour) have different influences on tone perception. Variations in the importance of these cues were noted across listeners.

6. References