Tone Features, Tone Perception, and Peak Alignment in Thai

Elizabeth Zsiga and Rattima Nitisaroj

Department of Linguistics

Georgetown University, Washington, D.C.

Suggested running head: Peak Alignment in Thai

Address for correspondence:

Elizabeth Zsiga

Dept. of Linguistics

Box 571051

Georgetown University

Washington, DC 20057-1051

Acknowledgements: Special thanks are due to Ken DeJong, Dan Loehr, Alice Turk, and Draga Zec, for advice that greatly improved this manuscript.

Abstract

This paper investigates the relationship between the phonological features of tone and tone perception in Thai. Specifically, it tests the hypothesis (proposed by Morén & Zsiga 2006) that the principle perceptual cues to the five-way tonal contrast in Thai are high and low pitch targets aligned to moras. Results of four perception studies, one using natural speech and three using digitally-altered speech, are presented in support of the hypothesis. It is argued that, by associating tones to moras, a straightforward mapping from the abstract autosegmental features H and L to the production and perception of Thai tones, a heretofore elusive goal, can be accomplished. This result has consequences for theories of contour tone perception, the distinctive features of tone, and the alignment of pitch targets to the segmental string.

Key Words: Thai, tone, tone perception, tone features

BACKGROUND: TONE FEATURES AND TONE PERCEPTION

Representations for Thai Tones

Phonologists generally agree that contour tones should be represented as a sequence of H and L autosegments associated to a tone-bearing unit (TBU), which might be a vowel, syllable, or mora (e.g., Leben 1973, Gandour 1974a, Anderson 1978, Yip 1989, 1995, 2002, Duanmu 1994, Zhang 2002, Gussenhoven 2004). Particularly for Asian languages, however, mapping these H and L autosegments into actual F0 contours and their perceptual correlates has proved problematic. Experimental studies of contour tone production and perception have generally either argued against a compositional analysis of contour tones (e.g., Abramson 1975, 1978; Gandour 1978, 1983; Gandour & Harshman 1978; Xu 1998, 2004) or have not addressed the question of abstract phonological representation at all (e.g., Shen & Lin 1991, Gandour et al. 2000, Wayland & Guion 2003, Liu & Samuel 2004). In this paper, we present perceptual evidence in support of a mora-based autosegmental representation for Thai tones, a representation that we argue provides a straightforward mapping from phonological representation to pitch contour.

Thai provides an interesting test case concerning the relationship between tone features and tone perception, because its five-way tonal contrast, which has been described and transcribed as high, low, mid, falling, and rising, seems a perfect candidate for autosegmental representation, but its phonetic contours do not correspond at all neatly to this representation. The five-way tonal contrast of Thai is illustrated by the minimal sets in Table 1.

TABLE 1 ABOUT HERE

This set of contrasts lends itself easily to a representation in terms of H and L autosegments (Leben 1971, 1973, Gandour 1974a) as shown by the representations in (1).

(1) Autosegmental associations

<u>mid</u>	<u>high</u>	<u>low</u>	<u>falling</u>	<u>rising</u>
	H	L	H L	L H
V	V	V	V	V

No consensus has been reached, however, on the identity of the TBU in Thai. The earliest autosegmental works on Thai (Leben 1971, 1973, Gandour 1974a) associate tones to vowels, as shown in (1). Later phonological studies, including Yip (1982) and Zhang (2002), assume that Thai tones are properties of syllables. Yip (2002) suggests either the syllable or mora as plausible candidates for the Thai TBU. Morén and Zsiga (2006) argue for the representation in (2), a proposal that we will term the "moraic alignment hypothesis."

(2) Moraic alignment hypothesis

a. Thai tones in phrase-final position (including citation forms)

<u>Mid</u>	<u>High</u>	<u>Low</u>	<u>Falling</u>	<u>Rising</u>
	H	L	H L	L H
μ μ	μ μ	μ μ	μ μ	μ μ

b. Thai tones in non-phrase-final position

Mid	High	Low	Falling	Rising
	H	L	H	L (H)
μ μ	μ μ	μ μ	μ μ	μ μ

Mid, high, and low tones have the same representation in both final and non-final position: mid tones have no phonological tone specification, high and low tones have a single tonal autosegment associated to the rightmost mora of the syllable. The falling and rising tones may differ by phrasal position. In final position, including citation forms, falling and rising tones have autosegments associated to both moras, but in non-final position the second specification may be deleted. Falling tones are consistently simplified in non-final position; rising tones are simplified only for some speakers. Morén and Zsiga (2006) present evidence for the representations in (2) based on lexical phonological distributions, and on the acoustics of pitch contours in citation form and connected speech. The present study further tests whether the moraic alignment hypothesis is consistent with the perceptual cues to tone identity that are in fact utilized by Thai speakers.[1] Specifically, do Thai speakers rely on high and low pitch targets, aligned with moraic structure, to encode the tonal contrasts of their language?

The rest of the paper is organized as follows. The rest of this introductory section reviews previous work concerning 1) the role of the mora in Thai phonology, and the phonological plausibility of the mora as the TBU, 2) the acoustics of Thai pitch contours and the mapping from autosegments to pitch targets, in both citation form and connected speech, 3) tone perception. The next sections present the experimental data, from four new perceptual experiments. Experiment 1 uses natural speech, with tokens both in isolation and in a sentence

[1] While we argue that moraic alignment works best for Thai, we make no claim that moraic alignment is universal, but follow Yip (2002) in assuming that tones may associate to different prosodic units in different languages.

context. Experiments 2, 3, and 4 use digitally-altered speech. Experiment 2 tests responses to different pitch contours in citation form; Experiment 3 tests responses to different pitch contours in sentence contexts; Experiment 4 examines the results of shifting peak alignment. Finally, the paper concludes with a general discussion, supporting the hypothesis that association of tonal autosegments to moras in Thai best accounts for not only for the phonological and acoustic, but also the perceptual data.

The Mora in Thai Phonology

The reader is referred to Morén and Zsiga (2006) for a detailed phonological analysis of Thai syllable structure and tone distributions. A few central points regarding the role of the mora in Thai phonology are reviewed here. All stressed syllables in Thai are bimoraic (Bennett 1994). Most Thai words are in fact monosyllabic; compounds and polysyllabic words (mostly borrowings) show final stress. The bimoraic minimum is evidenced in that the rhymes of monosyllabic words and stressed final syllables may consist of a long vowel ([na:], “rice field”) or a short vowel plus coda ([laŋ], “crate”, [làk] “stake”), but may not consist of a short vowel with no coda (*[na]). The patterning of CVC with CV: indicates that coda consonants are moraic. Long vowels also occur in closed syllables ([la:ŋ] “omen”, [là:k] “various”), in which case the vowel and final consonant share the second mora (Bennett 1994, consistent with Broselow et al. 1997). Monomoraic CV syllables occur only in unstressed (non-final) position in polysyllabic words and compounds, such as [sàˈrǐ:] “sari” or [p^{h}ráˈc^{h}âu] “deity,” lit. “ruler of monks.” These light CV syllables are augmented with a final glottal stop, adding a second mora, if they occur in isolation: [p^{h}ráʔ] “monk” (Gandour 1974b).

