Comparison of F0 range in Spontaneous Speech in Kammu Tonal and Non-Tonal Dialects

Karlsson, Anastasia; House, David; Svantesson, Jan-Olof; Tayanin, Damrong

2011

Link to publication

Citation for published version (APA):
COMPARISON OF F₀ RANGE IN SPONTANEOUS SPEECH IN KAMMU TONAL AND NON-TONAL DIALECTS

Anastasia Karlsson¹, David House², Jan-Olof Svanessson¹, Damrong Tayanian¹

¹Dept. of Linguistics and Phonetics, Lund University; ²Dept. of Speech, Music and Hearing, KTH, Stockholm
anastasia.karlsson@ling.lu.se, davidh@speech.kth.se, jan-olof.svanessson@ling.lu.se

ABSTRACT
The aim of this study is to investigate whether the occurrence of lexical tones in a language imposes restrictions on its pitch range. Kammu, a Mon-Khmer language spoken in Northern Laos comprises dialects with and without lexical tones and with no other major phonological differences. We use Kammu spontaneous speech to investigate differences in pitch range in the two dialects. The main finding is that tonal speakers exhibit a narrower pitch range. Thus, even at a high degree of engagement found in spontaneous speech, lexical tones impose restrictions on speakers' pitch variation.

Keywords: pitch range, tone, timing, intonation, Kammu, Khmu.

1. INTRODUCTION
There is recurrent interest in comparing the F₀ range of different languages in the broad context of investigating language-specific use of F₀. There has been general speculation that different pitch ranges and other characteristics of F₀ can comprise a part of the phonetic structure of a language and thus differ systematically between languages (see [7] for a review). One question concerns the influence of lexical tone on intonation, and this has generated the hypothesis that tone languages may have an overall larger F₀ range than non-tonal languages by virtue of the additive effect of the lexical tones being superimposed on the intonation contour. Several studies have supported this hypothesis, while in other studies no difference in pitch range between tonal and non-tonal languages was found. In some studies, the opposite tendency has been observed where tone languages display a smaller F₀ range.

In many of the studies supporting the hypothesis, Mandarin Chinese has been compared with English. In a study of broadcast news speech [11], it was found that Mandarin has a wider pitch range and more F₀ fluctuations than English. This is discussed in terms of the effect of lexical tones.

In [12], where a bilingual Chinese-English corpus was used to develop a mixed-language speech synthesis system, the pitch range of the English words was larger in the bilingual corpus than in the English one. These results are discussed in terms of the influence of the Chinese lexical tones on the corpus.

In [7], Mandarin was found to have a larger pitch range than English in single-word utterances. However, this effect was not seen in prose passages. These results highlight the effect of speech material. Eady [1] found no difference in F₀ standard deviations between English and Mandarin.

Another interesting and relevant area of study is the modification of F₀ which takes place in infant directed speech. [2] reported an exaggeration of F₀ range in infant directed speech in Mandarin. However, in a study comparing infant directed speech in Australian English to Thai [8], it was found that F₀ range was more exaggerated in Australian English than in Thai. These results are discussed in terms of restriction on pitch excursions in infant directed speech due to lexical tone.

Lexical tone can thus be seen to either restrict F₀ range or enhance it, varying across language, speech material, and speaking style. By investigating a language in which lexical tone is a characteristic of one dialect but absent from another dialect, we aim to study the effect of lexical tone on F₀ range.

Kammu is a Mon-Khmer language spoken by some 600,000 people mainly in Northern Laos, but also in adjacent areas of Vietnam, Thailand and China. One of its main dialects has lexical tones (high or low) on each syllable, while the other main dialect lacks lexical tones. The tones have developed by the merger of voiceless and voiced initial consonants. Other differences between the dialects are marginal, and speakers of different
dialects understand each other without difficulty [9, 10].

Earlier studies of Kammu have shown a compressed F0 range in the tonal dialect as compared to the non-tonal dialect in read speech [3, 5, 6]. In this study we aim to investigate F0 range differences in spontaneous speech material.

2. METHOD

In spontaneous speech, speakers tend to show a higher degree of engagement and to be more relaxed, so this kind of material should give us a more realistic picture about the actual freedom in pitch variations for tonal speakers.

To investigate F0 range differences in Kammu tonal and non-tonal dialects we chose spontaneous monologues about growing and preparing rice, an activity that most Kammu speakers are well acquainted with. We were able to obtain material with a very homogeneous informational and lexico-syntactical structure where new information is placed at the end of the utterance and is focussed. F0 range was investigated on both the phrasal and lexical level.

2.1. F0 range on the phrase level

F0 range variation on the phrase level was captured by measuring the difference between the F0 maxima and minima in each prosodic phrase following [4]. The prosodic phrase is defined by its tonal shape, having a high boundary tone on its right edge. This boundary tone occurs for both tonal and non-tonal speakers. In the tonal dialect it is suppressed when the phrase final word has low lexical tone.

Ten tonal and ten non-tonal speakers were included in the study. F0 maxima and minima were measured in 168 phrases for the non-tonal and 110 phrases for the tonal speakers.

2.2. F0 range and timing on the lexical level

To investigate differences in F0 range between the two types of dialects on the lexical level, the utterance final words were chosen. The utterance final word is produced with a tonal gesture (rise-fall) having several pragmatic functions. It signals sentence accent, focus, speaker engagement and discourse structure [3, 4, 5]. Because of the between-speaker homogeneous structure of the rice monologues we were able to obtain a fairly good overlapping material with the same utterance final words for the speakers. Two events were investi-
speakers. As seen in the tables, women have narrower pitch ranges than men in both groups. Wilcoxon’s rank sum test shows that this difference is significant ($p < 0.02$ in both groups).

