Production and Perception of Thai Lexical Tone and Intonation in Children
Chutamanee Onsuwan, Juthatip Duangmal, and Praneerat Panpraneet
Department of Linguistics, Faculty of Liberal Arts, Thammasat University
consuwan@tu.ac.th

Paper presented at the 12th International Conference on Thai Studies
22-24 April 2014
University of Sydney
ABSTRACT

Previous developmental studies suggested that in tonal languages, both lexical tone and intonation are mastered very early in children. Unlike intonation which generally signals speaker’s attitudes and emotions, lexical tone distinguishes words. Focusing on fundamental frequency ($F_0$) and pitch as main acoustic and perceptual cues, this work explores Thai children’s production and perception of five lexical tones (Experiments 1 and 2) and those of affective utterances (Experiments 3 and 4).

Experiment 1 investigates the production of five lexical tones by comparing $F_0$ measures and contours taken from recordings of 15 Thai children aged 2–3;11, 4–5;11 and 6–7;11. In Experiment 2, lexical tone identification of 75 Thai young children aged 2–3;11, 4–5;11 and 6–7;11 was compared using a 2-choice picture identification task. The findings suggested that at the age of 2–3, Thai children were able to produce and differentiate the five lexical tones, but they did not achieve mastery of these contrasts until the later age. In Experiment 3, Thai children’s production of affective utterances (happy–sad and interested–bored) was examined. $F_0$ measures were taken from recordings of 12 Thai children aged 5–7;11 and 8–10;11. Finally, in Experiment 4, sensitivity to the affective intonations in 40 children aged 5;0–7;11 and 8;0–10;11 was investigated with a 2-choice picture identification task. The results suggested that by the age of 5 Thai children’s ability to produce and their sensitivity to affective intonations were already developed and mastered.

Index Terms: Thai children, lexical tone, affective intonation, fundamental frequency
Production and Perception of Thai Lexical Tone and Intonation in Children

INTRODUCTION

Background and Previous Studies
Lexical tone and intonation are different pitch patterns in speech, generally correlated with speaker’s fundamental frequency (F0). While lexical tone distinguishes words, intonation could be associated with the domain larger than a single word (i.e., phrases and sentences) and signals grammatical meaning (e.g., statement, question), speaker’s attitudes and emotions (e.g., interested, bored) (Abercrombie 1967). All spoken languages have intonation and around 70% of those have lexical tone (Ladd 1992; Yip 2007).

Thai is a tonal language. Every word has distinctive pitch pattern which denotes a change in lexical meaning. There are five tones: three so-called ‘level/static’ tones, “mid” (ˉ), “low” (ˋ), “high” (ˊ) and two so-called ‘contour/dynamic’ tones, “falling” (ˆ), and “rising” (ˇ) (Abramson 1962). Like tones in other tonal languages, whether spoken in isolation or in connected speech, F0 height and shape have been found to be major cues to differentiating the five tones. Production and perception of Thai tones have been extensively studied (e.g., Abramson 1962; Potisuk, Gandour & Harper 1994), but only a few studies discuss tonal development in Thai children. Tuaycharoen (1977) conducted a longitudinal study where she observed acquisition course of lexical tones in a Thai child. She stated that tonal acquisition began very early and completed at around 23 months. But to date, not much has been known about Thai children’s perception and production of lexical tones, especially from different age groups.

Unlike Thai lexical tones, Thai intonation patterns have received much less attention. This may be due to the fact that naturally produced utterances with different intonation patterns cannot easily elicited in a laboratory setting. Luksaneeyanawin (1983) explored various intonation patterns of Thai adults’ one-word utterances. To the best of our knowledge, no extensive study has been done on the acquisition of different intonation patterns in Thai children.

Present Study
To fill the gap mentioned above, the goal of this work is to investigate Thai children’s production and perception of five lexical tones (Experiments 1 and 2) (children aged 2–3;11, 4–5;11 and 6–7;11) and those of affective utterances (Experiments 3 and 4) (children aged 5–7;11 and 8–10;11). It is our hope that this work will be among the first studies which offer some insights into Thai children’s production and perception of lexical tones and affective intonation patterns.

