

Temporal Distribution of Tonal Information in Continuous Cantonese Speech

Ying Wai WONG

Department of Linguistics and Modern Languages
Chinese University of Hong Kong
ywwong@cuhk.edu.hk

Abstract

According to previous studies on contextual tonal variations, as a result of interaction between tone-bearing syllables, carryover effect is much greater than anticipatory effect. Consequently, for a given tone in continuous speech, the most consistent portion of F_0 (fundamental frequency) contour resides in the latter part of the host syllable. Perception tests on Cantonese level tones were conducted in this study to investigate whether such asymmetry in terms of F_0 contour consistency is exploited by our tone perception system. Given conflicting tonal information within the target syllable, our tests give evidence that the latter portion is more prominent in determining tonal identities. This finding coincides harmoniously with several related results like syllable structure typology and contour tone restrictions, and more importantly, leads us to deeper understanding of temporal organization of human speech sounds. Topics relevant to the current study such as role of intensity profile, etc. and further extensions are also presented.

1. Introduction

As background for the current tone perception study, previous related results, both from production and perception perspectives, are briefed in Sections 1.1 and 1.2. Following that, Section 1.3 integrates those previously obtained findings to establish the current study.

1.1. Tone production

Just like most tone languages, produced in isolation, the six Cantonese lexical tones can be identified quite robustly from their distinct F_0 contours. The tone identification task, however, becomes more perplexing when the target syllable appears in continuous speech, where its F_0 contour is heavily influenced by surrounding tonal environment, as reported for a variety of languages like Cantonese [1, 11], Mandarin [17], Thai [5] and Vietnamese [7]. Cantonese low-falling tone (T_4), as an example, rarely reaches the floor of ones' F_0 range, as it usually does while in isolation. Majority of the studies agree on the observation that carryover and anticipatory contextual effects differ in terms of their nature and magnitude. Specifically, carryover effect is assimilatory while anticipatory effect is dissimilatory, with the former effect being greater in magnitude. Consequently, for a given tone, initial portion of the corresponding F_0 contour exhibits greater variation than that of the latter portion.

1.2. Tone perception

A number of tone perception studies [2-4, 9, 11, 14] have been carried out previously, focusing on the intrinsic cues of

Cantonese tones. Often, continua along acoustic dimensions like *level* and *slope* of F_0 contours of the target syllable were created for tone identification and discrimination tasks to evaluate their significance in affecting tone perception.

Similar to tone production, continuous speech introduces further complexities into tone perception mechanism. Referring to some such studies [8, 10, 12, 13, 15, 16], context was reported to be heavily relied on in tone perception, parallel to vowel perception. Usually, the context provides a reference of F_0 range for normalizing F_0 values from variations due to intra- and inter-speaker factors [15]. Moreover, there are circumstances, like the one in Wong & Diehl's [15] study, in which the context can even *override* intrinsic acoustic cues of the target syllable in tone identification.

1.3. The current study

Given the directionally asymmetric degree of contextual effects mentioned in Section 1.1, in continuous speech, the latter portion of F_0 contours for each syllable are more consistent as a result. From economy point of view, less memory is required for storing templates of the latter portion of F_0 contour, compared to that of the initial portion which shows more variations. Does our tone perception system take advantage of this asymmetry to rely primarily on the latter portion? A series of experiments were conducted to verify such tone perception phenomena in continuous speech, taking into consideration both *intrinsic* (i.e. the target syllable) and *extrinsic* (i.e. the context) cues, as pointed out in Section 1.2.

There were two perception tests carried out in this study, a pre-test validation task (Experiment 1) and the main test (Experiment 2), as presented below, with details like stimulus preparation, manipulation and test procedure.

2. Experiment 1

To elicit subjects' performance in perceiving natural utterances, we find a validation task necessary to confirm whether our test sentences, after F_0 manipulation, still offer a sense of natural utterances.

2.1. Stimulus recording

To facilitate our study of temporal distribution of tonal information, we choose the syllable [ji], represented in *Jyutping* as *ji*, which is voiced throughout its whole duration. In current study, we focus only on the three Cantonese level tones (high level T_1 , mid-level T_3 and mid-low level T_6), and those tones are associated with the target syllable *ji* which is embedded in a carrier sentence *ngo5 ji4 gaal duk6 __ zib6 bei2 nei5 teng1* (Now, I read to you the character __). To make this test sentence sound more naturally, it is designed such that the target syllable is surrounded by voiceless

portions respectively due to preceding voiceless velar stop [k] from *duk6* and following voiceless affricate [ts] from *zi6*. This minimizes the audible F_0 discontinuities at both boundaries of the target syllable resulting from F_0 manipulation.