Thus, in stressed syllables, two moras are always available to carry the contrasts shown in (2). Special distributional constraints hold of obstruent-final syllables, which Morén and Zsiga (2006) analyze in terms of constraints on tone-mora association. It has been argued (Yip 1995, Odden 1995) that representations in which the mora is the TBU overgenerate possible contrasts, because they allow too many possible tone-to-TBU mappings. However, Morén and Zsiga argue for a non-representational solution to the overgeneration problem, proposing a set of alignment constraints that allow all and only the contrasts in (2) to surface.

Finally, in unstressed monomoraic syllables, the possible tonal contrasts are reduced from five to two or three. Sources differ as to whether the distinction between H and M is neutralized in unstressed syllables (see Hiranburana (1971), Leben (1973), Luksaneeyanawin (1983), Potisuk et al. (1994, 1996), Gandour et al. (1999), and Nitisaroj (2006) for extensive discussion). All sources agree, however, that falling and rising tones do not occur on unstressed CV syllables, supporting a view in which two moras are necessary for the realization of complex tones.

Pitch Contours in Citation Forms

As noted above, at least since Abramson (1962) the tonal contrasts of Thai have been labeled as high, low, mid, falling, and rising. However, the actual phonetic shapes of the individual tones, even in citation form (as documented, for example, by Abramson 1962, Erickson 1974, Gandour et al. 1991), do not match these phonological labels. Example pitch contours of the Thai tones in citation form are shown in Figure 1. The tokens are all the syllable /na:/[2], with meanings for each tone as indicated in Table 1, spoken by a 28-year-old female speaker of the Bangkok dialect. (Tokens were digitized and analyzed in Praat (Boersma &

[2] Notation with phonemic brackets (e.g., /na:/) is used when the segmental string /n/ /a:/ is intended, regardless of tonal pattern. Notation with phonetic brackets (e.g., [na:]) is used when the mid-toned syllable is intended.

Weenik 2003), using an autocorrelation algorithm with a 20 ms analysis window. Pitch tracks begin at vowel onset.) As is clear from the figure, none of the tones are actually level. The mid tone comes closest, falling about 20 Hz over the course of the syllable but remaining in the middle of the pitch range. The high tone is a scooped contour: falling slightly and remaining as low or lower than the mid tone for the first half of the syllable, then rising steeply in the second half. The low tone falls steadily, reaching the bottom of the pitch range at the end of the syllable. The falling tone is realized as a rise-fall contour, and the rising tone as a fall-rise contour.

FIGURE 1 ABOUT HERE

Creating these complex contours from simple H and L associated to the syllable is certainly possible. Such phonetic mapping rules would be complex, however. A single H associated to the syllable would have to be mapped into a level-rising scooped contour, while an H linked as part of a falling tone would correspond to a quick rise to the top of the pitch range. The complexity of attempting to map H and L autosegments onto the actual contours of the tones of Thai leads Abramson (1979) to reject a compositional analysis of the contours. He argues (p. 7) that the data "lend no phonetic plausibility" to arguments for the specification of rising and falling tones as sequences of H and L autosegments. "For phonetic support of the argument one would expect to be able to devise a formula by which the dynamic tones were obviously to be derived from the shape of the static tones. Even the citation forms, let alone the F0 curves of running speech, provide no acoustic basis for such a claim. It seems psychologically far more reasonable to suppose that the speaker of Thai stores a suitable tonal shape as part of his internal representation of each monosyllabic lexical item" (p.7).

Morén and Zsiga argue, however, that a mora-based representation as in (2) does allow a straightforward mapping from autosegments to the acoustics of Thai tones. During moras with no phonological tone, pitch falls gradually to or within the mid range. Phonologically-specified moras reach a high or low pitch inflection at their right edge. Thus mid tones, with no phonological specification, remain fairly level throughout their duration. High tones are specified only on the second mora, and thus remain in the mid range during the first half of the syllable, begin to turn upward at syllable midpoint, and reach their high point at the rightmost edge of the second mora. Low tones are also specified only on the second mora, and they also reach a pitch extremum only at the right edge. Falling and rising tones show a pitch inflection at or near the syllable midpoint (Gandour et al. 1999, Potisuk et al. 1997): falling tones are high at syllable midpoint and low at syllable endpoint, while rising tones show the reverse pattern. [3]

Pitch extrema are realized at the right edge of the specified mora: high and low tones reach their targets at the right edge of the second mora, which corresponds to the right edge of the syllable. Falling and rising tones reach pitch extrema at the right edge of the first mora (middle of the syllable) and right edge of the second mora (right edge of syllable). Morén and Zsiga encode this alignment as part of the Thai-specific phonology-phonetics mapping, treating the mora as the smallest unit in the prosodic hierarchy (Cohn 2003), and proposing that phonological tones are aligned with the right edge of the mora as part of their phonetic realization, parallel to alignment algorithms that have been proposed for larger prosodic units including syllables and phrases (see Gussenhoven 2004 and refs therein.) This rightward alignment is also consistent, however, with a universal trend toward late realization of phonetic

[3] These specifications are consistent with Erickson (1976, 1994), whose EMG studies showed that, in Thai, contraction of the crico-thyroid was found for any pitch raising, and contraction of the strap muscles was found for pitch lowering to the bottom of the pitch range, but that in the absence of any raising or lowering gesture pitch fell gradually within the mid range. See Morén and Zsiga (2006) for further discussion of tonal underspecification and contextual variation.

pitch targets within a specified domain (Xu 1999a, 2004). Crucial to the present argument, however, is the point that moras act as separate domains for pitch alignment. This is particularly clear in the case of the high tone, which behaves exactly like an unspecified mid tone during the first half of the syllable, then changes direction in the second half and reaches a high point only at the right edge.

Pitch Contours in Connected Speech

Further complicating the issue of Thai tonal representation and perception is the fact that some of the tones of Thai, particularly the contour tones, change shape from citation form to connected speech, even in stressed syllables (Abramson 1979, Gandour et al. 1994, Potisuk et al. 1997, Kallayanamit 2004, Morén & Zsiga 2006, Nitisaroj 2006). This difference was alluded to in the quote from Abramson (1979) above. While all previous studies of Thai connected speech report some simplifications, the greatest changes have been reported in recent studies examining the speech of younger female speakers from Bangkok (i.e., Potisuk et al. 1997, Kallayanamit 2004, Morén & Zsiga 2006, Nitisaroj 2006). These studies demonstrate that in non-final position in Thai connected speech, the pitch excursion in the second half of contour tones is curtailed. Examples from Morén and Zsiga (2006) are shown in Figure 2. (The speaker on the right is the same speaker as in Figure 1.) Examples are all syllables of the form CVN (consonant followed by a short vowel followed by a nasal). Tokens were analyzed in the same way as those in Figure 1.

