Journal of Chinese Linguistics Vol. 34, No. 1, 134-154, January 2006

# TEMPORAL AND TONAL ASPECTS OF CHINESE SYLLABLES: A CORPUS-BASED COMPARATIVE STUDY OF MANDARIN AND CANTONESE[1]

**Gang Peng**

*The Chinese University of Hong Kong*

## ABSTRACT

Previous studies on temporal and tonal aspects of languages are usually based on limited data from a small number of subjects. It is difficult to know whether these findings can really represent the general temporal and tonal aspects of continuous speech, or just the speech of the specific subjects involved. Because of this difficulty it may not be appropriate to directly apply these findings to current speech technologies.

In this study, large vocabulary continuous speech databases for Mandarin and Cantonese, recorded from large populations of subjects, are used to investigate the temporal and tonal aspects of syllables in continuous speech. Our findings include the following. No obvious temporal compensation effect has been found between the syllable initial (null initial being the sole exception) and the syllable final; Cantonese syllables exhibit less variation than Mandarin, achieving better isochrony.

Then, following the example of the vowel balloons pioneered by Peterson and Barney, the tones of Mandarin and Cantonese with two parameters: $F_0$ height and $F_0$ slope, have been analyzed. Some linguistic hypotheses for tone development are advanced in the discussion of the tone balloons.

## 1. INTRODUCTION

In the phonetic literature, there are two related concepts which refer to the temporal aspects of speech. One is the concept of isochrony, which posits that there is a syllable-like unit which maintains a rather constant time in continuous speech. The other is the concept of temporal compensation, which posits that as the number of elements in the syllable increases, some or all of the elements shorten in order to maintain an approximately constant duration for the syllable.

It was noticed by Lehiste (1972) that the duration of a base word may be considerably reduced, if a derivational suffix is added to it. For instance, when the base words *stead*, *skid* and *skit* are compared with the derived words *steady*, *skiddy* and *skitty*, the duration of the base part of the derived word is considerably shortened. Therefore, even with the addition of a fairly long suffix –*y*, the overall duration of the derived words was not much longer than that of the base words. Several studies examined the temporal compensation within the syllable. Mixdorff et al. (2002) showed that nasals are about 35 ms longer on average after a short vowel than after a long vowel in some word pairs in standard Thai. Zee (1995) found the duration of nasal part of syllable /an/ is much longer than that in syllable /aan/ from two Cantonese speakers. In this study we are concerned with the same temporal questions in speech. However, in contrast with the previous report on laboratory speech, our methods are computational, based on extensive corpora recorded from large numbers of speakers.

Along with the temporal aspects of speech, tone is an essential component of Chinese, in which voice pitch is used to build words much as consonants and vowels do. For instance, in Mandarin, the syllable /ma/, when pronounced with a high level pitch pattern, means “mother”; when pronounced with a rising pattern, the meaning is “hemp”; when pronounced with a falling-rising pattern, the meaning is “horse”; when pronounced with a falling pattern, the meaning is “to scold”. For a more detailed discussion of Mandarin tones, see Wang (1973).

In this paper, the temporal structure of Chinese syllables will be investigated according to different types of initials, finals and tones. Then, the tones will be analyzed and examined following the example of the vowel balloons pioneered by Peterson and Barney (1952).

### 1.1. Organization of the paper

This paper consists of four major sections. In this section, we will further discuss the following topics: syllable structure of Chinese syllables, data collection and some formulae for duration statistics. The next section describes the individual effects of initials, finals and tones, on syllable duration. Then, both Mandarin and Cantonese tones will be examined in section 3. Finally, some general discussion and conclusions will be made in section 4.

<table>
<tr><td colspan="4">Tone</td></tr>
<tr><td rowspan="2">[Initial]</td><td colspan="3">Final</td></tr>
<tr><td>[Medial]</td><td>Nucleus</td><td>[Ending]</td></tr>
</table>

Figure 1. Syllable structure of Chinese dialects. Elements enclosed in [ ] are optional (Wang, 1973).

### 1.2. Syllable structure

Traditionally, as depicted in Figure 1, the Chinese syllable is divided into three components: a tone, an optional initial consonant, and a final which consists of an optional medial, a nucleus and an optional ending. The stop ending is not used in Mandarin, while the retroflex does not occur in Cantonese. Furthermore, in Mandarin, there are three widely used medials, /i/, /u/ and /y/. On the other hand, in Cantonese, there is only one medial, /u/, and this medial can only occur after the velar initials. In order to avoid this kind of skewed distribution, most phonologists view this medial, /u/, as a part of the labialized velar initials. So phonologically, we can say there is no medial, but phonetically, there is still one medial in Cantonese.