30 test sentences, comprising the target syllable specified as the three level tones (T_1 , T_3 , T_6) embedded in the test sentence mentioned above, each with 10 repetitions, were prepared. The recordings were obtained from a male native Cantonese speaker (the author) in a randomized order. Recording session was conducted in a quiet room, with a Sony ECM-MS957 electret condenser microphone, at 22kHz sampling rate. After pitch analysis by PRAAT, F_0 contours for the target syllable for the three level tones, together with their corresponding neighboring syllables, are plotted in Fig. 1a. The F_0 traces of syllables differ from their canonical forms considerably: Assimilatory carryover and dissimilatory anticipatory effects as previously reported (e.g. [17]) can be observed at onset and offset of voiced portions of *zi6* and *duk6* respectively. F_0 perturbation due to voiceless consonants [18] are present at both ends of the target syllable. Inspecting their F_0 levels, T_6 (mean: 110 Hz) locates closer to T_3 (mean: 120 Hz) than T_1 (mean: 137 Hz) does, matching their tone letter labels quite well (T_1 : 55, T_3 : 33, T_6 : 22).

Then, one utterance for each tone, each from a pool of 10 repetitions, was selected based on *least-square error* criterion with reference to the mean curve shown in Fig. 1a. The three selected test sentences are graphed in Fig. 1b. Comparing them with the ones in Fig. 1a, the curves generally show similar trends in terms of their F_0 movement, though with trajectories in Fig. 1a appearing smoother due to averaging.

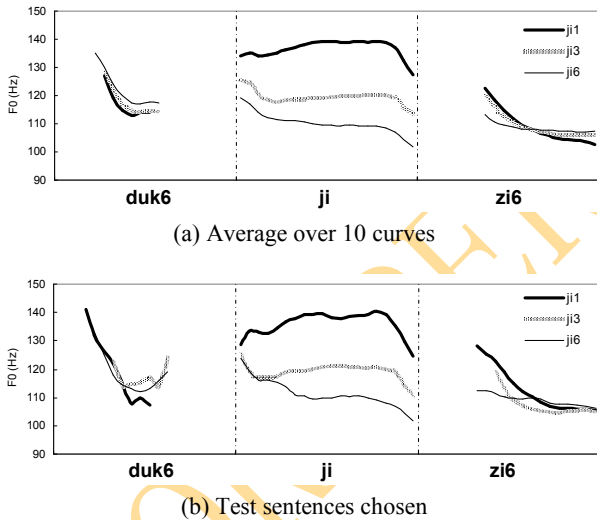


Figure 1: F_0 traces for the target syllable “ji”, together with two neighboring syllables “duk6” and “zi6”.

2.2. Stimulus preparation

Observed from Fig. 1b, the target syllable *ji*, when specified with different tones, is associated with different context in terms of F_0 trajectories. To counter-balance such context effect, all the three test sentences were used in our test.

We imposed the F_0 contours of the target syllable obtained in Fig. 1b on each of the three test sentences to minimize unnaturalness due to discontinuities in formant movement and intensity profile, especially for Experiment 2.

Summing up, we have 9 combinations (3 carrier sentences \times 3 F_0 contours) up to now. Multiplied by 5 repetitions, a total of 45 tokens, grouped into 5 blocks, were presented to each subject for a tone identification task.

2.3. Procedure

11 subjects (8M3F), all native Cantonese speakers, participated in the test. Tests were carried out in a quiet room, with speaker volume tuned at a comfortable level. The subjects were requested to identify the tones mentioned in given test utterances by marking the corresponding Chinese characters on a given answer sheet. For each utterance, three characters were given as choices, which were the syllables *ji* associated with the three level tones (T_1 , T_3 and T_6). The stimuli, arranged in a randomized order, were presented with intervening silence intervals of 4s (i.e. ISI = 4s), and subjects had to make their responses within that period.