FIGURE 2 ABOUT HERE

Comparison of Figures 1 and 2 shows that the high, mid, and low tones have essentially the same shape in both citation form and connected speech, though they are shorter in duration in connected speech. Differences are seen, however, in the contour tones. Figure 1 illustrated that in citation form or in phrase-final position, the falling tone is realized with a rise-fall contour. In non-final position in connected speech (Figure 2) there is no actual fall: the pitch rises and remains high. There is also variation in the realization of rising tones by different individuals. As shown in Figure 2, the non-final rising tone has a fall-rise contour for the speaker shown on the left, but for the speaker shown on the right the same token in the same environment is realized with low-falling pitch for most of the syllable duration, with a very slight upturn only at the end. The acoustic studies of connected speech cited above note that although contour tones are simplified, the five-way contrast is maintained in stressed syllables, and no tonal distinctions are neutralized. This claim is tested perceptually in Experiment 1 below. (As noted above, whether or not tonal contrasts are neutralized in unstressed syllables remains a debated question which is not addressed here.)

Morén and Zsiga (2006) analyze contour tone simplification as shown in (2b): in non-final position in connected speech, the second tone of a contour (the one associated to the weaker mora), may be deleted. The L of the falling tone is deleted for both speakers; the H of the rising tone is deleted only for the first speaker. As shown in (2b), moraic association allows the five-way contrast to be maintained even with simplification: high and falling tones are distinguished in that the high tone has an H autosegment associated to the second mora, and the "falling" tone an H autosegment associated to the first mora.[4]

A phonological tone simplification without neutralization can not be captured by an autosegmental representation in which contour tones are represented as H and L autosegments associated to the syllable or vowel, as in (1) above. If simplification of the falling tone, for example, is modeled as deletion of the L, then the distinction between falling and high would be phonologically neutralized, and Figure 2 shows that this is not the case. Nor can connected speech simplifications be easily modeled as phonetic reduction. If only phonetics is involved, one would have to assume a reduction that selectively targets only the offsets of the contour tones, not other high and low points in the pitch trajectories, and that reduces just those points to such an extent that falling tones end at the top of the pitch range rather than at the bottom. One would have to argue for a phonetic implementation in which the phonological L of the rising tone causes a pitch drop to the bottom of the range, while the phonological L of the falling tone has no phonetic realization at all. Potisuk et al. (1997) also rule out an analysis involving tone delay, in which the L of the falling tone might be realized on a subsequent syllable, by demonstrating that syllables following falling tones in connected speech are not significantly different from syllables following a mid tone.

Thus, Morén and Zsiga (2006) argue for the representations in (2) based on phonological tone distributions, and the acoustics of pitch trajectories in citation form and connected speech. It is argued that alignment of tones to moras accounts for gaps in the distribution of tones on obstruent-final syllables, and allows for a straightforward mapping from autosegmental representation to pitch contour in both citation form and connected speech. Missing from that

[4] The lack of an actual fall in pitch on "falling" tones in connected speech indicates that this linguistic label is a misnomer; it is not used by Thai speakers themselves, who label the tones with terms that correspond to military ranks.

study, however, was any perceptual data. It is to the issue of tone perception that the next section now turns.

Perceptual Studies

Slope vs. endpoints. The moraic alignment hypothesis makes specific predictions about the perceptual cues to Thai tones, as shown in (3):

(3) Perceptual cues predicted by moraic alignment

a. simple tones in both final and non-final position:

high tone: high at syllable endpoint (H at right edge of second mora)

low tone: low at syllable endpoint (L at right edge of second mora)

mid tone: absence of any high or low points (no autosegmental specification)

b. complex tones in final position:

falling tone: high at syllable midpoint, low at syllable endpoint

(H at right edge of first mora, L at right edge of second mora)

rising tone: low at syllable midpoint, high at syllable endpoint

(L at right edge of first mora, H at right edge of second mora)

c. complex tones in non-final position:

falling tone: high at syllable midpoint (H at right edge of first mora)

rising tone: low at syllable midpoint (L at right edge of first mora)

In predicting the importance of high and low pitch targets aligned to specific sub-syllabic constituents, the moraic alignment hypothesis differs from previous studies of contour tone perception. Most perceptual studies of Thai tones, and of contour tones in general, have emphasized the importance of pitch movement rather than endpoints in tone identification and discrimination. (One counterexample is Mixdorff et al. (2002), discussed below.) Pike (1948) suggested that in contour tone languages such as Thai and Chinese, the unit of contrast (the "toneme") was the upward or downward pitch movement itself, not any particular high or low point. Other more recent perceptual studies have supported this generalization.

Gandour and colleagues, in a series of experiments (Gandour 1978, 1981, 1983, 1984; Gandour & Harshman 1978) manipulated aspects of the pitch contours of syllables in citation form, and obtained judgments of the similarity between contours from speakers of various languages, both tonal and nontonal, including Thai, English, Chinese, and Yoruba. Gandour concludes that both pitch height and direction of pitch movement are crucial perceptual cues. Gandour (1983) reports that pitch direction is particularly salient for Thai speakers, whereas "extreme endpoint", which might seem to correlate most directly to H and L autosegments, is claimed to be salient only for non-linguistic pitch perception. Gandour (1978) explicitly argues that if it is a goal of phonology to "develop an empirically grounded set of phonetic features" (p. 42), then phonological tone representations must use perceptual features including direction and slope of pitch change. Such perceptual features have been successfully adopted by other researchers on tone perception, including Massaro et al. (1982), Lin and Repp (1989), House (1990), Gandour et al. (2000), Wayland and Guion (2003), and Xu (1998, 2004), among many others.

The importance of contour slope for tonal identification in Thai is supported by the data reported in Abramson (1978). Using the syllable [k^ha:] as a base, Abramson synthesized different straight-line pitch trajectories over the course of the syllable, and asked 37 Thai listeners to identify the tone of the synthesized syllable. (Recall from Table 1 above that [k^ha:] is a meaningful lexical item when used with all five of the contrastive tones of Thai.) Level, gradually sloping, and sharply sloping trajectories were included. While level trajectories were usually identified as high, mid, or low tones, adding some gradual movement to the trajectory over the course of the syllable made the "static" tones less confusable and more acceptable to the listeners. The trajectory most reliably identified as a high tone began at the middle of the pitch range and rose 30 Hz over the course of the syllable; the trajectory most reliably identified as low began in the middle of the range and fell 30 Hz over the course of the syllable. Abramson concludes that the tonal contrasts of Thai are defined in terms of slope and direction of pitch change: mid tones are flat, high and low tones are associated with gradual upward or downward movement, while rising tones (and by extension, falling tones, though these were not overtly tested in these experiments) require "abrupt" movement.

