Realization and Perception of Tones in Mono- and Polysyllabic Words in Thai

Hansjörg Mixdorff, Patavee Charnvivit, Sudaporn Luksaneeyanawin***

*Faculty of Computer Science and Media, TFH Berlin University of Applied Sciences, Germany
mixdorff@tfh-berlin.de
**Centre for Research in Speeech and Language Processing, Chulalongkorn University, Bangkok, Thailand
patavee@gmail.com, Sudaporn.L@chula.ac.th

Abstract. *This paper presents results concerning the relationship between word structure in terms of number of syllables and tonal realization in Thai. It examines whether the fact that a word is longer implies certain tonal reductions. Our hypothesis is that a mono-syllabic word will be uttered more carefully than a polysyllabic word due to the potentially larger number of possible confusable words. We also examine whether the total number of syllables in a word has an effect, creating more tonal reductions in longer than in shorter words, as well as the target word's position in a sentence. A list of frequent Thai words was statistically analyzed regarding the occurrences of syllable/tone combinations in words of varying length. Six sets of syllables were selected which can carry at least four of the five tones of Thai and occur as monosyllabic words, as well as in varying positions in two- to four-syllable prosodic words. A set of two- to four syllable words was then developed, each of which were inserted into a carrier sentence and recorded. The target syllables were then extracted from their original context and presented to native speakers of Thai who had to decide which tone they perceived. The results of the perception test indicate that the size of the word indeed influences the recognition rate which is also higher for stressed than for unstressed syllables.*

1. Introduction

Syllabic tones in tone languages are connected with distinct *F0* patterns (rising, falling etc.). Thai has five different lexical tones: mid (0), low (1), falling (2), high (3), and rising (4) (commonly used tone indices are given in parentheses). Research has shown that these tones are associated with distinct *F0* patterns which are strongly influenced by the tonal context (tonal co-articulation between subsequent syllables) (Abramson, 1979) focus, as well as sentence intonation (Luksaneeyanawin, 1998), (Mixdorff et al., 2002). Typical words of Thai can contain up to five syllables. The question therefore arises whether the longer words depend as much on the correct tonal realization for disambiguation like, for instance, monosyllables. Since a very long word has a smaller potential for confusions than a short one, this might lead to a less careful realization of tones. The current study therefore addresses the issue whether the degree of reduction is directly related to the number of syllables in a word. We extract monosyllabic tokens from their contexts and present them to native speakers of Thai in a perception test, in order to quantify the tonal reductions.

2. Speech Material

We conducted a statistical analysis of a frequency list of Thai words containing 9536 entries compiled at the Centre for Research in Speeech and Language Processing, Chulalongkorn University,

Bangkok (CRSLP). Table 1 displays some statistics of the database regarding words of between one and five syllables. Although the frequency list contains entries with up 10 syllables these larger units are rather idioms than common words. As can be seen, the predominant word structure is disyllabic.

Table 1. *Statistics of Thai word list. Frequency and percentage of words with 1-5 syllables.*

number of syllables	frequency	percentage
1	2040	21.4
2	3999	41.9
3	2000	21.0
4	816	8.6
5	324	3.4

We selected the following six syllables which exist as meaningful monosyllabic words carrying at least four of the five tones (transcription given in Thai SAMPA, developed at CRSLP):

syllable	khaa	sii	naa	thaa	phaa	caa
existing tones	0-4	0-4	0-4	0, 2-4	0-2, 4	0-2, 4

Then a set of two- to four-syllable words was developed in which the target syllable occurs at varying locations. Since the last syllable in a Thai word by default carries the word stress (Mixdorff et al. 2003) we made sure that this condition also occurred. The ultimate list of words contained a total of 203 words. Pairs of these words were inserted into the carrier sentence: "Phom maj daaj phuut waa X khrap, txx phuut waa Y khrap"- "*I didn't say X, but said Y"*, so that each word once occurred in the X position and once in the Y position, yielding a total of 406 word tokens. The resulting sentences were randomized and recorded on a PC at 16 kHz/16 bit by a male speaker of standard Thai. He was requested to read the sentences which were displayed to him on a computer screen fluently without pausing at the phrase break. The resulting speech data was annotated by means of forced alignment and labels checked. Table 2 displays the frequencies of 1-4 syllable words and the positions of the target syllables in these words. A *PRAAT* (http://www.praat.org) script was used to cut the target syllables from their contexts and weight them with a Hamming window in order to avoid discontinuities at the onsets of the stimuli.

