Abstract

As one of the major Chinese dialects, Hakka typically has a tone system with six lexical tones. The traditional 5-level notation of tones in Hakka varies in previous references due to its subjective and relative nature. In order to overcome the limitations of the traditional approach, the command-response model for the process of $F_0$ contour generation is employed to analyze quantitatively the tones in continuous speech of two varieties of Hakka, spoken in Meixian and in Shataukok, respectively. By providing both phonological descriptions to each tone type and quantitative approximations to continuous $F_0$ contours, the model-based approach provides an efficient connection between phonetics and phonology of Hakka tones.

Index Terms: Hakka, tone, $F_0$ contour, command-response model, command pattern

1. Introduction

Hakka dialects (客家話) is a spoken variation of Chinese by the Hakka group whose ancestors migrated from northern China to southern China since 1,700 years ago. It is the 5th largest Chinese dialects today, with a population of over 30 million speakers distributed widely in southern China (mainly in Guangdong, Fujian, and Jiangxi provinces) and Southeast Asia as well as in many overseas Chinese communities elsewhere in the world. Hakka preserves the phonetic features of Chinese spoken in ancient times better than Mandarin.

Among Hakka dialects, the Meixian (梅縣) variety [1-4] in the northeast of Guangdong has always been viewed as the archetypal, forming a sort of standard Hakka that is the most influential. Also, Hakka spoken in Hong Kong forms a largely homogenous variety. Originally named after Shataukok (沙頭角) [4, 5] – a small town well known to be on the border between Hong Kong and mainland China, this variety has also spread its influence to the bordering areas of Guangdong.

Since the most unique feature of Hakka is on tones [1], our present study will be focused on the tonal patterns. After introducing the traditional notation of the tones in these two varieties of Hakka, we will investigate how a fully quantitative approach can be introduced to provide phonetic as well as phonological descriptions of tones, and also to give accurate representations of continuous $F_0$ contours for Hakka speech.

2. Hakka tone systems

In terms of the traditional Chinese phonology, Chinese has four tone categories: Ping (level), Shang (elevating), Qu (departing), and Ru (entering), each of which can be divided into two registers according to the voicing property of the initial: Yin (upper) for voiceless initials and Yang (lower) for voiced initials, hence eight types of tones altogether.

In Meixian and Shataukok Hakka, the division in Yin and Yang registers is still preserved in the tone categories of Ping and Ru, but no longer exists in the tone categories of Shang and Qu. Thus, there are six lexical tones, as listed in Table 1. Syllables of entering (Ru) tones end with an unreleased stop /p/, /t/, or /k/, and hence are shorter than those of non-entering tones. Contrary to the tradition, Yin-Ru gives lower pitch than Yang-Ru in these two varieties of Hakka.

Conventionally, a 5-level tone code (or tone letter) system proposed by Chao [6] is used to represent the $F_0$ features of tones, though for a specific tone language/dialect the code may vary somewhat from one reference to another. The five levels, from 1 to 5, indicate the relative pitch values from low to high, respectively. Several proposals for the code of lexical tones in Meixian Hakka [1-4] and Shataukok Hakka [4, 5, 7] are listed in Table 1, where the underlines denote a faster rate of change (e.g., 32 has a comparable duration to 3 but shorter than 32).

Besides, tone sandhi also occurs regularly in continuous speech of Hakka. In both the two varieties, T1, T4, and T3 before a lower tone (i.e., before T2, T3, or T5; and for T4, also before T4) are converted into 35, 55, and 44 (i.e., T1), respectively [8]. Henceforth, the former two sandhi tones are denoted as T1’ and T4’, respectively, which also need to be modeled in describing $F_0$ contours of continuous speech.

In spite of wide adoption by phoneticians, the tone code system apparently has inherent limitations. First, the codes are usually defined perceptually, and are subjective and relative, which explains why the proposed codes vary with references. Such discrepancies may reflect phonetic differences between individual speakers, or the diversity in subjective judgments by researchers, but are not phonologically meaningful. Second, in continuous speech the actual $F_0$ values change with tonal context, word stress, and phrase intonation, and differ notably from their citation form in isolated syllables. Third, the 5-level codes are discrete and not fully quantitative, and hence cannot characterize the continuous and dynamic nature of $F_0$ values.

<table>
<thead>
<tr>
<th>Tone name</th>
<th>Tone num.</th>
<th>5-level code</th>
<th>Example char.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meixian</td>
<td>Shataukok</td>
<td>Ref.[1] [2] [3] [4,5] [7]</td>
<td></td>
</tr>
<tr>
<td>Ping</td>
<td>T1</td>
<td>44 44 33 44 33</td>
<td>马先生</td>
</tr>
<tr>
<td>Shang</td>
<td>T3</td>
<td>31 31 42 31 32</td>
<td>把学生</td>
</tr>
<tr>
<td>Qu</td>
<td>T5</td>
<td>53 5 52 42 53</td>
<td>罪犯</td>
</tr>
<tr>
<td>Ru</td>
<td>T6</td>
<td>5 5 44 5 5</td>
<td>禽兽</td>
</tr>
</tbody>
</table>

In Table 1, the traditional description of tones in Meixian and Shataukok is illustrated as a reference.
To overcome these limitations, we apply the command-response model for the process of $F_0$ contour generation [9], to analyze tones in continuous speech of Meixian and Shataukok Hakka in a fully quantitative way.