Thus, there has been a divide between researchers on contour tone perception (as cited above), who have assumed or argued for contour tones as "single dynamic elements" (Xu 2004:15), and researchers on contour tone phonology (e.g., Yip 2002 and others cited above), who have assumed or argued for contour tones as sequences of H and L autosegments. Anderson (1978:154), countering Gandour's proposal for perceptual features for tone, goes so far as to argue that perception results may be irrelevant to phonological feature specification. Yet Anderson also notes (p. 135) that "insofar as possible. . . we should attempt to establish a correspondence between the features [of tones] and the unitary, independently controllable parameters of articulation, acoustics, and/or perception. . . ." It is the goal of the present study to investigate whether the moraic alignment hypothesis for Thai tones can offer a clear correspondence between the autosegmental features required to account for phonological patterning (as in (2)) and the crucial perceptual cues to tone identification (as proposed in (3)).

Peak alignment. Researchers on pitch perception, whether they support a unitary or compositional analysis of pitch contours, have found the timing of pitch inflections to be of crucial importance in cueing both tonal and intonational contrasts (e.g., Bruce & Garding 1978, Garding et al. 1986, Silverman & Pierrehumbert 1990, Shen & Lin 1991, Prieto et al. 1995, Hermes 1997, Xu 1998, 1999a, 1999b, Arvaniti et al. 1998, Ladd et al. 1999, 2000, Gussenhoven 2004 and references therein). While researchers on Thai tonal perception (specifically Abramson and Gandour as cited above) have focused on direction and slope of pitch change, the timing of pitch inflections has also been found to be relevant. The role of peak timing in Thai is supported by a finding reported by Gandour et al. (1991), in a study of variability in the production of citation forms by 20 Thai speakers. The study found that high tones may sometimes be produced in citation form with a final fall (perhaps intonational), so that for some speakers both high and falling tones occur with a rise-fall contour. Gandour et al. propose that, for at least some speakers, the most reliable difference between the two tones lies in the <u>timing</u> of the peak in the tonal contour, early for the falling tone and late for the high tone. They conclude that " it is not unreasonable to suggest that the timing of the turning point may serve as a possible cue underlying the distinction between the Thai falling and high tones" (p. 358).

A study of the perception of synthesized Thai speech by Mixdorff et al. (2002) also supports the importance of peak alignment in the perception of Thai tones. Mixdorff et al. model the pitch changes in Thai tones in terms of "tone commands" which specify high and low inflection points in the F0 contour. Their proposed tone commands are shown in (4).

(4) Tone commands (Mixdorff et al. 2002)

a. mid = no tone command

b. low = low tone command

c. rising = low followed by high

d. high = high tone command **late in the syllable**

e. falling = high tone command **early in the syllable**

In the Mixdorff perceptual study, the cues to Thai tones are argued to be specific high and low points, sometimes aligned to specific points in the syllable, rather than whole contours or slopes. The study also shows that simplification of the falling tone (with only a single "high tone command") is more consistent than simplification of the rising tone, which was found to require both a high and low tone command.

Goals of the Present Study

Extending previous work on the phonology, acoustic realization, and perception of Thai tones, the goals of the present study are to test the following hypotheses:

1. Previous research has shown that falling and rising tones are simplified in non-final position in connected speech. It is hypothesized that such tonal simplification does not result in perceptual neutralization, but that all five contrastive tones remain distinct in stressed syllables. This hypothesis is tested in Experiment 1, in which Thai listeners are asked to identify lexical items produced in natural speech with both full contours (Figure 1) and simplified contours (Figure 2b).

2. Morén and Zsiga (2006) propose the representations in (2) in order to account for the phonological and acoustic patterning of Thai tones. Based on these representations, they further propose that the crucial perceptual cues to Thai tonal contrasts are high and low pitch inflections at syllable midpoint and endpoint, as in (3) above. It is hypothesized that the presence of pitch inflections at syllable midpoint and endpoint will provide more consistent cues to tone identification, in both citation form and connected speech, than will overall contour shape or slope. This hypothesis is tested in Experiments 2 and 3, in which Thai listeners are asked to identify the tones of syllables in which the pitch has been digitally altered in stepwise fashion at midpoint and endpoint. Experiment 2 presents syllables in citation form; Experiment 3 presents syllables in connected speech.
3. Morén and Zsiga (2006) further predict that pitch targets are aligned to the right edges of moras. It is hypothesized that tonal identifications will change if peak alignment is varied, even if overall shape and slope remain constant. This hypothesis is tested in Experiment 4, in which Thai listeners are asked to identify lexical items in which peak alignment but not contour shape is varied.

EXPERIMENT 1: NATURAL SPEECH

This experiment tests whether tonal contrasts are neutralized in non-final stressed syllables in connected speech, even with radical simplifications such as those shown in Figure 2.

Stimuli for Experiment 1

The stimuli used for this perceptual study are a subset of the data analyzed in Morén and Zsiga (2006); example tokens are graphed in Figures 1 and 2b above. A female speaker of the Bangkok dialect (in her late twenties and a graduate student at Georgetown University at the time of the study) recorded the stimuli. Three repetitions of each of the five possible tonal shapes on the syllable /na:/ (Table 1) were recorded in each of four different contexts, as shown in (5).

(5) Phrases for Experiment 1

a) Citation: in isolation, preceded and followed by a pause;

b) M__M: in a carrier phrase, preceded and followed by a mid-toned syllable;

[níd bɔ̀:k na: ______ kʰɯ: kʰamtɔ̀:p]

N. tell N. ______ be answer

Nid told Naa that ______ was the answer.

c) H__H: in a carrier phrase, preceded and followed by a high-toned syllable;

[níd bɔ̀k ná: ______ lɛ́ʔ lam kʰɯ: kʰamtɔ̀:p]

N. tell aunt ______ and stalk be answer

Nid told her aunt that ______ and "stalk" were the answers.

d) L__L: in a carrier phrase, preceded and followed by a low-toned syllable.

[tɤ:m kʰam naɪ cɕɔ̂ŋwâ:ŋ ráwà:ŋ ________ kàp náp]

fill word in blank between ________ and count

Fill in the blank between ________ and "count".

The speaker read the materials from index cards on which the sentences were written in Thai script. (The tonally contrasting words are all orthographically distinct.) The different contexts were recorded in separate blocks, the better to ensure that the target word would receive stress, but sentences were randomized within the blocks. Sentences containing /na:/ as the target syllable were interspersed with sentences with other target syllables, as part of the larger acoustic study reported in Morén and Zsiga (2006). A total of 60 tokens of /na:/ were recorded by this speaker (5 tones x 4 contexts x 3 repetitions). Tokens were digitized at 40K and individual tokens of each sentence and each target word were extracted using the Praat signal analysis program for the Macintosh (Boersma & Weenik 2003).