Table 1 summarizes the types of tones, initials and finals for Mandarin and Cantonese. Checked syllables end exclusively in the consonants /-p, -t, -k/; while other syllables are unchecked. Meanwhile, the tones on checked syllables are checked tones; the tones on unchecked syllables are unchecked tones. The initials are divided into seven groups according to the manner of articulation, which will be shown later. The nasal finals end in nasal endings; the rhotacized final is /er/ only; the checked finals end in stop endings; all other finals are plain finals. Complete lists of both Mandarin and Cantonese tones, initials and finals can be found in Peng and Wang (2004).

| | Mandarin | Cantonese |
|---|---|---|
| Tones | 4 tones plus neutral tone | 6 unchecked tones, and 3 checked tones |
| Initials | 7 types of initial consonants | 7 types of initial consonants |
| Finals | 3 types of finals: plain, nasal, and rhotacized | 3 types of finals: plain, nasal, and checked |

Table 1. Summary of syllable components of Mandarin and Cantonese.

### 1.3. Data collection

In our study we focus on two dialects of Chinese, Mandarin and Cantonese. In contrast with previous phonetic studies based on manual measurements by the phonetician, we use computer programs [HTK] (Young et al., 2002) for segmentation of continuous speech. The Mandarin database (Peng et al., 2004) consists of 510,791 syllable tokens (39,519 utterances from 38 female and 38 male speakers), while the Cantonese database [CUSENT] (Lee et al., 2002) consists of 215,604 syllable tokens (20,378 utterances from 34 female and 34 male speakers). In general we find good consistency between the computer segmentation and that of the phonetician. However, it is difficult for HTK programs to accurately locate the onset of the leading syllable and the offset of the final syllable of a phrase or an utterance, which segment speech signals only based on the statistics of the target units with specific contexts. So the leading and final syllables are deleted, decreasing the total number of tokens to 425,674 and 164,191 for Mandarin and Cantonese respectively. And one more reason for removing the final-position syllables is that we would like to exclude the final-position lengthening effect at the current stage.

Due to the extremely skewed distribution of the neutral tone in Mandarin, the tokens with neutral tone are deleted except calculating the duration statistics according to different tone types. (There are 392 tone-ignored syllables, i.e., base syllables, among the above 425,674 tokens. But only 16 base syllables are associated with the neutral tone.) Then the number of Mandarin tokens is further reduced from 425,674 to 400,090. Moreover, the tone 3 always changes to tone 2 when followed by another tone 3 due to the tone-sandhi rule (Wang and Li, 1967), so the tokens from tone 3 are deleted when followed by another tone 3. So the tokens are reduced from 425,674 to 416,341 for duration statistics according to tone identities. We repeat that 400,090 tokens are used for tone irrelevant duration statistics (we need not consider tone-sandhi effect, but we have to delete tokens with the neutral tone), while 416,341 tokens are used for tone relevant duration statistics (we have to consider tone-sandhi effect, but we should include tokens with the neutral tone.).

1.4. Some formulae for duration statistics

In the formula

$$d_{ij} = \frac{1}{K}\sum_{k=1}^{K} d_{ijk} , \qquad (1)$$

$K$ is the number of tokens for the $j^{th}$ syllable, and $d_{ijk}$ is the duration of the $k^{th}$ token of the $j^{th}$ syllable for the $i^{th}$ speaker. The Formula (1) is used to avoid the bias in occurring frequency for target syllables. For tone irrelevant statistics, the tonal differences will be ignored. Accordingly, the syllable above will be referred as to base syllable. So $d_{ij}$ is the average duration of the $j^{th}$ syllable for the $i^{th}$ speaker.

In the formula

$$d_i^c = \frac{1}{J}\sum_{j=1}^{J} d_{ij}^c , \qquad (2)$$

$J$ is the number of syllables in a certain cohort $c$, which could be a collection of syllables with the same types of initials, finals or tones, for the $i^{th}$ speaker; $d^c{}_{ij}$ is the average syllable duration calculated by Formula 1, whose carrying syllables belong to the above certain cohort $c$; $d^c{}_i$ is the average syllable duration within the certain cohort for the $i^{th}$ speaker.

In the formula

$$d^{c} = \frac{1}{I}\sum_{i=1}^{I} d_i^{c} , \qquad (3)$$

$I$ is the number of speakers, $d^c$ is the average duration of a certain cohort over all speakers.

## 2. TEMPORAL ASPECTS OF CHINESE SYLLABLES

Early research examining syllable duration utilized highly controlled experimental paradigms in order to deduce the effects of specific factors. Numerous studies concluded that the vowel duration is predicted by numerous factors, including vowel intrinsic duration, post-vocalic consonant voicing, position in word, lexical stress, and phrase final position (House, 1961; House and Fairbanks, 1953; Lehiste, 1972; Lehiste and Wang, 1976; Lieberman, 1960; Peterson and Lehiste, 1960). As for tone languages, such as Mandarin and Hakha Lai[2], the initials, finals, and tones have been demonstrated to play an important role in determining syllable duration (Feng, 1985; Maddieson, 2004).

### 2.1. Temporal variation according to types of initials

| Abbreviation | Manner of articulation | Mandarin | Cantonese |
|---|---|---|---|
| 1. Su | Unaspirate stops | b, d, g | b, d, g, gw |
| 2. Sa | Aspirate stops | p, t, k | p, t, k, kw |
| 3. Au | Unaspirate affricates | z, zh, j | z |
| 4. Aa | Aspirated affricates | c, ch, q | c |
| 5. F | Fricatives | s, sh, x, f, h | s, f, h |
| 6. N | Voiced initials | m, n, r, l | m, n, ng, l |
| 7. Z | Zero (null or glides) | ϕ[3], y, w | ϕ, j, w |

Table 2. Types of initials according to different manners of articulation. (Letters used here are defined according to Hanyu Pinyin for Mandarin and YuetPing (LSHK, 2002) for Cantonese.)