PRODUCTION OF LEXICAL TONES: EXPERIMENT 1

Experimental Setup
A picture-naming test was conducted on 15 Thai children (girls only) from 3 age groups: 2;0–3;1, 4;0–5;11, and 6;0–7;11. A baseline group consists of 3 Thai adult females (born and grew up in and around Bangkok). Each test session lasted about 5–7 minutes.
There are two sets of pictures. Both sets consist of 10 identical target words and 5 filler words, but the pictures are in different order. The target words include ปู [pū:] ยา [jà:] (mid tone); กล่อง [kłòŋ] เข่า [kʰàw] (low tone); กล้อง [kłɔŋ] เสื้อ [sùa] (falling tone); ม้า [má:] วัด [wát] (high tone); and เสือ [sùa] ม้า [má:] (rising tone). All words were chosen to meet the requirement that they are commonly used among the children in these age groups.

During the recording procedure, each participant was shown a picture from the two sets one by one and was asked to say the corresponding word. Each word was repeated 3 times. The best two tokens for each word were selected and evaluated (and time-normalized) using Praat speech analysis software. Together, 360 speech tokens were analyzed (10 target words × 2 repetitions × 18 participants).

Experimental Results
Fundamental frequency (F0) (in Hertz) values taken from the speech tokens are shown here in Figures 1–5. Each F0 curve is an average across all subjects in the same age group. It should be noted that the baseline F0 curves are similar to those reported in the previous studies of Thai tones (e.g., Potisuk, Gandour & Harper 1994). The results show that Thai children’s production of each tone generally follows the same pattern as that of the adults (baseline). There seems to be minor differences among the production of children from three different age groups. In fact, across the three age groups, F0 curves display the most similar patterns in mid and low tones.
Figures 1–5. Fundamental frequency (in Hertz) of each lexical tone (from left to right; top to bottom; mid, low, falling, high, and rising) taken from the words produced by 4 subject groups.

Moreover, some similarities and differences among tone production of children and adults can be observed in their fundamental frequency range in Figure 6. Tones that show the widest $F_0$ range for the adults are falling and rising; and for the children rising and falling. The narrowest range is found in mid tone for all age groups.

The findings confirm that Thai two-year-olds are able to produce five tonal contrasts (Tuaycharoen 1977) and their tone production is comparable to the adults’. Among the five tones, mid and low tones (level/ static tone) seem to be mastered sooner than others (Tuaycharoen 1977).

PERCEPTION OF LEXICAL TONES: EXPERIMENT 2

Experimental Setup
A picture-identification test was conducted on 75 Thai children (boys and girls) from 3 age groups: 2;0–3;1, 4;0–5;11 and 6;0–7;11. A baseline group consists of 5 Thai adults, aged 20–50 years.
The speech materials consist of 30 monosyllabic words; all are basic words in the language. Altogether, there are 10 target tone pairs (20 words) in which members differ only in lexical tone. Five pairs of filler words (10 words) are also included.

<table>
<thead>
<tr>
<th>Tone pair</th>
<th>Target word</th>
</tr>
</thead>
<tbody>
<tr>
<td>mid–low</td>
<td>1) ปู ‘crab’ [pû:] – ปู ‘paternal grandfather’ [pû:]</td>
</tr>
<tr>
<td>mid–falling</td>
<td>2) ยา ‘medicine’ [jâ:] – หญ้า ‘grass’ [jà:]</td>
</tr>
<tr>
<td>mid–high</td>
<td>3) นา ‘rice filed’ [nâ:] – น้า ‘aunt’ [nâ:]</td>
</tr>
<tr>
<td>mid–rising</td>
<td>4) ไฟ ‘fire’ [fâj] – ไม ‘mole’ [fâj]</td>
</tr>
<tr>
<td>low–falling</td>
<td>5) กล่อง ‘box’ [klâŋ] – กล้อง ‘camera’ [klâŋ]</td>
</tr>
<tr>
<td>low–high</td>
<td>6) วัด ‘cold’ [wàt] – วัด ‘monastery’ [wàt]</td>
</tr>
<tr>
<td>low–rising</td>
<td>7) เข้า ‘knee’ [kæaw] – เขา ‘mountain’ [kæaw]</td>
</tr>
<tr>
<td>falling–high</td>
<td>8) เต้น ‘dance’ [tên] – เต็นท์ ‘tent’ [tên]</td>
</tr>
<tr>
<td>falling–rising</td>
<td>9) เสืือ ‘shirt’ [stûa] – เสือ ‘tiger’ [stûa]</td>
</tr>
<tr>
<td>high–rising</td>
<td>10) หมา ‘horse’ [mâ:] – หมา ‘dog’ [mâ:]</td>
</tr>
</tbody>
</table>

Table 1. Ten target tone pairs (20 words); members of each pair differ only in their lexical tone.