3. The command-response model for $F_0$ contour generation for tone languages

Figure 1 shows the diagram of the model for tone languages. It describes $F_0$ contours in the logarithmic scale as the sum of phrase components, tone components and a baseline level $lnF_0$. The phrase commands produce phrase components through the phrase control mechanism, giving the wide-range shape of $F_0$ contour, while the tone commands generate tone components through the tone control mechanism, characterizing the local $F_0$ changes. Both mechanisms are assumed to be critically-damped second-order linear systems. Tone languages usually require both positive and negative tone commands due to faster local $F_0$ changes. For a given tone language, a specific set of tone command patterns needs to be defined.

The model has been successfully applied to several Chinese dialects including Standard Mandarin, Cantonese, Shanghainese, and Lanqi [10]. In those dialects, tones can usually be characterized phonologically by the ternary polarities (+, 0, −) of tone commands (here 0 indicates null tone command) in the earlier and the later parts of a syllable; exceptionally, we also need two phonologically distinct levels (in terms of command amplitude) for negative tone commands to distinguish the tones of Yin-Ping and Yang-Qu in Cantonese.

4. Speech data

Since the pitch values of tones are relative, investigation of $F_0$ contours in isolated syllables cannot reveal the relationship between tones in continuous speech as well as the interaction between tones and sentence intonation. Hence, we used continuous speech but conducted a controlled experiment.

The speech materials for Meixian Hakka and Shataukok Hakka consist of a fixed carrier frame as follows respectively:

Meixian: “E1 zak5 he4 _ sid.”

Shataukok: “Ngia1 zak5 he4 _ su4.”

(“This is the character _”)

which share the same meaning and the same tone composition in each syllable (the first syllables /e/ and /ngia/ are lexically in T3, but change into T1 due to tone sandhi since they precede a T5 syllable). Each of the following target syllables (carrying each lexical tone) is embedded at the underlined position:

<table>
<thead>
<tr>
<th>T1</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>/sa1/</td>
<td>/sa4/</td>
</tr>
</tbody>
</table>

They share a rather similar syllable composition and hence the segmental effect is minimized. Another set of syllables /fui(k)/ is also used, but since the results are consistent, here we only show the analysis on /sat(i)/. The target syllables are embedded at an utterance-medial position, because $F_0$ at the beginning of an utterance is usually the most variable, while the utterance-final tone patterns can be affected by the final $F_0$ lowering. For reference, these syllables are also recorded in the isolated form.

It is noted that in the above sentence /sa4/ is converted into T4’ since it precedes another T4 syllable. Also, the sandhi T1’ needs to be analyzed. Hence, we added the following two sentences by changing the sentence-final syllable:

Meixian/Shataukok: “E1/Ngia1 zak5 he4 sa1’ teu2/tiu2.”

(“This is a place in HK.”)

Meixian/Shataukok: “E1/Ngia1 zak5 te4 sa4 diau1.”

(“This is a Chinese novel.”)

The target syllable /sa1/ is converted into a rising tone T1’, while /sa4/ keeps the lexical T4 without any tone sandhi.

The informants of Meixian Hakka include four native speakers from the urban area of Meixian (two male and two females), while the informants of Shataukok Hakka include two native speakers from New Territories, HK (one male and one female; the male lives abroad now). Each sentence was uttered four times by each informant.

The $F_0$ values are extracted by a modified autocorrelation analysis of the LPC residual signal. Since most the utterances share a fixed carrier frame and differ only in the target syllables, tone command patterns for each tone type can first be inferred heuristically from visual comparison of the pooled $F_0$ curves of all the tone types in the target syllable, following the method given in [11]. Namely, a flat curve implies null or one tone command, while a curve involving a rising/falling section suggests at least one tone command; and a flat curve lying right in the middle can usually be assumed to have null tone command, while those above and below correspond to positive and negative tone commands, respectively. The validity of the assumed command patterns is then verified by analysis-by-synthesis of the entire $F_0$ contours [11].

5. Analysis of tones in Meixian Hakka

Tone patterns are observed to be consistent between the four speakers. By comparison of time-normalized $F_0$ curves of all the tone types in the target syllable, we first make heuristic assumptions to tone command patterns, which can then be adjusted iteratively during analysis-by-synthesis of entire $F_0$ contours, until a set of patterns consistently giving both highly accurate $F_0$ fitting and concise tonal distinction is obtained.

