COMPARISON OF F_0 RANGE IN SPONTANEOUS SPEECH IN KAMMU TONAL AND NON-TONAL DIALECTS

Anastasia Karlsson[a], *David House*[b], *Jan-Olof Svantesson*[a] & *Damrong Tayanin*[a]

[a]Department of Linguistics and Phonetics, Lund University, Sweden;
[b]Department of Speech, Music and Hearing, KTH, Stockholm, Sweden
anastasia.karlsson@ling.lu.se; davidh@speech.kth.se; jan-olof.svantesson@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 F_0 range of different languages in the broad context of investigating language-specific use of F_0. There has been general speculation that different pitch ranges and other characteristics of F_0 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 F_0 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 F_0 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 F_0 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 F_0 standard deviations between English and Mandarin.

Another interesting and relevant area of study is the modification of F_0 which takes place in infant directed speech. [2] reported an exaggeration of F_0 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 F_0 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 F_0 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 F_0 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 F_0 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 F_0 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 F_0 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. F_0 range was investigated on both the phrasal and lexical level.

2.1. F_0 range on the phrase level

F_0 range variation on the phrase level was captured by measuring the difference between the F_0 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. F_0 maxima and minima were measured in 168 phrases for the non-tonal and 110 phrases for the tonal speakers.

2.2. F_0 range and timing on the lexical level

To investigate differences in F_0 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 investigated: F_0 range and the timing of the focal gesture. For this purpose we measured the start value, the maximum and the final value of F_0 in the rhymes of utterance final words. Durations of vowels and sonorant codas were also measured.

The timing of the gesture was captured by measuring the duration of the rhyme up to the point where the F_0 maximum value occurs.

Kammu has contrasting vowel length, and the rhymes are analyzed into four groups: VN, VVN, VC, VVC (V = short vowel, VV = long vowel, C = stop, N = sonorant). The material includes 7 non-tonal and 13 tonal speakers.

3. RESULTS

3.1. F_0 range on the phrase level

The F_0 range over a phrase, i.e. the difference between the F_0 maximum and minimum in semitones, is shown as a mean value for each speaker, in Tables 1 and 2, for non-tonal and tonal speakers, respectively.

Table 1: F_0 range in prosodic phrases (means in semitones). Non-tonal speakers.

speaker	sex	mean range	*N*
1	m	15.29	30
3	m	15.02	21
7	m	14.00	14
9	m	13.45	21
8	m	12.33	18
4	m	10.72	33
5	m	9.40	12
10	f	6.48	5
12	f	6.40	7
11	f	6.08	7

Table 2: F_0 range in prosodic phrases (means in semitones). Tonal speakers.

speaker	sex	mean range	*N*
22	m	11.21	8
13	m	9.73	10
19	m	9.57	14
20	m	9.37	11
18	f	9.04	13
24	m	8.58	8
26	f	8.46	8
17	f	6.94	13
25	f	6.90	11
21	f	6.01	14

Although the F_0 range on the phrase level shows great variation within each dialect, some observations can be made. Pitch variation is larger in non-tonal speakers, means ranging between 6.08 and 15.29 St, compared to 6.01-11.21 St in tonal 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).

type	rise			fall		
	0	H	L	0	H	L
VN	3.8	0.6	0.9	8.8	3.9	5.3
VVN	4.6	0.9	1.0	11.3	6.2	4.8
VC	3.9	2.4	0.6	0.2	0.0	0.7
VVC	4.2	0.7	1.4	1.1	0.7	0.5

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

		VN	VVN	VC	VVC
vowel duration	0	116	228	108	200
	H	126	260	109	212
	L	132	250	86	232
coda duration	0	237	201	—	—
	H	176	160	—	—
	L	228	170	—	—
peak location	0	123	158	83	132
	H	43	77	84	69
	L	78	120	38	75
number of words	0	16	13	11	13
	H	10	22	9	12
	L	20	22	3	4

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.