2.4. Results

Fig. 2 shows the tone identification results from Experiment 1. In the figure, a column specified with C_i and S_j represent responses (shaded differently according to the answered tones) resulting from the carrier sentence from originally tone-*i* recording, but with F_0 contour of the target syllable *ji* replaced by a T_j contour. Under this naming scheme, columns C1-S1, C2-S2, C3-S3 are untouched utterances, and they all yield 100% identification of their corresponding tones. For other columns, responses agree mostly with the F_0 contour of the target syllable, regardless of the carrier sentence used. Comparing those percentages, C1-S6 and C6-S1 are the lowest (both being 93%) among the nine. Incidentally, among the three level tones, T_1 and T_6 are farthest from one another acoustically, pointing to influence possibly due to the context (i.e. neighboring syllables of the target *ji*).

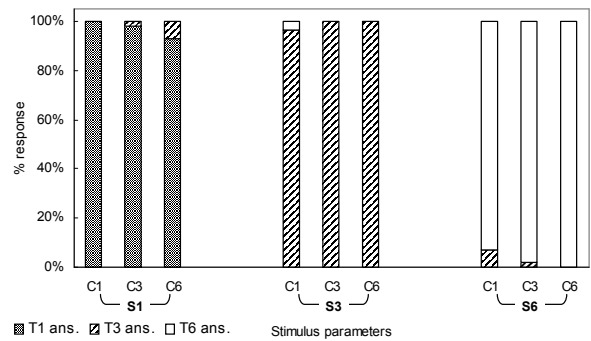


Figure 2: Identification results in Experiment 1, where responses mostly correspond to F_0 contour of the target syllable.

The aforementioned observations are confirmed by a two-factor (*carrier sentence*, F_0 contour) repeated measures ANOVA conducted, which indicates a main effect only for F_0 contour, $F(2,20) = 2375.127$, $p = 0.000$. Conversely, neither *carrier sentence* ($F(2,20) = 1.151$, $p = 0.336$) nor interaction between these two factors ($F(4,40) = 1.885$, $p = 0.132$) show significant results.

To sum up, though minor context effect does exist, Experiment 1 confirms the dominant role of tonal contours of the target syllable in tone identification, enabling us to

investigate the relative importance of different portions of the *target syllable* itself effectively.

3. Experiment 2

Results from Experiment 1 suggest that with minor context effect, F_0 imposition on the target syllable in carrier sentences successfully led subjects to identify the tone associated with the F_0 contour. We proceeded to determine the most prominent portion, in our case, which of the two temporally divided halves, of the target syllable in tone perception.

3.1. Stimulus preparation

We divide temporally the target syllable into two equal portions. F_0 contours of conflicting tonal identities were then imposed on the two parts, producing 6 combinations (T_1 - T_3 , T_1 - T_6 , T_3 - T_1 , T_3 - T_6 , T_6 - T_1 , and T_6 - T_3). Fig. 3 illustrates 3 such F_0 traces in a T_6 carrier sentence. Jumps in F_0 can be observed at the boundary between the halves, however, these were not noticed by majority of the subjects for this test.

Summing up, 18 combinations (3 carrier sentences x 6 F_0 contour combinations) were obtained. With 5 repetitions for each, a product of 90 tokens were prepared for each subject.

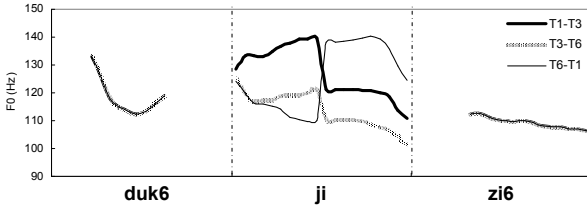


Figure 3: Examples of F_0 contours after manipulations for Experiment 2.

3.2. Procedure

The same 11 subjects proceeded to Experiment 2. Test procedure was basically the same except that this time, 10 blocks, each consisting of 9 stimuli, were presented.

3.3. Results

Inheriting the assumption from Experiment 1 that context effect only exists minimally, tone identification results for utterances from different carrier sentences are pooled together, as shown in Fig. 4. To facilitate discussions, hereafter, *first tonal target* (FTT) and *second tonal target* (STT) refer to tones specified by F_0 contours of the 1st and 2nd halves of the target syllable, such that, for instance, the T_1 - T_3 trace in Fig. 3 is said to have an FTT of T_1 and an STT of T_3 . In Fig. 4, responses are grouped into 6 columns according to the tonal targets (F for FTT, S for STT) of the stimulus sentence. For this experiment, there was no stimulus with two tonal targets being the same (e.g. F1-S1) as only the comparative prominence of the two portions is our focus.