As shown in Table 2, all initials are divided into seven categories according to the manner of articulation. As shown in Figure 2, the duration of the initial depends largely on the type of consonant it is. For Mandarin, the order of increasing duration is as follows; zero, voiced initials, unaspirated stops, unaspirated affricates, aspirated stops, fricatives, and aspirated affricates, which is quite consistent with Feng's (1985) study with the order: voiced initials, unaspirated stops, unaspirated affricates, fricatives, aspirated stops, and aspirated affricates. In our study, aspirated stops and fricatives exhibit approximately equal duration. For Cantonese, the order of increasing duration is only slightly different. Then the order of increasing duration is as follows; zero, voiced initials, unaspirated stops, aspirated stops, unaspirated affricates, fricatives, and aspirated affricates. Consequently, longer initial duration results in longer syllable duration for both Mandarin and Cantonese[4].

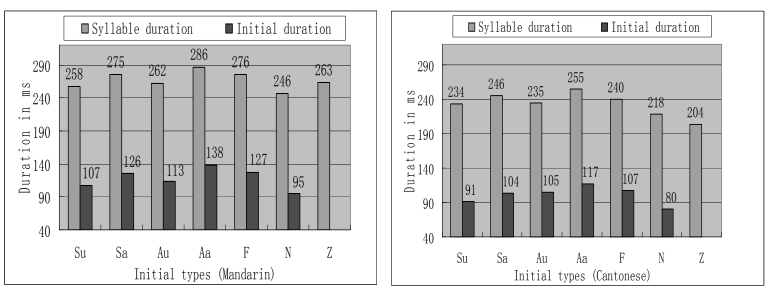



Figure 2. Syllable and initial duration according to different types of initials. The light bars indicate syllable duration, while the dark bars indicate the corresponding initial duration.

2.2. Temporal variation according to types of finals

As shown in Figure 3, the duration of the finals depends on the nature of the ending. For Mandarin, the order of increasing duration is; plain, nasal, and rhotacized [note, there is only one syllable type, i.e. /er/]. For Cantonese, the order of increasing duration is; checked, plain, and nasal. Consequently, longer final duration results in longer syllable duration for both Mandarin and

Cantonese (except for the rhotacized final in Mandarin).

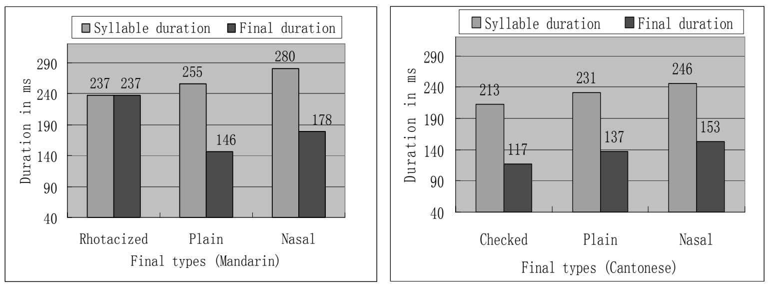



Figure 3. Syllable and final duration according to different types of finals (nature of the endings). The light bars indicate syllable duration, while the dark bars indicate the corresponding final duration.

As shown in Figure 4, in both Mandarin and Cantonese, the duration of the finals also depends on the height of the nucleus vowel. In increasing duration, they are: high vowel, mid vowel, and low vowel.

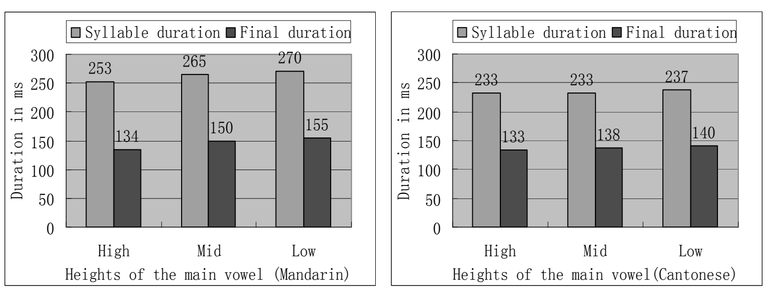
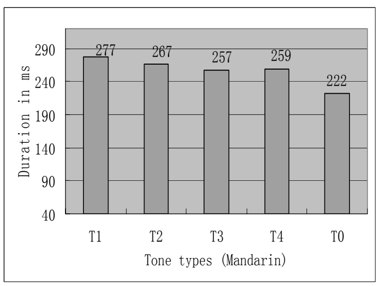



Figure 4. Syllable and final duration according to different types of finals (height of the vowel nucleus). The light bars indicate syllable duration, while the dark bars indicate the corresponding final duration.