Presentation of the stimuli for Experiment 1

The stimuli were played to 10 listeners, all native speakers of the Bangkok dialect, between the ages of 25 and 30.[5] The stimuli were presented in 3 different conditions, with tokens randomized within each condition.

(6) Presentation conditions:

a) citation condition: citation forms

b) sentence condition: full sentences (with the different tonal contexts interspersed)

c) excised condition: target syllables extracted from the full sentences

[5] Originally, the same ten speakers, 6 women and 4 men residing in Bangkok, participated as listeners in all three conditions in Experiment 1. Subsequently, an omission was discovered in the presentation of the citation form stimuli to five of the ten listeners, which made the data from that condition unusable. Data from five other speakers of the dialect, who had been tested on the same citation forms in a pilot version of this study, was substituted, for the citation condition only. Identifying citation form tones was a very easy task for all the Thai listeners, and all listeners performed at or near ceiling.

Conditions (b) and (c) used the same tokens, except that in condition (b) the whole sentence was heard, and in condition (c) only a single syllable was heard. All listeners heard the citation forms first, then the full sentences, then the extracted syllables. Within each condition, two different randomizations were used, for five subjects each. Participants were seated in a quiet room and heard the stimuli over headphones as they were played from a laptop computer. There was a two-second delay between the presentation of the stimuli.

For each of the three conditions, listeners were given an answer sheet on which the five words corresponding to the syllable /na:/ were written in numbered rows, in Thai script. In conditions (a) and (c), listeners were instructed to circle on the answer sheet the word they heard. In condition (b), they were given a separate sheet on which the context sentences were written, with blanks where the target words would occur. Listeners read through the sentences prior to beginning that portion of the experiment, and kept this sheet at hand to refer to as necessary. For condition (b), the listeners were asked to circle the word that filled in the blank in the sentence they heard.

Results of Experiment 1

Listeners achieved 98.67% correct in identifying tones in citation form. Eight subjects made no errors in identification; two subjects made one error each: a high tone misidentified as falling, and a rising tone misidentified as high.

Results from the sentence condition are shown in the confusion matrix in Table 2. Overall, listeners achieved 77% correct, less accurate than the citation forms, but well above chance (20%). By far, the most common confusion was for low tones to be misidentified as rising. Rising tones were also often misheard as low; mid tones were sometimes misheard as high. Falling tones were heard most accurately, being correctly identified 96% of the time, compared to 59% correct for the low tone. When the data are broken down by surrounding context, the L__L context is most accurate, with 91% of tokens accurately identified in this context, vs. 64% of tokens in the H__H context.

TABLE 2 ABOUT HERE

TABLE 3 ABOUT HERE

Results from excised condition are shown in the confusion matrix in Table 3. When attempting to identify the tones of excised syllables, listeners achieved only 55.2% correct. This is on average well above chance, but some tones in some conditions, particularly mid tones, were identified at only chance levels. The falling tones were again the most accurately identified.

Discussion of Experiment 1

For Thai listeners, identifying naturally-produced tones in citation forms is an easy task. The results reported here are nearly identical to the 98.6% correct reported by Abramson (1975).

Lower scores in the sentence context are to be expected. The speaker whose productions were used spoke very quickly, with a lot of tonal reduction and coarticulation, as shown in Figure 2. The listeners had particular trouble with the rising tones, which may be attributed to the fact that simplification of rising tones is variable (Potisuk et al. 1997, Morén & Zsiga 2006, Kallayanamit 2004), and this speaker used a more extreme reduction than average. The speaker may in fact have been approaching physical limitations on the speed of pitch changes, particularly for implementing pitch rises, which have been shown to take longer than pitch falls

(Sundberg 1979; see also Zhang 2002, Xu 2004 and references therein). The pitch rises on the high and rising tones, which don't begin until the middle of the syllable, are relatively small. The pitch rise of the falling tone begins early, and thus reaches the top of the pitch range, but only late in the syllable. Listeners may also have had difficulty in deciding whether the final rise that distinguishes the low and rising tones, as well as the mid and high tones, should be attributed to contextual effects (the higher onset of a following tone) or to the identity of the target tone (Gow 2003). In the L__L context, where the following tone begins (and stays) low, any final rises can be unambiguously assigned to the target tone, and thus listeners are most accurate in this context.

While coarticulatory effects do influence perception in connected speech, the tonal contrasts are not neutralized, and listeners still perform well. Falling tones, in particular, are identified nearly as well as in citation forms, despite the fact that the "falling" tonal contour in fact rises for almost all of the syllable duration.

There are many reasons for lower scores in the excised context. These syllables were extremely short, some as short as 50 to 80 ms, and several of the listeners reported having difficulty even hearing them as speech. Single syllables are of course usually pronounced with citation contours, which was not the case here. The syllables showed contextual effects, but with the contexts missing it was even more difficult to distinguish rising vs. low and high vs. mid. It is also possible that surrounding syllables carried information that was used in the sentence context to help identify the target tone. Nonetheless some crucial information was still conveyed in the excised condition, and listeners identified most of the tones at rates greater than chance.

The following experiments turn to the question of "What is the crucial perceptual information that allows Thai listeners to identify tones in both citation form and connected speech?" Are there any cues that remain stable across different contexts? The following experiments use tokens for which the pitch contours have been digitally altered.

EXPERIMENT 2

Experiment 2 investigates cues to perception in the citation forms.

Stimuli for Experiment 2

A young female speaker of the Bangkok dialect of Thai (not the same speaker as the one participating in Experiment 1) recorded citation form pronunciations of the syllables /k^ha:/ and /lau/, with all five tonal patterns. These syllables provide minimal sets of lexical items for all five tones, as shown in Table 1. The phrases were digitized at 40K, and the pitch pattern of the utterance was analyzed in Praat using an autocorrelation algorithm (Boersma 1993). For this speaker, the average pitch for a mid tone was determined to be approximately 210 Hz, the peak of a falling tone to be 250 Hz, and the low point of a rising tone to be 170 Hz. Thus the values 250 Hz, 210 Hz, and 170 Hz were chosen as the top, middle, and bottom of the pitch range.

The Praat resynthesis algorithm (Boersma & Weenik 2003) was used to create stimuli in which the pitch over the course of the syllable was systematically altered. The speaker's natural pronunciations of the words [k^ha:] and [lau], with mid tones, were used as the basis of the synthesis. For the resynthesis, pitch values were specified at the onset, midpoint, and endpoint of the syllable, with linear interpolation between specified points. The syllable midpoint was defined as the end of the first mora, that is, the point half way from the release of the onset consonant of the target syllable to the closure for the onset consonant of the following syllable. Effects of moving the pitch inflection to other points in the syllable are tested in Experiment 4, described below. Here, it should be noted that for the /k^{h}a:/ syllables, approximately the first 40 ms of the pitch track were thus obscured by aspiration. At all three target points, pitch values were varied in steps of 20 Hz (170, 190, 210, 230, 250). There was thus a total of 250 stimuli (5 onsets x 5 midpoints x 5 endpoints x 2 syllables).