The results of analysis on eight utterances (for each target tone, respectively) by a male speaker are shown in Fig. 2. The crossed symbols indicate the measured $F_0$ values, while the solid, dotted, and dashed lines indicate the approximated $F_0$ contours, the baseline frequencies, and the contributions of phrase components, respectively. The differences between the approximated $F_0$ contours and phrase components correspond to the tone components. In addition to an utterance-initial phrase command (larger) before the beginning of the utterance, the speaker tends to place a second phrase command (smaller) immediately before the embedded target syllable, giving an unconscious emphasis on the target syllable. Especially, the baseline $F_0$ and the amplitudes of the 2nd phrase commands in the eight utterances are almost set equal, such that the tone command patterns can be derived in an equal condition.

The analysis results indicate that the command patterns for the tones in Meixian Hakka are:

<table>
<thead>
<tr>
<th>T1</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>negative</td>
<td>overly negative</td>
</tr>
</tbody>
</table>

2634
Figure 2: Analysis-by-synthesis of $F_o$ contours of the utterances of Meixian Hakka by a male speaker.

T3: negative only in the later part  0 –
T4: positive to null/negative       + 0/–
T5: negative (short)            [– –]
T6: positive (short)           [+ +]
sandhied T1: negative to positive – +
sandhied T4*: positive          + +

As shown, these patterns are valid not only for the embedded target syllables but also for other syllables in the carrier frame. Thus, there are at most two separate tone commands within a syllable. In terms of the polarities (+: positive; 0: zero; –: negative) of tone commands respectively in the earlier and the later parts of a syllable, which do not necessarily correspond to equal duration, the tones can be represented phonologically by a set of symbols as listed above in the rightmost column. Namely, two virtual commands are used to give a consistent representation. The brackets indicate entering tones which are shorter than non-entering tones. Especially, both T1 and T2 have a negative command, but T2 shows larger absolute command amplitude than T1. Hence, a special symbol “–” is employed for T2 to indicate an overly negative command.

It should be noted that the utterance-final T4 can give two types of tone command patterns, i.e. ‘+ 0’ and ‘+ –’. This kind of variation, however, is non-distinctive and does not affect tone identification. It also exists in other Chinese dialects. For example, in Shanghaiese Yin-Ping tone is usually denoted as 53 (i.e. ‘+ 0’), but can also be uttered more steeply as 52 or 51 (i.e. ‘+ –’) especially at the end of an utterance [12].

The above symbols are qualitative and for phonological purpose only. With a set of quantitative command parameters (both timing and amplitude), the model can then provide fully quantitative representations to continuous $F_o$ contours. For this purpose, quantitative distinction between tones is also needed. For example, the command onset time for T5 is apparently later than that for T6. Due to space limitations, however, such quantitative analysis will be presented in our future work.

An approximate correspondence can be seen between the patterns and the traditional notations, except that ‘–’ for T1 notably does not coincide with 44 or 33 which is derived from isolated syllables. We found that for all the four speakers T1 uttered in isolation is always higher than in connected speech. This conspicuous distinction was ignored in previous works.

6. Analysis of tones in Shataukok Hakka

Tone patterns are observed to be largely consistent across the two speakers, except that the male living abroad has already merged T3 into T1. The results of analysis-by-synthesis on $F_o$ contours of seven utterances (for each target tone except T4, which is not given here but can be derived from the utterance-final syllable /su4/) by the female speaker are shown in Fig. 3, as can be interpreted similarly as Fig. 2. The analysis indicates that the command patterns for tones in Shataukok Hakka are:

<table>
<thead>
<tr>
<th>Tone</th>
<th>Command Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1:</td>
<td>null 0 0</td>
</tr>
<tr>
<td>T2:</td>
<td>overly negative = =</td>
</tr>
<tr>
<td>T3:</td>
<td>negative – –</td>
</tr>
<tr>
<td>T4:</td>
<td>positive to null/negative + 0/–</td>
</tr>
<tr>
<td>T5:</td>
<td>null (short) [0 0]</td>
</tr>
<tr>
<td>T6:</td>
<td>positive (short) [+ +]</td>
</tr>
</tbody>
</table>

sandhied T1: negative to positive – +
sandhied T4*: positive + +

Likewise, there are at most two separate tone commands
Let the abscissa and the ordinate indicate the polarity of the tone command occurring in the earlier and the later parts of a syllable, respectively. A novel phonological structure of the tone systems of Meixian Hakka and Shataukok Hakka can then be given in Fig. 4. Unlike the traditional approaches, this representation is directly associated with a generative model. In the constellation, the hollow and solid circles denote lexical and sandhi tones, respectively, while the triangles denote a steeper variant of T4. The points occupied simultaneously by two types of tones are denoted by double circles—the two tones differ in duration (the entering tone is shorter). The two varieties of Hakka differ only slightly in T1, T3, and T5.

Based on the command-response model, we have analyzed the tones in continuous speech of two varieties of Hakka (Meixian and Shataukok) in an objective and quantitative way. By incorporating the distinctions in polarity and amplitude of tone commands in different parts of a syllable, the model provides both phonological descriptions to each tone type and fully quantitative representations to continuous $F_0$ contours.

8. References