2.3. Temporal variation according to types of tones

As shown in Figure 5, in both Mandarin and Cantonese, the duration of the syllables varies according to the tones. For Mandarin, in increasing duration, they are: neutral tone, tone 3, tone 4, tone 2, and tone 1. While for Cantonese, in increasing duration, they are: tone 7, tone 9, tone 5, tone 8, tone 6, tone 4, tone 3, tone 2, and tone 1. In Cantonese, tones 7, 8, and 9 are checked tones, whose duration is shorter than that of their counterparts, tones 1, 3, and 6, respectively.



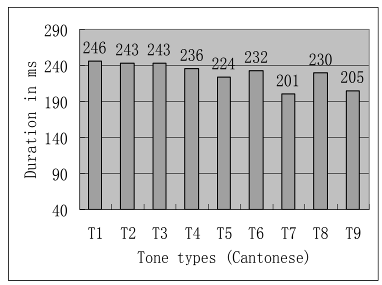




Figure 5. Syllable duration according to different tones.

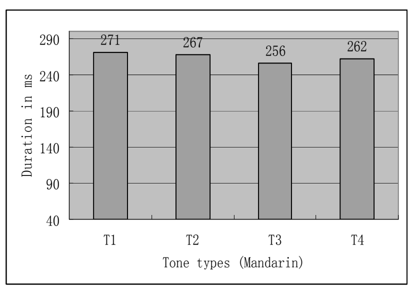



Figure 6. Syllable duration according to different tones (only the sixty selected syllables are involved).

Previous studies indicated that the tone 2 exhibits the longest duration, while tones 1, 3, and 4 have similar duration in non-final position for Mandarin (Feng, 1985). However, only highly controlled data from a small number of subjects (2 young, 2 middle-aged, and 2 seniors) were used. In order to make sure that our results are not due to the intrinsic duration of different base syllables. (As we know from above, different kinds of syllables usually have different duration), sixty base syllables (totally 153,901 tokens) whose tone associations include all four tones (excluding the neutral tone) are selected for computing the duration of different tones. As shown in Figure 6, these results are quite consistent with those in Figure 5 for Mandarin.

### 2.4. Discussion

As described above, different types of initials, finals and tones exhibit different durations. Where do these durational differences come from? Duration is a function of maintaining oral configuration over time during the production of the target sound, which can be an initial, a final or a tone in this context. The amount of time depends on the structure of the oral configuration involved in producing the target sound, which result in different intrinsic duration for different kinds of sounds. And usually, the more stages or components in a sound, longer duration will be needed to produce that sound. For instance, the aspirated stops and aspirated affricates are longer than their corresponding unaspirated versions, because they both are associated an additional stage of 'aspiration' than their counterpart; the nasal finals include both final nuclei and nasal endings, so they exhibit much longer duration than other finals do. However, the unaspirated stop property of initials and unreleased stop property of finals shorten their duration. For instance, the unaspirated affricates show shorter duration than that of fricatives, although the unaspirated affricates include the stage of producing a fricative and all stages of an unaspirated stop; the duration of the checked finals, including their counterparts and the unreleased stops, is shorter than that of their counterparts.

We summarize the duration depicted in Figures 2, 3, and 4 into Tables 3, 4, and 5, respectively. As shown in Table 3, the initial duration varies largely according to different manner of articulation for both Mandarin and Cantonese.

In Mandarin, the corresponding final durations are quite similar (zero initial being the sole exception). Consequently, there is no temporal compensation effect between initial and final duration in this case for Mandarin. For Cantonese, although the corresponding final duration exhibits some variations, they are quite small (143-130 = 13ms) compared to that (117-80 = 37ms) of initial duration. Especially, the same final duration (138ms) is found for the two extreme cases, whose corresponding initials are aspirated affricates and voiced initials (one has the longest duration, while the other has the shortest duration). Therefore, there is also no compensation effect between initial and final duration in this case for Cantonese. (If we partition this problem into only two classes: zero initials and non-zero initials, then the syllables with zero initials exhibit 0ms initial duration, but have the longest final duration. In this case, we may say there are temporal compensation effects in both Mandarin and Cantonese.)

| | Initial types | Su | Sa | Au | Aa | F | N | Z |
|---|---|---|---|---|---|---|---|---|
| Mandarin | Syllable (ms) | 258 | 275 | 262 | 286 | 276 | 246 | 263 |
| | Initial (ms) | 107 | 126 | 113 | 138 | 127 | 95 | 0 |
| | Final (ms) | 151 | 149 | 149 | 148 | 149 | 151 | 263 |
| Cantonese | Syllable (ms) | 234 | 246 | 235 | 255 | 240 | 218 | 204 |
| | Initial (ms) | 91 | 104 | 105 | 117 | 107 | 80 | 0 |
| | Final (ms) | 143 | 142 | 130 | 138 | 133 | 138 | 204 |

Table 3. Syllable, initial and final duration according to different types of initials.