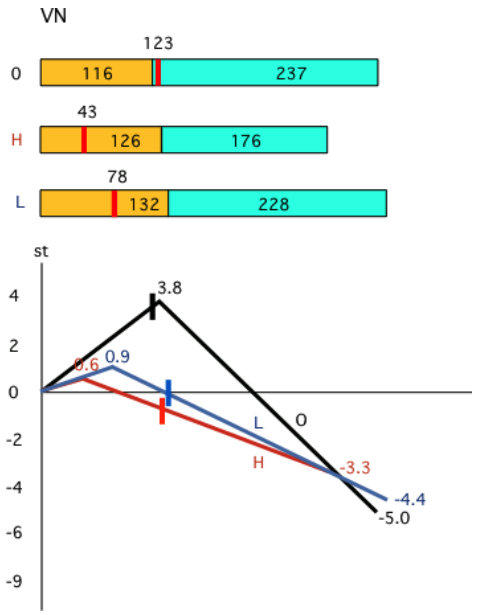
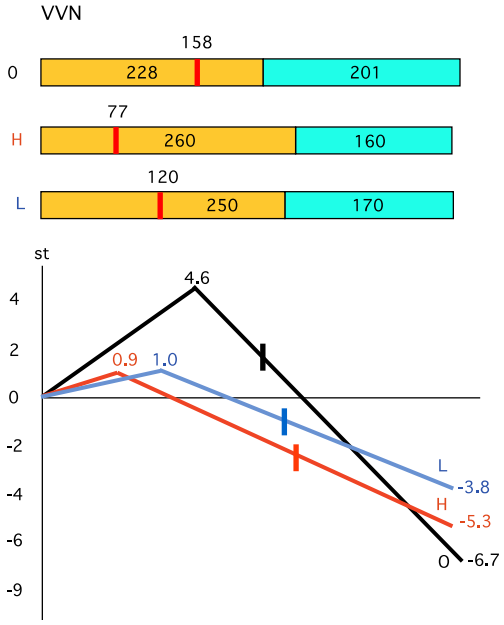


Figure 2: Duration of vowel and coda consonant (ms), timing of F_0 maximum (ms) and pitch range (semitones) in VVN rhymes.



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 F_0 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 F_0 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 F_0 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. F_0 range is narrower in the dialect with lexical tones.

Non-tonal speakers use broader F_0 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 F_0 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

[1] Eady, S.J. 1982. Differences in the F0 patterns of speech: Tone language versus stress language. *Language and Speech* 25, 29-42.

[2] Grieser, D.L., Kuhl, P.K. 1988. Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology* 24, 14-20.

[3] House, D., Karlsson, A., Svantesson, J-O., Tayanin, D. 2009. The phrase-final accent in Kammu: effects of tone, focus and engagement. *Proceedings of Interspeech 2009* Brighton, 2439-2442.

[4] Karlsson, A. 2011. Prosodic features of Kammu tonal and non-tonal dialects: Read and spontaneous speech. In Endo, M., Saitô, Y. (eds.), *Tone, Accent and Intonation in Eastern Eurasian Languages*, 19-28. The 18th Meeting of the Linguistic Circle for the Study of Eastern Eurasian Languages, Aoyama Gakuin University, Tokyo.

[5] Karlsson, A., House, D., Svantesson, J-O., Tayanin, D. 2010. Influence of lexical tones on intonation in Kammu. *Proceedings of Interspeech 2010* Makuhari, Japan, 1740-1743.

[6] Karlsson, A., Svantesson, J-O., House, D., Tayanin, D. 2011. Tone restricts F_0 range and variation in Kammu. *Proceedings of Fonetik 2011* KTH, Stockholm.

[7] Keating, P., Kuo, G. 2010. Comparison of speaking fundamental frequency in English and Mandarin. *UCLA Working Papers in Phonetics* 108, 164-187

[8] Kitamura, C., Thanavishuth, C., Burnham D. and S. Luksaneeyanawin, S. 2001. Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development* 24, 372-392.

[9] Svantesson, J-O. 1983. *Kammu Phonology and Morphology.* Lund: Gleerup.

[10] Svantesson, J-O., House, D. 2006. Tone production, tone perception and Kammu tonogenesis. *Phonology* 23, 309-333.

[11] Yuan, J., Liberman, M. 2010. F0 declination in English and Mandarin broadcast news speech. *Proceedings of Interspeech 2010* Makuhari, Japan, 134-137.

[12] Zhang, Y., Tao, J. 2008. Prosody modification on mixed-language speech synthesis. *Proc. 6th International Symposium on Chinese Spoken Language Processing, ISCSLP2008* Kunming, 253-256.