Table 2*: Frequencies of words with 1-4 syllables used in the study (in total 203) and the positions of the target syllable (1-4).*

Number syllables	frequency	1	2	3	4
1	27	27			
2	66	30	36		
3	58	15	18	25	
4	41	8	17	10	6

In the ensuing perception test we intended to examine which of the following factors had an influence on the tonal realization:

1. the size of the word
2. the position of the syllable in the word, especially the distinction between word-final (primarily stressed) syllables and others
3. the position of the word in the sentence (medial or final)

We assumed that, provided more tonal reductions occurred, they should lead to a reduced rate of tone recognition once a syllable was excised from its context and presented in isolation.

3. The Perception Test

Experiments were conducted using the *DMDX* software (http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm) and employed scripts provided by Caroline Jones (MARCS, University of Western Sydney, Australia) that were slightly modified. In the experiment participants were requested to identify the presented stimulus by choosing one of four or five mono-syllabic words written in the Thai script.

Some of the syllables were chosen for a practice session preceding the experiment proper, and the following 400 stimuli in one trial series were divided into two groups of 200 split by a short break. Within each trial series tokens were presented in randomized order and each series took about 30-40 minutes to complete.

Participants listened to the stimuli over headphones connected to a PC soundcard. Each trial started with a preparation phase of one second during which the word 'ready' was displayed. Then the stimulus was presented, followed by the written forms of the possible choices. The word choices were arranged in ascending order by the numbering conventions of their associated Thai tones. Following the presentation of the syllables, participants made a forced choice by hitting the appropriate number key on the keyboard.

In the practice trials (using one syllable set), feedback concerning response accuracy was given, but in the main test not.

Participants were 20 students and members of staff of Chulalongkorn University (3 male, 17 female), Bangkok, aged 20-45. They reported to have normal hearing.

4. Results

An earlier study on the perception of Thai syllables uttered in isolation in various audio-only and audio plus video conditions)(Mixdorff et al., 2005) will serve as a benchmark for comparing the results of the current experiment. The pooled result of the perception experiment showed the picture displayed in Table 4. As can be seen, recognition rates drop considerably, especially for tones 0 and 4 which are often confused with tone 1. In Figure 2 we see *F0* contours calculated and plotted with *PRAAT* that pertain to syllables intended to carry tone 0 (top) and tone 4 (bottom), respectively. All contours have a similar, slowly falling characteristic which are commonly associated with tone 1 (with which these tokens were consistently confused). This indicates some kind of tonal neutralization occurring in the context of a word. In the case of tone 4, the rise part is delayed into the following syllable, accounting for the confusion with tone 1. These results are consistent with the findings reported in Luksaneeyanawin, (1995).

Table 3 displays the percentage correct and the corresponding confusion matrix on clean audio. As can be easily seen, the recognition rate is close to 100% and the largest number of confusions occurs between tones 0 and 1. Figure 1 shows prototypical contour of all tones taken from Abramson, 1979.

The pooled result of the perception experiment showed the picture displayed in Table 4. As can be seen, recognition rates drop considerably, especially for tones 0 and 4 which are often confused with tone 1. In Figure 2 we see *F0* contours calculated and plotted with *PRAAT* that pertain to syllables intended to carry tone 0 (top) and tone 4 (bottom), respectively. All contours have a similar, slowly falling characteristic which are commonly associated with tone 1 (with which these tokens were consistently confused). This indicates some kind of tonal neutralization occurring in the context of a word. In the case of tone 4, the rise part is delayed into the following syllable, accounting for the confusion with tone 1. These results are consistent with the findings reported in Luksaneeyanawin, (1995).