There is only one syllable type, i.e. /er/, for rhotacized final in Mandarin. It will be ignored in this discussion. As shown in Table 4, the final duration depends on the nature of the final endings. In Mandarin, the finals with nasal endings exhibit 32ms (178-146) longer than the finals with plain endings do, but the corresponding initial duration only shows 7ms difference. Therefore, there is no obvious compensation effect between initial and final duration in this case. In

Cantonese, the finals with nasal endings exhibit 36ms (153-117) longer than the finals with checked endings (-p, -t and -k) do, but the corresponding initial duration only shows 3ms difference. Therefore, there is no obvious compensation effect between initial and final duration in this case.

| | Ending types | Rhotacized /checked | Plain | Nasal |
|---|---|---|---|---|
| Mandarin | Syllable (ms) | 237 | 255 | 280 |
| | Initial (ms) | 0 | 109 | 102 |
| | Final (ms) | 237 | 146 | 178 |
| Cantonese | Syllable (ms) | 213 | 231 | 246 |
| | Initial. (ms) | 96 | 94 | 93 |
| | Final. (ms) | 117 | 137 | 153 |

Table 4. Syllable, initial and final duration according to different types of finals (nature of the endings).

| | Heights | High | Mid | Low |
|---|---|---|---|---|
| Mandarin | Syllable (ms) | 253 | 265 | 270 |
| | Initial. (ms) | 118 | 115 | 115 |
| | Final. (ms) | 134 | 150 | 155 |
| Cantonese | Syllable (ms) | 233 | 233 | 237 |
| | Initial (ms) | 100 | 95 | 97 |
| | Final (ms) | 133 | 138 | 140 |

Table 5. Syllable, initial and final duration according to different types of finals (heights of the nucleus).

As shown in Table 5, the final duration depends on the heights of final nucleus. But we note that this dependency is much less in Cantonese. In Mandarin, the finals with low nucleus exhibit 21ms (155-134) longer than the finals with high nucleus do, but the corresponding initial duration only shows 3 ms difference. Therefore, there is no obvious compensation effect between initial and final duration in this case. In Cantonese, the duration difference is too narrow (only 7ms [140-133] difference between the duration of finals with high and low nucleus) to consider temporal compensation effect.

We conclude from the above discussions: in both Mandarin and Cantonese, the duration of the initials and the finals does not influence each other. Therefore, we find no evidence of temporal compensation.

Recall Figures 2, 3, 4, and 5, we note the ranges of variation of syllable duration are 40ms (286-246) and 51ms (255-204) according to different types of initials; 43ms (280-237) and 33ms (246-213) according to different endings of finals; 17ms (270-253) and 4ms (237-233) according to different heights of the final nucleus; 55ms (277-222) and 45ms (246-201) according to different tones, for Mandarin and Cantonese, respectively. If we ignore the case of zero initial which is an extreme case, the ranges of variation of syllable duration are 40ms (286-246) and 37ms (255-218) according to different types of initials, for Mandarin and Cantonese, respectively. Consequently, the ranges of variation in all cases are smaller in Cantonese than that in Mandarin, which is the physical basis of the perception of isochrony. Especially the range of variation according to the heights of final nucleus in Cantonese (237 - 233 = 4ms) is less than one quarter of that of in Mandarin (270 - 253 = 17ms). So we can say Cantonese is more isochronous than Mandarin.

## 3. TONES OF CHINESE SYLLABLES

In this section, following the example of the vowel balloons pioneered by Peterson and Barney (1952), we have analyzed the tones of Mandarin and Cantonese with two parameters; $F_0$ height and $F_0$ slope. We draw minimum ellipses of varying areas to cover 90% of the points, as shown in Figure 7[5]. Please note: the raw $F_0$ values are first transformed to log-scale 5-level[6] values according to the Formula 1 described on page 9 of our previous study (Peng and Wang, 2005); then we calculate the $F_0$ height (average $F_0$ value) and $F_0$ slope of the middle segment of each $F_0$ contour (10% of both leftmost and rightmost $F_0$ values will not be used to calculate the height and slope, in order to decrease the tone-irrelevant variation.); each point in the tone charts represents the average $F_0$ height and $F_0$ slope for a certain tone of a certain speaker; each circle in the ellipses represents the corresponding center of the ellipse.