Note that a subset of these patterns, those with straight-line trajectories across the syllable /k^{h}a:/, partially replicate the stimuli used by Abramson (1978). Abramson used a greater number of straight-line trajectories, but did not include any trajectories where pitch changed direction mid-syllable.

Presentation of stimuli for Experiment 2

The stimuli were played to 10 listeners, 5 men and 5 women between the ages of 22 and 30, all native speakers of the Bangkok dialect. Two different randomizations were used, for five subjects each. Tokens of /k^{h}a:/ and /lau/ were interspersed, with each stimulus presented once to each listener. Participants were seated in a quiet room and heard the stimuli over headphones as they were played from a laptop computer. There was a two-second delay between the presentation of each stimulus.

Listeners were given an answer sheet on which the five words corresponding to the syllable /k^{h}a:/ or /lau/ were written in numbered rows, in Thai script. (Each row consisted of only variants of /k^{h}a:/ or /lau/, corresponding to the syllable that would be presented.) Listeners were instructed to circle on the answer sheet the word they heard.

Results of Experiment 2: Citation Forms

Figure 3 graphs the effect of changes in the onset, midpoint, and endpoint values on the identification of tonal patterns. The differently-shaded areas indicate the percentage of total identifications made for each distinctive tone. Results for the /k^ha:/ syllables are shown in the left column, and for the /lau/ syllables in the right column. (For each graph, n = 1250, 125 patterns x responses by 10 subjects.) For example, Figure 3A shows that /k^ha:/ syllables with a pitch onset of 170 Hz were identified as high 32% of the time and as mid 20% of the time.

The patterns for /k^ha:/ and /lau/ are largely similar. Though there were some differences between the two syllables (discussed below), the identifications made for the two different syllable types were highly correlated (r^2 = .728).[6]

FIGURE 3 ABOUT HERE

Onset values had little effect on tonal identification. In Figures 3A and 3D, the width of the shaded bars remains fairly uniform across the different onset values, especially for the mid, rising, and falling tones. All five tones are well represented at each onset value, although there is a general tendency throughout to choose high and low tones more often than rising and falling. One difference due to onset is evident, however, in the percentages of high vs. low tones. Contours with low onsets were more likely to be identified as high tones, and contours with high onsets were more likely to be identified as low tones, especially for the /lau/ syllables.

[6] To compute a correlation, patterns were first sorted according to the tone as which they were most often identified overall. Then, the number of times that tone was chosen for each syllable was recorded. For example, the pattern 210-230-250 (gradual rise from onset to endpoint) was identified as a high tone in 10 of 10 presentations on the syllable /k^ha:/, and in 8 of 10 presentations on the syllable /lau/. That pattern was thus scored as 10 for /kha:/ and 8 for /lau/. The correlation between the scores for the two syllables across all pitch patterns then computed. A high correlation indicates that the pitch pattern was identified as the same tone with the same (or similar) frequency on both syllables. A low correlation would indicate that the pattern was identified as a different tone depending on syllable type, or as the same tone but at a higher or lower frequency.

The effects of midpoint and endpoint are much greater. Figures 3B and 3E show the effect of changes in midpoint value. Patterns with low midpoint values are predominantly identified as low or rising tones. Patterns with high midpoint values are predominantly identified as falling tones, but may also be identified with mid or high tones. Contours with midpoint values in the middle of the pitch range may be identified as high, low, or mid, but almost never as rising or falling.

Figures 3C and 3F show the effect of pitch values at syllable endpoint. Low endpoint values are associated with low and falling tones. Contours with high endpoint values are predominantly identified as high tones (with rising tones second). Mid-range endpoint values are predominantly associated with mid tones, though identification with any of the five tones is possible.

Figure 3 shows that the interaction between midpoint and endpoint value (or alternatively, the direction and extent of pitch change in the second half of the syllable) is most important for tone identification. In order to make this clearer, the interaction of midpoint and endpoint values are graphed in Figure 4.

FIGURE 4 ABOUT HERE

Each grid in Figure 4 shows the number of times (out of 100 presentations) that a given midpoint-endpoint combination was identified as a particular tone. Data is pooled over subjects (10), syllables (2), and onsets (5). Each row represents a specific midpoint value, each column a specific endpoint value. Thus Figure 4A shows that the combination of midpoint 250/endpoint 250 (level high in second half of the syllable) was identified as a high tone in 63 of 100 presentations, but that the combination midpoint 210/endpoint 250 was identified as a high tone in 94 of 100 presentations. The shading in the grids is proportional to the strength of the identification. Although there are some areas of ambiguity (light gray squares), the distribution of the dark squares shows that each tone occupies a separate area of the space defined by midpoint/endpoint combinations.

The patterns most clearly identified as high tones (4A) have a midpoint value in the middle of the pitch range, and a high endpoint. A high level pattern (midpoint250/endpoint250) was identified as high in the majority of cases, but this pattern was also often identified as mid (4C) or even falling (4D). Further, more tones were identified as high than as any other, and any pattern with a rising slope in the second half of the syllable (including midpoint170/endpoint210) could be identified as high.

Low tones (4B) occupy the bottom corner of the grid. The patterns most consistently identified as low have both a low midpoint and low endpoint. Patterns with a midrange midpoint and low endpoint (midpoint210/endpoint170) were predominantly identified as low (69%), but could also be identified as mid (15%) or falling (14%).

Mid tones (4C) occupy the middle of the grid. Patterns most consistently identified as mid had both midpoint and endpoint values in the middle of the pitch range. However, any flat or slightly falling contour (except at the very bottom of the pitch range) could be identified as mid: there is a pattern of darker squares along the top-left to bottom-right diagonal in 4C.

Falling tones (4D) occupy the upper right corner of the grid. Patterns with a high midpoint and low endpoint were consistently identified as falling. In fact, any pattern with a high midpoint could be identified as a falling tone, with the strength of the identification weakening as the endpoint gets progressively higher.

Rising tones (4E) were chosen least often. Only patterns with a very low midpoint and very high endpoint were consistently chosen as rising. No pattern with a midpoint value above 170 Hz was identified primarily as a rising tone. As the midpoint becomes higher, patterns with a high endpoint become ambiguous between rising and high (compare the first columns of 4E and 4A). As the endpoint becomes lower, patterns with a low midpoint become ambiguous between rising and low (compare the bottom rows of 4E and 4B).

In order to facilitate comparison with previous studies, specifically Gandour (1978, 1983) and Abramson (1978), Tables 4 and 5 give two further views of the data. As was noted in above, Gandour tested the perception of various pitch contours, and concluded that slope of pitch change was the most salient perceptual for Thai listeners. Table 4 shows the results for all contours in the present study that included a steeply sloping contour, either a fall from the top of the pitch range to the bottom, or a rise from the bottom of the pitch range to the top, including those where the fall or rise took place in the first half of the syllable. (Because Figure 4 pools over different onset values, pitch change over the whole syllable is not inferable from Figure 4).