Table 3. *Proportion correct and confusion matrix on citation forms in percent taken from Mixdorff et al., 2005.*

Tone	0	1	2	3	4
percent correct	99.5	91.8	96.7	96.7	99.0

intended tone	perceived tone [%]				
	0	1	2	3	4
0	99.5	0.5	0.0	0.0	0.0
1	5.5	91.8	2.2	0.5	0.0
2	0.0	2.3	96.7	0.5	0.5
3	0.0	0.0	3.3	96.7	0.0
4	0.0	0.5	0.5	0.0	99.0

When we examine the relationship between correct responses and the length of the super-ordinate word, we find a highly significant negative correlation (ρ=-.33, $p<.01$) which is still highly significant (ρ=-.20) if we only include polysyllables in the analysis. The influence of word stress is also highly significant (ρ=.30 for all tokens, and .22 for polysyllables only). The associated mean recognition rates in polysyllables are 47.8% for stressed and 38.9% for unstressed syllables.

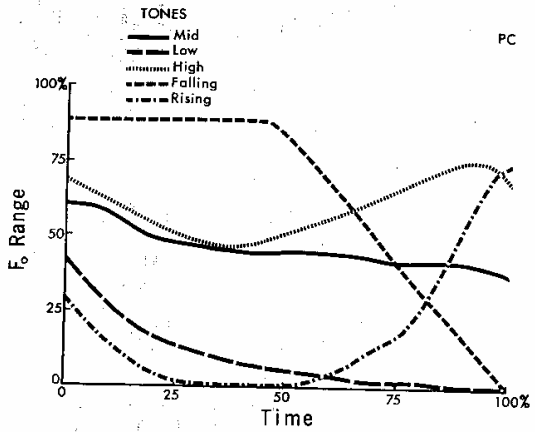

Figure 1. *Prototypical configurations of Thai tones in citation forms taken from Abramson, 1979.*

We subsequently checked whether the position in a word influenced the recognition rate beyond the stressed/unstressed distinction. Table 6 displays the results for two- to four-syllable words. The figures confirm that the tone on the stressed final syllable in a word is more reliably identified than on syllables in word-initial position. In four-syllable words this also concerns the penultimate syllable. Since tonal pattern and syllable duration are closely linked this relationship might also explain the other results. We therefore examined the syllable durations of the syllabic tokens used in the perception test depending on the syllable position (see Table 7).

Except for the penultimate syllable in four-syllable words the figures match those in Table 6. According to Luksaneeyanawin (1983) the possible stress patterns of disyllabic words are *strong-strong* or *weak-strong*, of tri-syllabic words *strong-weak-strong* or *weak-weak-strong* and of tetra-syllabic words either *weak-strong-weak-strong, strong-weak-weak-strong*, or *weak-weak-weak-strong*, depending also on the tempo. It seems therefore that the disyllabic words are mainly pronounced with *strong-strong pattern*, the tri-syllabic words with cratic rhythm (*strong-weak-strong*), and the tetra-syllabic words *strong-weak-weak-strong*, with the primary stress on the last syllables of the words.

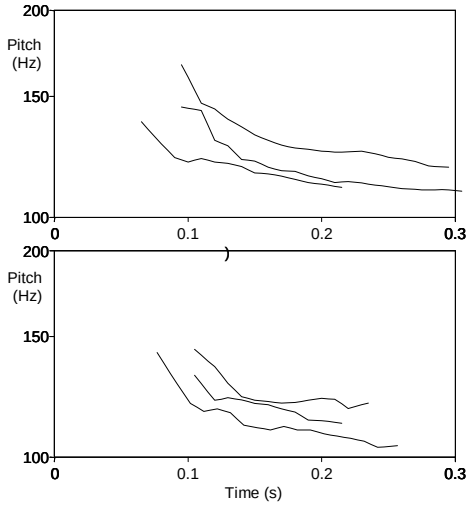

Figure 2. *Examples of F0 contours of stimuli that were wrongly classified as tone 1: Intended tone 0 (top), intended tone 4 (bottom).*

Table 4. *Proportion correct and confusion matrix, pooled result of current perception experiment.*

Tone	0	1	2	3	4
percent correct	41.4	70.4	46.2	63.9	40.6

intended tone	Perceived tone [%]				
	0	1	2	3	4
0	41.4	27.3	16.9	8.4	6.0
1	11.0	70.4	9.0	2.0	7.6
2	20.2	7.2	46.2	21.7	4.7
3	7.8	1.8	7.3	63.9	19.2
4	14.3	32.5	5.6	7.0	40.6