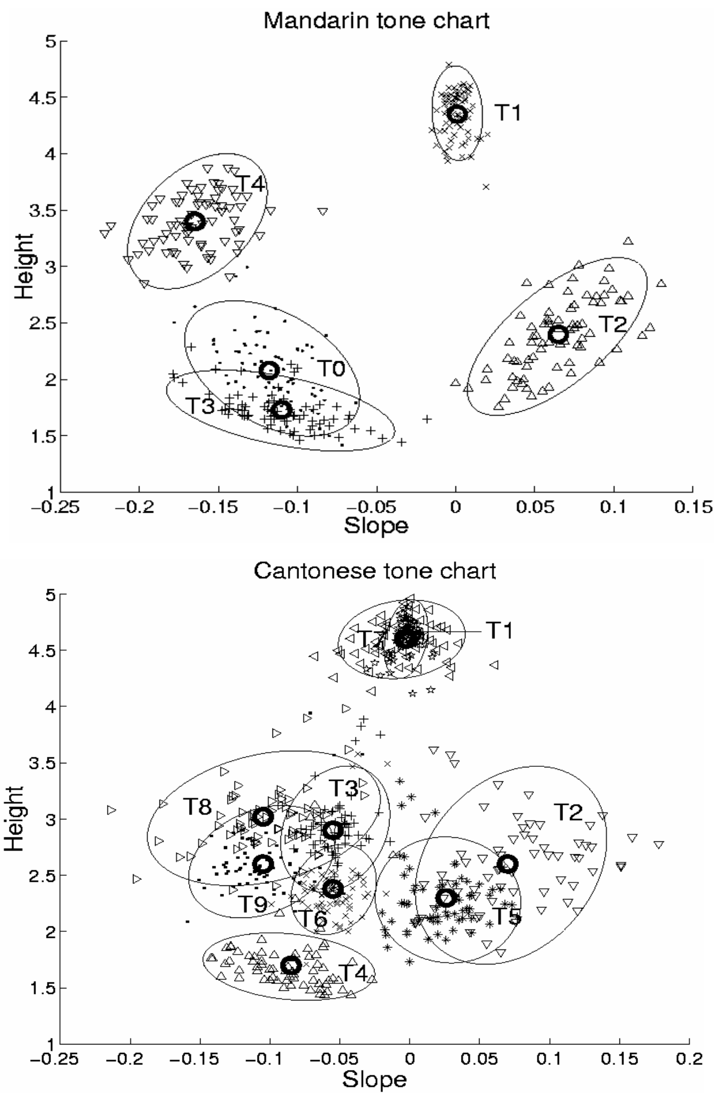



Figure 7. Tone charts for Mandarin and Cantonese.

We find that the Mandarin tones are relatively compact and discretely distributed[7]. These factors allow for successful tone recognition; our score is 83.06% (Peng et al., 2004). The tone value for tone 2 is 35, while the tone value for tone 4 at non-final position is 53. We may wonder why the position of tone 2 in Figure 7 is so much lower than that of tone 4, since they are usually described as 35 and 53 respectively. Xu (2001) concluded that peak delay occurs when there is a sharp $F_0$ rise near the end of a syllable, regardless of the cause of the rise. And the rising tone 2 in Mandarin always satisfies this peak delay condition. Although Xu (2001) studied the $F_0$ peak phenomena only for Mandarin, we hypothesize that peak delay occurs in Cantonese tones under the same condition. Therefore, the Cantonese high rising tone 2 (with tone value 35) locates at a little bit lower position than that of tone 3 (with tone value 33); low rising tone 5 (with tone value 23) locates at a little bit lower position than that of tone 6 (with tone value 22). However, whether there are systematic peak delays in Cantonese rising tones, i.e., tone 2 and tone 5, merits further study.

Nonetheless, the Cantonese tones are much messier. Tones 7, 8 and 9 are checked tones, which can be separated from their counterpart, i.e., tones 1, 3, and 6 respectively, by additionally considering their durations. Since their $F_0$ values correspond to the long tones 1, 3 and 6, respectively, they are usually labeled according to their unchecked tone counterparts. This is done in many transcription schemes, including that of the Linguistic Society of Hong Kong (LSHK, 2002), where only six distinct tones are labeled. Then we redraw the Mandarin and Cantonese tone charts as Figure 8.

As shown in Figure 8, there are no intersections in Mandarin tone chart any more. However in Cantonese, the two pairs, tones 2 and 5, and tones 3 and 6 are quite difficult to separate by both computer programs and human subjects (Peng and Wang, 2005). Therein, the results of the two perception experiments indicated that humans exhibited great difficulty in differentiating tones 3 and 6. These factors allow for less successful tone recognition for Cantonese; our score is 71.50% (Peng and Wang, 2005).