The word stimuli were selected from recordings of a 31-year-old native female, born and raised in Bangkok. All words were spoken in isolation.

During the test session, a 2-choice identification task was carried out using pictures. Participants heard five simple words as practice trials and then 30 words (20 target words and 10 filler words) one at a time over headphones. In each trial (one word presented), they were asked to identify which word they had heard by pointing at one of the 2-picture choices. The test was performed individually on each participant about 10 minutes in a quiet room. Together, 1600 responses (20 target words × 80 participants) were collected. Percent correct responses were tallied and confusion matrices were constructed, following Onsuwan et al. (2012).

**Experimental Results**

The results (Figure 7) show that among the children, the oldest group performs best at 98%, followed by the middle group, and the youngest group. Importantly, the level of differences between the 3 groups is statistically significant \[F(2, 72) = 84.78, \ p < 0.05\].
Confusion matrices shown in Tables 2–4 present percent correct responses (boldfaced; diagonal wise) as well as misidentified responses. For example, in the topmost table, the 2–3 year-olds identified mid tone stimuli as mid tone (correct response) 72 percent of the time; they identified mid tone as low tone (misidentification) 4 percent.

<table>
<thead>
<tr>
<th>2–3 yrs</th>
<th>Mid</th>
<th>Low</th>
<th>Falling</th>
<th>High</th>
<th>Rising</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid</td>
<td>72</td>
<td>4</td>
<td>10</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Low</td>
<td>20</td>
<td>51</td>
<td>7</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>Falling</td>
<td>5</td>
<td>3</td>
<td>87</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>High</td>
<td>9</td>
<td>4</td>
<td>7</td>
<td>77</td>
<td>3</td>
</tr>
<tr>
<td>Rising</td>
<td>16</td>
<td>14</td>
<td>1</td>
<td>7</td>
<td>62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4–5 yrs</th>
<th>Mid</th>
<th>Low</th>
<th>Falling</th>
<th>High</th>
<th>Rising</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid</td>
<td>97</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>10</td>
<td>81</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Falling</td>
<td>0</td>
<td>1</td>
<td>96</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>High</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>93</td>
<td>1</td>
</tr>
<tr>
<td>Rising</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>87</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6–7 yrs</th>
<th>Mid</th>
<th>Low</th>
<th>Falling</th>
<th>High</th>
<th>Rising</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid</td>
<td>99</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>3</td>
<td>96</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Falling</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>High</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Rising</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>95</td>
</tr>
</tbody>
</table>

Tables 2–4. Confusion matrices (in percent) of children identification performance; aged 2–3 years old (top) 4–5 years old (middle) and 6–7 years old (bottom).

In general, low tone elicited a large number of confusion responses and tends to be misidentified as mid tone. This may be due to the fact that mid and low tones are realized as level tone with relatively similar F0 curves (Potisuk, Gandour & Harper 1994). However, they seldom confused rising tone with falling tone (and vice versa). The two tones possess very different contour F0 curves and should be difficult to be confused with one another.
The findings show that age difference plays an important role in tonal perception of Thai children. The ability to accurately identify the five lexical tones seems to gradually develop with age.

PRODUCTION OF AFFECTIVE INTONATION: EXPERIMENT 3

Experimental Setup
There are three subject groups in this study. Twelve Thai children (girls only) aged 5;0–7;11, and 8;0–10;11; six children in each group. A baseline group consists of three Thai adult females (aged 20–24); they are all undergraduate students majoring in Theatre and Communication Arts.

Of interest here are affective intonation patterns found in two pairs of affective polar opposites: happy–sad and interested–bored. For each pair, a target utterance was created: “ไม่รู้หรอก” ‘(I) don’t even know’ (for the happy–sad pair) and “แน่ๆเลย” ‘(that is) certainly so’ (for the interested–bored pair). Each utterance has three words and each word begins with a voiced consonant. To elicit affective utterances, eight different made-up mini-stories were role-played (using finger puppets) by the researcher and the child. At the end of each mini-story, a target utterance was successfully uttered and recorded.

Together, 120 speech tokens (2 target utterances × 4 affective types × 15 participants) were evaluated (and time-normalized) using PRAAT speech analysis software.