Generally speaking, identification responses follow quite neatly the *second tonal target* in all the six cases, with percentages ranging from 88% (for columns F3-S1 and F6-S1) to 97% (for columns F1-S3 and F1-S6), supporting our conjecture that tone perception primarily relies on the *latter portion* of the host syllable.

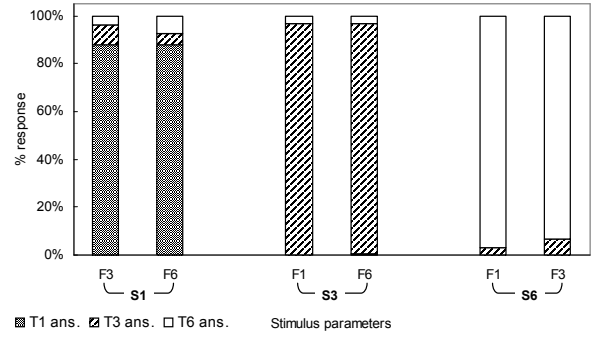


Figure 4: Identification results in Experiment 2, where responses are observed to be primarily determined by F_0 contours of the second half of the target syllable.

Regarding the responses deviated from the majority (e.g. T_3 responses in F1-S6), we can observe the patterns: (1) only T_3 responses for Fx-S6 sentences; (2) only T_6 responses for Fx-S3 sentences (with one exceptional instance at F6-S3). T_1 responses are not preferred possibly due to its comparatively long acoustic distance from the other two level tones; (3) both T_3 and T_6 responses for Fx-S1 sentences. Specifically, a larger proportion of responses for F3-S1 sentences go to T_3 , while a larger proportion of responses for F6-S1 sentences go to T_6 . These results are interpreted as evidence that the initial portion of the F_0 contours in a syllable also plays a role in tone perception.

To statistically evaluate the effect of different factors, repeated measures ANOVA tests are carried out. As our design does not include stimuli with the two tonal targets specified as the same tone (i.e. F1-S1, F3-S3 and F6-S6), a *full factorial* of the factors (*carrier sentence, first tonal target, second tonal target*) is not available for a three-factor ANOVA test. Instead, we break down the problem into 6 two-factor repeated measures ANOVA tests shown in Table 1. Due to space limitation, we only present p values here. Judging from the table, there is main effect only for the *second tonal target* factor, statistically supporting our observations from visual inspection.

Table 1: p values from two-factor repeated measures ANOVA test results: Asterisks (*) represent items reaching the significance level $p < 0.05$.

test #	factors	stimuli included	factor1	factor2	interaction
1	carrier, FTT	F3-S1, F6-S1	0.113	0.508	0.225
2	carrier, FTT	F1-S3, F6-S3	0.892	1.000	0.371
3	carrier, FTT	F1-S6, F3-S6	0.831	0.209	0.547
4	carrier, STT	F1-S3, F1-S6	0.913	0.000*	0.183
5	carrier, STT	F3-S1, F3-S6	0.106	0.000*	0.272
6	carrier, STT	F6-S1, F6-S3	0.153	0.000*	0.080

To summarize, through conflicting tonal information associated with two halves of a single syllable, Experiment 2 points to the dominance of the F_0 contour of the *latter portion* of a syllable in determining its tonal identity.

4. Discussions

4.1. Speech production phenomena

We compare our findings against various phenomena to gain insights into temporal organization of human speech sounds: First, CV is the most common structure among a variety of syllable types. Among voiceless consonants, stops leave a portion of silence due to vocal tract closure, and aperiodic air turbulence portions associated with fricatives and affricates cannot bear F_0 information. These all contribute to lower the likeliness of the initial portion of syllables as the major identity to bear tonal information.

Second, from previous studies on contextual tonal variations [1, 5, 17], contour tones are observed to exhibit their supposedly most characterizing rising and falling F_0 trajectories mainly in the 2nd half. This suggests that tonal information for contour tones are packed in the latter portion of syllables.

Third, a hierarchy of *tone-bearing ability* was proposed in a typological study of 105 languages on contour tone restrictions [6], which states the relative tolerance of contour tones on different syllable types. Observed from the hierarchy, ending segments (e.g. sonorant / obstruent / null consonant coda) of a syllable play a prominent role in determining tone-bearing ability. In particular, only a small proportion of languages surveyed allow contour tones on syllables closed with obstruents. This restriction turns out to be explainable if our observation of perceptual bias is correct: Phonetically, the ending obstruents represent a region where F_0 cues are not assessable, or only minimally retained in case of voiced ones. This is equivalent to loss of F_0 information at the ending portions of the host syllables [18], which is *perceptually salient* according to our results. Thus, this significant reduction of available F_0 information can be compensated by having less rich tonal information encoded. Absence of contour tones is a good strategy. Cantonese demonstrates this restriction well, where only 3 level tones (T_1 , T_3 and T_6) can be associated with syllables closed with voiceless stops.