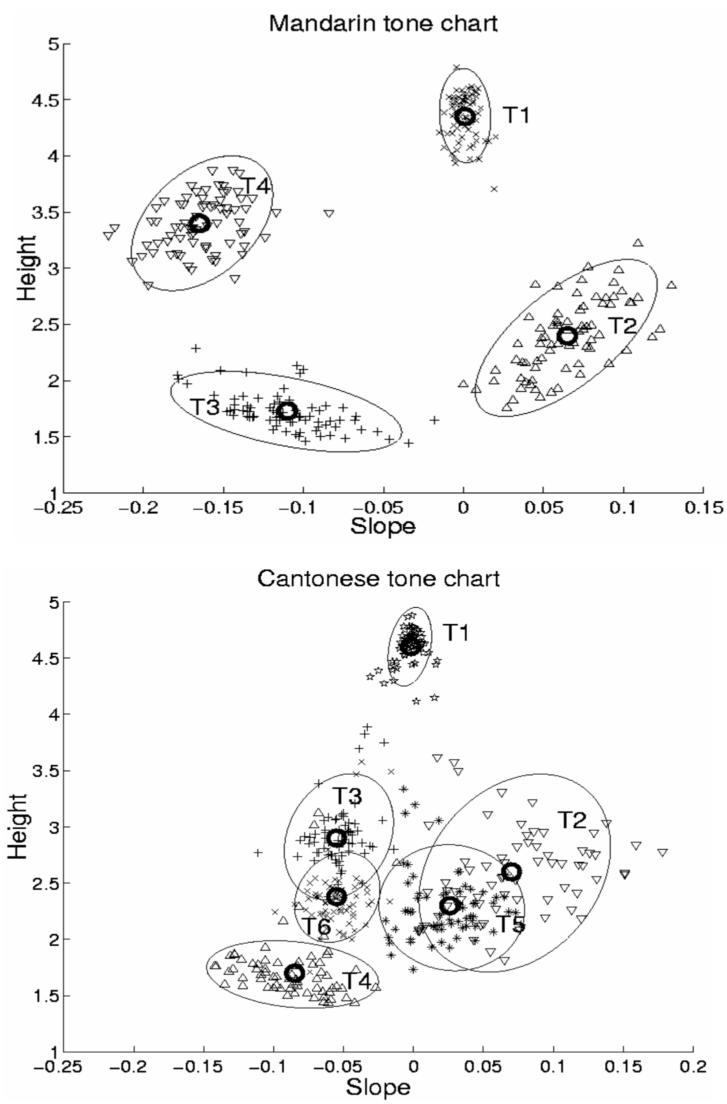



Figure 8. Tone charts for Mandarin and Cantonese, where the neutral tone for Mandarin and the three checked tones for Cantonese are deleted.

Where did the above intersections, i.e., the overlaps between tones 2 and 5 and tones 3 and 6, come from? One hypothesis is that those intersections may come from merger. A merger occurs when two phonemes become one phoneme. This has happened, for instance, for two English vowels, resulting in the homophony of words such as 'meet' and 'meat'. In our case, the two phonemes are tones 2 and 5, and tones 3 and 6. Merger can take place at least in two parallel ways, i.e., by speaker and by word (Wang, 1969). If they are in the process of merging, those two tones will intersect more and more. At the end of merging, they will be merged into one tone. Another hypothesis is that those intersections are due to the incomplete process of splitting; the two splitting tones intersect less and less; finally, there will be no intersection or only little intersection due to the difference among speakers. The first hypothesis is much more plausible regarding the tone development of Chinese dialects (Wang, 1986). This is another interesting question which merits further study. From an optimization perspective, Ke et al. (2003) studied the universal structure of tone systems by using genetic algorithm (Holland, 1975). Their results were not very consistent between their predicted systems and the observed systems. One possible reason for inconsistency is that these tone systems are constantly changing, which might be an optimization process. But further investigation along this line will be useful toward understanding of tone development.

## 4. SUMMARY AND CONCLUSIONS

The syllable duration varies according to different types of initials, finals, and tones; the duration of the initials and finals does not influence each other. Therefore, we find no evidence of temporal compensation; the ranges of variation in syllable duration according to initials, finals, and tones are smaller in Cantonese than that in Mandarin, which is the physical basis of the perception of isochrony. Finally, we have analyzed the tones of Mandarin and Cantonese with two parameters; $F_0$ height and $F_0$ slope. Then, the hypothesis of merger has been adopted to explain the large overlaps of Cantonese tones.

In this study, the continuous speech data was completely segmented by computer programs, unlike previous studies using manually segmented data. The advantage of this method is that huge amounts of data can be used. The weakness is that computer programs cannot guarantee the segmentation is performed as well as phoneticians can. However, the errors from computer programs are quite consistent because they segment speech based on statistical properties of speech units. Therefore, such errors might be minimized by averaging. Then, the results from this kind of study, on the one hand, would give some hints to the underlying linguistic meanings which further trigger linguistic studies of these aspects. On the other hand, they can be used to build models for speech synthesis or automatic speech recognition, serving as a bridge between linguistic knowledge and speech technologies.

## NOTES

1.The work described in this paper was supported by a CERG grant (No. 9040676) awarded to Guanrong CHEN of City University of Hong Kong, and two CERG grants (No. CUHK 1224/02H and CUHK 1127/04H) awarded to William S-Y. WANG of Chinese University of Hong Kong.

2. Hakha Lai is a tone language spoken in Northern Burma and India.

3. The symbol ‘ϕ’ is used here to indicate phonetic null.

4. Zero initial excluded and one other exception: aspirated stops in Cantonese exhibit shorter duration than that of unaspirated affricates and fricatives, but duration of the syllables with aspirated stops is longer than that of syllables with either unaspirated affricates or fricatives.

5. First, the center of an ellipse is calculated by averaging all samples from one tone class. Then linear regression is used to find the major axis of the ellipse (the minor axis is perpendicular to the major axis). By tuning the length of the major and minor axes of the ellipses, we can cover about 90 percent of samples of the target tone class. Finally, the ellipse which has the minimum area is selected for illustration. (MATLAB was used to draw the ellipses here.)

6. We use log-scale 5-level transformation not only because it works well in tone recognition (Peng and Wang, 2005), but also it is consistent with the time-honored selection of number of levels for linguists to transcribe tones (Chao, 1930).

7. The neutral tone intersects with tone 3. This is because the neutral tone is a

highly context-dependent tone. Due to its heavy context dependency, it varies greatly according to its preceding tone. But the overall height of the neutral tone is low. However, these can be separated by further considering their durations.