Table 5. *Proportion correct depending on the size of the super-ordinate word and number of stimuli N for each type of word (Σ=400).*

syllables in word	1	2	3	4
percent correct	62.2	46.8	41.3	36.9
N stimuli	54	139	119	88

We calculated correlations between the proportion correct and the syllabic duration of the tokens used in the perception experiment and found $\rho=.38$ ($p<.01$). This confirms that a main factor influencing the recognition rate besides the tone of the syllable and the length of the word is the syllabic duration. As could be expected the reaction time of the subjects is negatively correlated with the proportion correct ($\rho=-.19$, $p < .01$) and amounts to an average 3.0 s for tokens from monosyllabic words and 3.5 s for those from polysyllabic words.

Table 6. *Proportion correct depending on the position of the syllable in the super-ordinate word.*

syllables in word	position	Percent Correct
2	1	44.6
	2	48.6
3	1	37.6
	2	36.1
	3	47.4
4	1	36.6
	2	30.8
	3	43.1
	4	43.6

Table 7. *Mean syllabic duration depending on the position of the syllable, tokens used in the perception test.*

Syllables in word	position	Mean duration in ms
2	1	215
	2	241
3	1	208
	2	190
	3	255
4	1	205
	2	181
	3	195
	4	241

The position of the word in the sentence did not have any significant effect on the recognition rate. Phrase-medial and phrase-final tokens yielded a proportion correct of 45.9% and 44.2%, respectively.

5. Discussion and Conclusions

The current study investigated the relationship between the size of a word and the tonal realization. Syllabic tokens were excised from their original context and presented in a perception test. The outcome of the test showed considerably lower recognition rates than for citation forms, especially for tones 0, 2, and 4. The latter was frequently confused with tone 1. Monosyllables on the average were 20% more often correctly classified than syllables excised from polysyllables, and the number of syllables in the polysyllabic word had a highly significant influence on this rate. Furthermore the drop of the recognition rate also depends on the position of the syllable in the word which obviously also strongly influences the duration of the syllables. The stressed word-final syllables were therefore less often misclassified than others.

It must be noted that the database represented a compromise between the need to cover many different conditions and still employing meaningful words. As a consequence, we did not yield a perfect balance in the stimulus material regarding tone,size of word, and position in word. Furthermore, the reading style of the material might not exhibit as strong tonal reductions as could be found in spontaneous speech.

In the future we plan to perform a more balanced experiment by not limiting the stimulus set to certain syllables, but rather selecting the stimuli by the words to better cover all conditions of tonal combinations. An investigation of more spontaneous speaking styles also seems to be promising.

References

Abramson, A. S. "The coarticulation of tones: An acoustic study of Thai." In T. L. Thongkum, P. Kullavanijaya, V. Panupong, & K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer phonetics and phonology*: Indigenous Languages of Thailand Research Project, 1979.

Luksaneeyanawin, S.. Accentual system in Thai. In Intonation in Thai. Ph.D. dissertation, University of Edinburgh, 1983.

Luksaneeyanawin, S., "Tone transformation." Proceedings of the Second Symposium on Natural Language Processing. Kasetsart University, pp. 345-353, Thailand, 1995.

S. Luksaneeyanawin, “Intonation in Thai,” in Hirst, D. and Di Christo, A. (Ed.), Intonation Systems. A Survey of Twenty Languages. Cambridge University Press, Cambridge, 1998.

Mixdorff, H., Luksaneeyanawin, S., Fujisaki, H. and P. Charnvivit "Perception of Tone and Vowel Quantity in Thai." In Proceedings of ICSLP2002, Denver, USA, 2002.

Mixdorff, H. and Luksaneeyawin, S. et al. "Modeling Rhythmic Variation in Thai and its Application to Speech Synthesis." Proceedings of ICPhS2003, Barcelona, Spain, 2003.

Mixdorff, H., Charnvivit, P. and Burnham, D. "Auditory-Visual Perception of Syllabic Tones in Thai." In Proceedings of AVSP 2005, pp. 3 - 8, Parksville, Canada, 2005.