Experimental Results
Each $F_0$ curve in Figures 8–11 is an average of $F_0$ measures from the affective utterances across all subjects in the same age group.

Figures 8–9. Fundamental frequency (in Hertz) of two types of affective utterance (happy (left) and sad (right)) produced by 3 subject groups.

From Figures 8–9, it could be seen that in general, the sad intonation has a smoother and flatter pattern (with an exception for the 5–7 years group) than the happy intonation. In fact, $F_0$ range for the happy intonation seems to be wider than that of the sad intonation (with an exception for the 5–7 years group in the sad intonation). In general, $F_0$ curves of children’s utterances are quite similar to those of the adults; this is even more so in the case of the happy intonation.
Figures 10–11. Fundamental frequency (in Hertz) of two types of affective utterance (interested (left) and bored (right)) produced by 3 subject groups.

Similar to the happy–sad pair, Figures 10–11 show that the bored intonation has a smoother and flatter pattern than the interested intonation. Likewise, F0 range for the interested intonation is wider than that of the interested intonation. In both figures, F0 curves of children’s utterances are very similar to those of the adults.

In all, the results showed that affective intonation patterns of the Thai children follow those of the adults. The findings are not surprising as many Psycholinguistics research have shown that babies as young as 8 months started to produce clear intonation and to imitate intonation from their mother (e.g., Aitchison 2008).

PERCEPTION OF AFFECTIVE INTONATION: EXPERIMENT 4

Experimental Setup
A picture-identification test was conducted on 40 Thai children (boys and girls) from two age groups: 5;0–7;11 and 8;0–10;11 and on a baseline group consisting of three Thai adult females. There were two sets of 2-picture choices presented; each represents different facial expression and emotion: happy versus sad (Set 1) and interested versus bored (Set 2).

The speech materials consist of eight target affective utterances (two utterances for each polar affect (i.e., happy, sad, interested, and bored) and two filler utterances. The target utterances were selected from recorded materials of a Thai female adult (25 years old; born and raised in Bangkok) based on acoustic examination and impressionistic evaluation of three phonetic features, F0, intensity, and utterance duration, to ensure that they match well with the affective intonation patterns they are conveying.

The test session was conducted over the headphones. Participants listened to two different test sets (Set 1: happy versus sad, and Set 2: interested versus bored); each with six trials (four target utterances and two filler utterances). In each trial, an utterance was repeated three times in a row with an interval of 0.58 second. Each test set was also repeated twice. The participants were asked to point at one of the two facial expression pictures that matches with each utterance. The test was performed individually on each participant in a quiet room.

Together, 688 responses (8 target utterances × 2 repetitions × 43 participants) were collected.

Experimental Results
Figure 12 shows participants’ percent correct responses to the target utterances. It can be seen that the children had high ability to differentiate affective intonation patterns;
an average score of the two groups across two test sets was around 91%. Performance differences between the two groups of children were not statistically significant.

![Graph](image)

Figure 12. Percent correct responses of affective intonation identification across 3 age groups.

SUMMARY AND DISCUSSIONS

Taken together, the four experiments clearly show that the acquisition of pitch patterns (lexical tone and intonation) developed very early in Thai children even though the exact time course of development of the two phonetic features could be somewhat different. Due to difficulty in eliciting affective intonation utterances in young children (2–4 years old), this study could not make a parallel comparison of the production of tone and intonation from the Thai children of same age groups.

For lexical tone, the findings suggested that at the age of 2–3, Thai children were able to produce and differentiate the five tonal contrasts, but they did not achieve mastery of these contrasts until the later age (up to 5–7 years old). Age difference played an important role in tonal perception of Thai children and the ability to accurately identify 5 lexical tones seems to develop with age. Interestingly, it was found that Cantonese children could not perceive lexical tones in an adult-like level until they reached about the age of 10 (Ching 1984).

When it comes to affective intonation, the findings suggested that by the age of 5 Thai children’s ability to produce and their sensitivity to affective utterances were already developed and mastered. Age difference did not seem to influence the ability to differentiate intonation patterns. This finding is in agreement with those reported in many Psycholinguistics research on development of children intonation (e.g., Aitchison 2008; Crystal 1986) which stated that infants as early as 8–9 months, often started to produce clear intonation and to imitate intonation from their caretaker. Further studies on intonation patterns associated with grammatical meaning (e.g., statement and question) should provide an important aspect for the acquisition of intonation in Thai children.

REFERENCES