## REFERENCES


CHAO,Yuan Ren. 1930. “A system of tone letters.” Le Maître Phonétique, 45, 24-27.

FENG, Long 馮隆. 1985. “北京話語流中聲、韻、調的時長 (duration of initials, finals and tones in Beijing dialect).” 北京語音試驗錄 (notes of phonetic experiments in Beijing dialect), edited by Lin, T. and Wang, L. J. Beijing University Publishing, pp. 131-195.

HOLLAND, John H. 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor.

HOUSE, Arthur S. 1961. “On vowel duration in English.” J. Acoust. Soc. Am. 56, 1016-1018.

HOUSE, Arthur S. and FAIRBANKS, Grant. 1953. “The influence of consonant environment upon the secondary acoustical characteristics of vowels.” J. Acoust. Soc. Am. 25, 105-113.

KE, Jinyun, OGURA, Mieko and WANG, William S-Y. 2003. “Optimization models of sound systems using genetic algorithms.” Computational Linguistics, 29.1:1-18.

LEE, Tan, LO, W. K., CHING, P. C. and MENG, Helen. 2002. “Spoken language resources for Cantonese speech processing.” Speech Communication, 36(3-4):327-342, 2002.

LEHISTE, Ilse. 1972. “The timing of utterances and linguistic boundaries.” J. Acoust. Soc. Am. 51, 2018-2025.

LEHISTE, Ilse and WANG, William S-Y. 1976. “Perception of sentence boundaries with and without semantic information.” In Dressler, W. and Pfeiffer, O. (eds.), Phonologica 19, pp. 277-283.

LIEBERMAN, Philip. 1960. “Some acoustic correlates of word stress in American English.” J. Acoust. Soc. Am. 32, 451-454.

LINGUISTIC SOCIETY OF HONG KONG, (LSHK). 香港語言學學會. 2002.

粵語拼音字表(Hong Kong JyutPing Character Table), second ed. Linguistic Society of Hong Kong, 2002.

MADDIESON, Ian. 2004. "Timing and alignment: a case study of Lai." Language and Linguistics, 5.4: 729-755.

MIXDORFF, Hansjörg, LUKSANEEYANAWIN, S., FUJISAKI, H. and CHARNVIVIT, P. 2002. "Perception of tone and vowel quantity in Thai." Proceedings of the 7th ICSLP, 753-756.

PENG, Gang and WANG William S-Y. 2004. "An Innovative Prosody Modeling Method for Chinese Speech Recognition." International Journal of Speech Technology. 7:129-140.

_____. 2005. "Tone recognition of continuous Cantonese speech based on support vector machines." Speech Communication, 45:49-62.

PENG, Gang, ZHENG Hongying and WANG William S-Y. 2004. "Tone recognition for continuous Chinese speech: a comparative study of Mandarin and Cantonese." Proceedings of the 4th International Symposium on Chinese Spoken Language Processing, 233-236.

PETERSON, Gordon E. and BARNEY, Harold L. 1952. "Control method used in a study of the vowels." J. Acoust. Soc. Am. 24, 175-184.

PETERSON, Gordon E. and LEHISTE, Ilse. 1960. "Duration of syllable nuclei in English." J. Acoust. Soc. Am. 32, 693-703.

WANG, William S-Y. 1969. "Competing change as a cause of residue." Language, 45:695-708.

_____. 1973. "The Chinese language." Sci. Amer. 228:50-63.

_____. 1986. "A note on tone development." In Papers in Honor of Wang Li on His Eightieth Birthday, Hong Kong: Chinese Language Society.

WANG, William S.-Y. and LI, K.-P. 1967. "Tone 3 in Pekinese." Journal of Speech and Hearing Research, 629-636.

XU, Yi. 2001. "Fundamental frequency peak delay in Mandarin." Phonetica, 58: 26-52.

YOUNG, Steve J., EVERMANN, G., KERSHAW, D., MOORE, G., ODELL, J., OLLASON, D., POVEY, D., VALTCHEV, V., and WOODLAND, P. C. 2002. HTK: Hidden Markov Toolkit V3.2. Cambridge University.

ZEE, Eric. 1995. "Temporal organization of syllable production in Cantonese."

Proceedings of the 13th International Congress of the Phonetic Sciences, 250-253.

# 漢語音節時長與聲調分析：基于語料庫的普通話和香港話的比較研究

彭剛
香港中文大學

以往關於時長與聲調分析大體都是基于少量幾個發音人的有限語料。這樣就很難區分這些研究成果到底反映的是所研究語言的一般特征，還是少數幾個發音人的個人特征。由于這個原因，它們很難被直接運用到當今的語音科技。在本文中，我們利用錄自大量發音人的大型連續語音數據庫來考查在自然語流中普通話和香港話的時長與聲調特征。我們發現：在普通話和香港話中，聲母時長和韻母時長不存在明顯的補償作用（零聲母情況除外）；香港話音節時長變動小于普通話，因而具有相對比較好的音節等時性。然後，我們依照 Peterson 和 Barney 考查元音的例子，利用聲調曲線的高度和斜率這兩個參數分析了普通話和香港話的聲調。並就聲調的發展討論了幾種語言學假設。