TABLE 4 ABOUT HERE

Table 4 shows that if Thai tones are to be described in terms of rises and falls, it must be rises and falls in the second half of the syllable. All pitch patterns that fell from a high midpoint to low endpoint (4C) were primarily identified as falling, regardless of onset value and shape of pitch change in the first half of the syllable. The reverse holds for rising tones: all pitch patterns that rose from a low midpoint to high endpoint (4F) were primarily identified as rising. Presence of a steep rise or fall was not sufficient for identification as a rising or falling tone, if the steep rise or fall was not localized in the second half of the syllable. A straight-line rise from onset to endpoint was consistently identified as a high tone, and a straight line fall was consistently identified as low (4A,D). Only 2% of patterns with a steep fall from onset to midpoint (4B) were identified as falling. Consistent with Figure 4, these low-midpoint patterns were identified as either low or rising, depending on endpoint value. Only 2% of patterns with a steep rise from onset to midpoint (4E) were identified as rising. Again consistent with Figure 4, these high-midpoint patterns were primarily identified as either high or falling. On the other hand, presence of a steeply sloping trajectory in any part of the syllable was sufficient to rule out identification as a mid tone: none of the patterns in Table 4 was primarily identified as mid.

Also as noted above, Abramson (1978) tested the perception of straight-line pitch trajectories across Thai syllables. Table 5 shows the results for all straight-line trajectories tested in the present study. The final column of Table 5 shows results from Abramson (1978) for the comparable trajectory (if any), read from the graphs presented in that study and adjusted for the different pitch ranges used. Grayed cells indicate trajectories not tested by Abramson.

TABLE 5 ABOUT HERE

Table 5 shows that it is the level trajectories that are primarily identified as mid tones. Only level trajectories at the very top or bottom of the pitch range are primarily identified as high or low in the present study. Comparison of the final two columns shows that the present data on level trajectories is consistent with Abramson (1978). The crossover points from high to mid and mid to low do not occur at the exact same places, but this is only to be expected given the differences in pitch range between the two studies.

A gradually falling trajectory is also identified as mid, if it ends in the middle of the pitch range (250-230-210). Consistent with Abramson (1978), adding upward or downward movement to the trajectories increases their acceptability as “static” tones, and decreases confusability with mid tones. In both the present study and in Abramson, trajectories that rise from mid-range onset to high endpoint are identified primarily as high, and trajectories that fall from mid-range onset to low endpoint are identified primarily as low. In the present study, a high level pattern (250-250-250) is identified as a high tone only 55% of the time, but a contour that rises from the bottom of the pitch range to the top (170-210-250) is identified as a high tone 85% of the time. Similarly, a pattern that falls from the top of the pitch range to the bottom (250-210-170) is identified as low 100% of the time, but a level low (170-170-170) is identified as low only 80% of the time.

Because the trajectories reported in Table 5 are all straight-line interpolations across the syllable, the rising and falling trajectories tend to be in the middle of the pitch range at syllable midpoint. Thus, consistent with Figure 4, trajectories with low endpoints are identified as low tones and trajectories with high endpoints are identified as high tones. On the other hand, these straight line trajectories, with no mid-syllable inflection, are almost never identified as rising or falling tones, no matter how steep the slope. This final finding is the one point at which the data in the current study do not agree with Abramson’s (1978) data. Abramson found that a straight-line rise from the bottom to the top of the pitch range was consistently identified as a rising tone. In the present study, this pitch pattern was usually identified as high. Abramson did not test steeply falling trajectories.

Discussion of Experiment 2: Citation Forms

Consistency with moraic alignment. Results of Experiment 2 are consistent with the predictions of the moraic alignment hypothesis. As shown in Figure 4, pitch inflections at syllable midpoints and endpoints provide a good basis for categorizing the five contrastive tonal patterns in citation form. In general, the cues are as predicted. Falling tones must have a high midpoint and low endpoint, rising tones a low midpoint and high endpoint. High tones must have a high endpoint and midrange midpoint, and mid tones have values that remain in the middle of the pitch range. One unexpected finding is that patterns with a low midpoint and low endpoint were most consistently identified as low tones; patterns with a mid-range midpoint and low endpoint were ambiguous between low and mid. While the importance of a low endpoint is predicted by the moraic alignment hypothesis, it was predicted that a mid-low pattern would be unambiguously identified as a low tone. The fact that the low tone reaches its target earlier than predicted may be attributed to two phonetic tendencies: the tendency for pitch to fall somewhat even in the absence of any active lowering gesture (Erickson 1976), and the tendency for pitch falls to be implemented more quickly than pitch rises (Sundberg 1979). Thus, even if active lowering doesn't begin until the second mora, the speaker may reach the bottom of the range before the end of the syllable.

Also unexpected was the finding that patterns with a low midpoint and mid-range endpoint were ambiguous between rising and high. The pattern midpoint170-endpoint 210, for example, was identified as a rising tone in 49% of cases, but as high in 33% of cases. Since such a token has neither a mid-range midpoint nor high end-point, the moraic alignment hypothesis predicts this token should not be identified as a high tone at all. These unexpected findings indicate that overall contour shape and slope are playing a role in Thai tone perception, as discussed further below, though they are not the dominant cues.

Comparison with previous studies. Participants in this experiment reported that they found the task very difficult. Many, if not most, of the contours presented were very different from the tonal contours of naturally-produced Thai speech (Figure 1). As was shown in Figure 4, listeners often disagreed on the tone identity of a particular pattern. (Only 8 of 125 patterns, 6.4%, were identified as the same tone all 20 times they were presented.) Participants reported that they found the /lau/ contours particularly difficult and confusing. This asymmetry was unexpected. If anything, it might be predicted that the /k^{h}a:/ syllables would be harder to identify, because the first part of the pitch contour was obscured by aspiration, conveying less information to the listener. However, it appears that listeners found the extra information in the /lau/ syllables confusing rather than helpful. They reported that they "heard two different tones," and had to choose between them.[7]

Perturbations caused by the laryngeal configurations of onset stops are one reason why paying attention to the pitch trajectory in the second half of the syllable, not the first, makes sense as a perceptual strategy. House (1990) shows that, in general, abrupt spectral changes associated with syllable onsets mean that tones are better perceived later in the syllable rather than earlier. Further, acoustic studies (Gandour et al 1994; Potisuk et al. 1997) have found significant perseverative coarticulation of tones in Thai connected speech. Gandour et al. (1994) found that the tone of a preceding syllable could affect the pitch trajectory of the following syllable for up to 60% of the target syllable's duration. Thus, pitch trajectories in the first half of

[7] A flat contour at the beginning of the syllable, which would be more clearly audible in the /lau/ than in the /k^{h}a:/ cases, may have predisposed the listeners to choosing mid more often for /lau/, a pattern that can be discerned by comparing the two columns in Figure 3.

a syllable are much more variable and contextually-influenced than those in the second half of a syllable, making the second mora a much more reliable carrier of tonal contrast. When the laryngeal configuration of the second mora is also contextually constrained (as in obstruent-final syllables), possible tonal contrasts are reduced (from five to two) lightening the perceptual load (see Morén & Zsiga 2006 for discussion).