### 3.2. $F_0$ range and timing on the lexical level

Pitch range and timing on the lexical level are shown in Tables 3 and 4, and words with VN and VVN rhymes are illustrated in Figures 1 and 2, respectively. In these tables and figures, 0 stands for non-tonal speakers, and H and L for tonal speakers’ words with high and low tone.

As seen in Table 3, non-tonal speakers exhibit a larger average $F_0$ rise for all types of rhymes. The difference is significant (Wilcoxon’s rank sum test, $p < 0.001$) for all rhyme types except VC. The tonal fall on syllables with a sonorant coda is also larger for non-tonal speakers ($p < 0.01$ for VN and $p < 0.001$ for VVN). In words with a non-sonorant coda the fall is truncated.

**Table 3**: $F_0$ range in utterance final words (means in semitones).

<table>
<thead>
<tr>
<th>type</th>
<th>rise</th>
<th>fall</th>
</tr>
</thead>
<tbody>
<tr>
<td>VN</td>
<td>3.8</td>
<td>0.6</td>
</tr>
<tr>
<td>VVN</td>
<td>4.6</td>
<td>0.9</td>
</tr>
<tr>
<td>VC</td>
<td>3.9</td>
<td>2.4</td>
</tr>
<tr>
<td>VVC</td>
<td>4.2</td>
<td>0.7</td>
</tr>
</tbody>
</table>

**Table 4**: Timing of the rhymes of utterance final words (means in ms).

<table>
<thead>
<tr>
<th>vowel duration</th>
<th>VN</th>
<th>VVN</th>
<th>VC</th>
<th>VVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>16</td>
<td></td>
<td>228</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>126</td>
<td>260</td>
<td>109</td>
<td>212</td>
</tr>
<tr>
<td>L</td>
<td>132</td>
<td>250</td>
<td>86</td>
<td>232</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>coda duration</th>
<th>0</th>
<th>237</th>
<th>201</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>126</td>
<td>160</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>228</td>
<td>170</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>peak location</th>
<th>0</th>
<th>123</th>
<th>158</th>
<th>83</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>43</td>
<td>77</td>
<td>84</td>
<td>69</td>
</tr>
<tr>
<td>L</td>
<td>78</td>
<td>120</td>
<td>38</td>
<td>75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>number of words</th>
<th>0</th>
<th>16</th>
<th>13</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>10</td>
<td>22</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>L</td>
<td>20</td>
<td>22</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Timing is also different for the two dialects (Table 4). The non-tonal speakers have a later tonal peak than tonal speakers; the difference is significant at the 0.05 level except for VC. In VN words, the peak tends to be realized within the coda for non-tonal speakers. Tonal speakers show a tendency for later timing of the tonal peak in words with low lexical tone compared to words with high tone. These tendencies are found for VN, VVN and VVC words, though variation is rather large, and the difference is not statistically significant. The pattern in VC words is less clear.

**Figure 1**: Duration of vowel and coda consonant (ms) in VN rhymes for non-tonal and tonal speakers. Timing of the $F_0$ maximum is shown with a thick line. The pitch range in semitones is displayed in the lower plot. The three points are $F_0$ start, $F_0$ maximum and $F_0$ minimum in the rhyme of the syllable. The boundary between vowel and consonant is shown with a thick line.

**Figure 2**: Duration of vowel and coda consonant (ms), timing of $F_0$ maximum (ms) and pitch range (semitones) in VVN rhymes.
4. DISCUSSION

The main result of this study is that the tonal dialect displays a narrower pitch range compared to the non-tonal dialect. This is consistent with recent findings for read speech [3, 5, 6]. Thus, even though there is a higher degree of engagement in spontaneous speech, lexical tones impose restrictions on pitch variation. This is consistent with the prosodic feature hierarchy proposed for Kammu [5] with lexical tones being superordinate to sentence intonation. It is interesting that our result is opposite to what was found for Chinese [11, 12]. One explanation could be that Kammu has a simpler tone system with only two level tones, while Chinese has a more complex system with contour tones. In Kammu, the difference between the low and high tone is often relatively small [10], which may also restrict the use of large pitch excursions.

However, these results are in line with those found for infant-directed speech in [8] where F₀ range was more exaggerated in Australian English than in Thai. It could be that in more engaged speech, e.g. infant-directed and spontaneous, lexical tones become more restrictive in their influence on the intonation contour.

In Figures 1 and 2 it can be seen that the words with low tone actually have a greater relative excursion than high-tone words. In a study concerning focus in Kammu [5] it was shown that an expanded F₀ range on focus is greater when focus is on words with low tone compared to words with high tone. This may be due to restrictions on the addition of a high focal accent to the high lexical tone.

Timing of the focal gesture is also different between the two dialects. Non-tonal speakers have a later timing, and in words with VN rhyme they use the entire rhyme by placing the F₀ peak on the nasal. In the tonal dialect focal gesture is restricted to the vowel kernel, which may be the reason for generally longer vowel duration in the tonal dialect. This is consistent with observations from a larger read material.

To sum up, we find systematic differences between the two types of dialects. Occurrence of tones influences both overall pitch range and timing of the focal gesture. F₀ range is narrower in the dialect with lexical tones.

Non-tonal speakers use broader F₀ range. They produce focal accent with a larger excursion and later timing. For the tonal speakers, the high degree of engagement found in spontaneous speech seems not to override lexical tones: the necessity to uphold the identities of the tones restricts the variation of F₀ range.

5. ACKNOWLEDGEMENTS

The work reported here was made within the project Separating Intonation from Tone, financed by the Bank of Sweden Tercentenary Foundation.

6. REFERENCES