4.2. Intensity profile

Inspecting the intensity profile, all of the three test syllables (T_1 , T_3 and T_6) have offset portions with relatively higher intensity than their onset portions. As the intensity is directly related to the level of sound pressure perturbation stimulating our auditory system, this appears to be a possible explanation to the perceptual bias on the latter portion at a first glance. However, according to an undergoing test in which the intensity level is equalized over the whole duration of syllable, the bias is still quite significant, though the quantitative difference, if any, is yet to be measured.

5. Conclusions and future work

In terms of temporal location within a syllable, our study unveils that the *latter portion* of a syllable is perceptually more salient in identifying tones in continuous speech. Contextual tonal cues, instantiated as F_0 contours on surrounding syllables, contribute as well in hinting tonal identities of the target syllable. From the perception perspective, our findings fit well with commonly known speech production phenomena like contour tone restrictions, syllable structure typology, and etc.

As intensity profile of the target syllable used in this study directly correlates with the relative prominence observed, extended studies are needed to evaluate the effect of such factor. Furthermore, to lower the tone perception task difficulty, we used only 3 level tones as choices, which limit the extent to which our conclusions can be generalized. Separate studies are necessary to investigate the situation of the remaining two rising and one falling tones in Cantonese.

6. Acknowledgement

The author wishes to thank Thomas Lee, William Wang, Eric Zee, Yi Xu and members of Language Engineering Laboratory and Language Acquisition Laboratory for helpful comments on earlier versions of this paper.

7. References

- [1] Chang, C. Y., 2003. *Intonation in Cantonese*. LINCOM Studies in Asian Linguistics. Vol. 49, Munich: Lincom Europa.
- [2] Fok, C. Y.-Y., 1974. *A perceptual study of tones in Cantonese*. Center of Asian Studies. University of Hong Kong.
- [3] Francis, A. L.; Ciocca, V. C.; Ng, B. K. C., 2003. On the (non)categorical perception of lexical tones. *Perception and Psychophysics* 65(6), 1029-1044.
- [4] Gandour, J. T., 1983. Tone perception in Far Eastern languages. *Journal of Phonetics* 11, 149-175.
- [5] Gandour, J. T.; Potisuk, S.; Dechongkit, S., 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22, 477-492.
- [6] Gordon, M., 2001. A typology of contour tone restrictions. *Studies in Language* 25, 405-444.
- [7] Han, M. S.; Kim, K. O., 1974. Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics* 2, 223-232.
- [8] Leather, J., 1983. Speaker normalization in perception of lexical tone. *Journal of Phonetics* 11, 373-382.
- [9] Li, P. Y., 2004. Perceptual analysis of the six contrastive tones in Cantonese. In *TAL-2004*, 119-122.
- [10] Lin, T.; Wang, W. S.-Y., 1985. Shengdiao ganzhi wenti [Tone perception]. *Zhongguo Yuyan Xuebao* 2, 59-69.
- [11] Liu, J., 2001. Tonal behavior in some tone languages. *Ph.D. Dissertation*. City University of Hong Kong.
- [12] Moore, C. B.; Jongman, A., 1997. Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America* 102, 1864-1877.
- [13] Vance, T. J., 1976. An experimental investigation of tone and intonation in Cantonese. *Phonetica* 33, 368-392.
- [14] Vance, T. J., 1977. Tonal distinctions in Cantonese. *Phonetica* 34, 93-107.
- [15] Wong, P. C. M.; Diehl, R. L., 2003. Perceptual normalization of inter- and intra-talker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research* 46, 413-421.
- [16] Xu, Y., 1994. Production and perception of coarticulated tones. *Journal of the Acoustical Society of America* 95, 2240-2253.
- [17] Xu, Y., 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 61-83.
- [18] Xu, Y.; Wallace, A., 2004. Multiple effects of consonant manner of articulation and intonation type on F_0 in English. *Journal of the Acoustical Society of America*, 115, Pt. 2, 2397.