However, when speakers had to decide on tone identification in ambiguous tokens, overall pitch direction over the course of the syllable, and thus onset value, seems to have played a role. For example, those patterns most strongly identified as high tones are those that approximate the shape of the tone in natural speech: midpoint in the mid range, and endpoint in the high range. However, any contour that had a generally rising shape, and that did not clearly meet any of the criteria outlined above for the other tones, was likely to be identified as a high tone. Similarly, gradually falling contours could be ambiguous between mid and low. The cases where overall slope apparently matters more than specific midpoints or endpoints, however, are confined to the most ambiguous patterns, where none of the criteria listed in (4) are met (specifically when the midpoint is extreme but the endpoint is not), and where the trajectories are least similar to those of naturally-spoken Thai. For example, the pattern 170-170-190 was surprisingly identified as a high tone in the plurality of cases (9/20), apparently because its overall shape (flat in the first half, rising in the second) approximates that of the high tone, though in the wrong half of the pitch range. This pattern was, however, also identified as low five times, rising three times, mid twice, and falling once.

The results of Experiment 2 show clearly that it is the midpoint and endpoint values of a pitch contour that are the most robust cues to tonal identification in Thai citation forms, though onset and overall slope are not irrelevant, and may play a decisive role in ambiguous cases. The cues for each tone are listed in Table 6, along with the proposed moraic representations.

TABLE 6 ABOUT HERE

These findings are consistent with the moraic alignment hypothesis, under the assumptions that phonological H and L tones are realized as pitch targets aligned at the right edge of the associated mora and that pitch on toneless moras is realized as a fall to or within the mid range. As Anderson (1978) has noted, however, it is always possible to convert endpoints into a slope, or to define a slope by its endpoints. Therefore, the findings above are also consistent with the findings of previous perceptual studies (e.g., Abramson 1978, Gandour 1983) that have found pitch height and contour shape (or slope) to be the crucial cues to tone identification. Making the translation from endpoints to slope, one can say that a rising tone is cued by a steep rise, a falling tone by a steep fall, a high tone by a gentle rise, a low tone by a gentle fall, and a mid tone by a flat contour. It must be noted, however, that specification of a particular shape is not sufficient to predict tonal identity. The timing of the pitch change with respect to sub-syllabic constituency must also be specified, as was demonstrated by the data in Table 4.

The reason for the difference between the present study and Abramson (1978) regarding responses to straight-line rising trajectories (Table 5C) is not clear. Abramson's listeners consistently identified as rising a trajectory that rose linearly from bottom of pitch range to top. In the present study, such a trajectory was almost never identified as a rising tone, but rather as high. It may well be that the rising tone in Thai is undergoing a diachronic change that is leading to a difference in perception over the 30 years (a generation, since both studies tested college-age students) that separates the two studies. Several recent acoustic studies (including Gandour et al. 1991, Potisuk et al. 1997, Morén & Zsiga 2006 (Figure 1)), have found that even in citation form the final rise of the rising tone does not reach the top of the pitch range, as was reported in Abramson (1962). Thus "rising" tones may be becoming less strongly associated with steep rises and high endpoints, with perceptual attention shifting to the importance of a low midpoint. As was shown in Figure 4, no contour with a midpoint above 170 Hz was identified primarily as rising in the present study, no matter how steep the pitch trajectory or how high the endpoint.

The hypotheses concerning the importance of the alignment of pitch peaks with moraic boundaries, and with movement in the first half vs. the second half of the syllable, are investigated more fully in Experiment 4. Experiment 3 now turns to the question of whether the proposed perceptual cues remain constant across citation forms and connected speech.

EXPERIMENT 3: CONNECTED SPEECH

Stimuli for Experiment 3

Experiment 3 investigates perception of Thai tones in connected speech. A young adult female speaker of the Bangkok dialect (the same speaker as in Experiment 2) recorded ten repetitions of the phrase shown in (6), a shorter version of one of the sentences used in Experiment 1, filling in the blank each time with one of the five tonal variants of /na:/ and /lau/. (The syllable /k^{h}a:/ was not used so as to avoid the effects of the initial aspirated stop.) Note that the target syllable is surrounded by syllables with mid tones.

(7) Template for pitch manipulation.

/tɤ:m ____ naɪ cç͡ɔ̂ŋwâ:ŋ /

fill ___ in blank

"Fill in the blank with ____"

The phrases were digitized at 40K, and the pitch patterns were analyzed in Praat using an autocorrelation algorithm with a 20 ms window (Boersma 1993). As in Experiment 2, the average pitch for a mid tone for this speaker was determined to be approximately 210 Hz, the top of the pitch range approximately 250 Hz and the bottom of the pitch range approximately 170 Hz. To create the stimuli for Experiment 3, the pitch patterns of the utterances produced with [lau] and [na:] (mid tones) were digitally altered using the Praat reysnthesis algorithm in the following ways.

On the syllable /tɤ:m/, consistent with the speaker's natural productions, pitch was set to fall from 250 Hz to 210 Hz during the vowel, then to remain at 210 Hz through the end of the [m]. (The high onset on this mid tone may be attributed to the effect of the preceding voiceless stop.) Thus all target syllables, the second word of the utterance, began in the middle of the pitch range.

Pitch targets were then set at the midpoint and end of the target syllable. Each of these targets was varied through nine steps of 10 Hz each, from 170 Hz to 250 Hz, for a total of 81 distinct patterns.[8] Finally, pitch was set at 210 Hz at the midpoint and end of the syllable following the target syllable ([naɪ]). This allowed for a fairly gradual reset to mid following the target syllable, again consistent with this speaker's natural productions and the patterns reported in the literature. The pitch pattern on the last word of the utterance ([cɕɔ̂ŋwâ:ŋ]) was left unaltered. Because the altered and unaltered portions of the pitch contour were separated by a

[8] Because onset was not varied, it was possible to use a smaller step size than was used in Experiment 2 (10 Hz rather than 20 Hz) in creating the tokens, allowing for a finer-grained analysis without exceeding the number of tokens a participant was able to tolerate listening to in a single session.

voiceless affricate, and because the values in the altered portion were chosen to correspond to this speaker's voice, there was no obvious discontinuity.

FIGURE 5 ABOUT HERE

Figure 5 shows an example of one of the stimuli, in which the target midpoint is set at 190 Hz and the target endpoint at 230 Hz, approximating the shape of a naturally-produced high tone. (This contour was in fact identified as high in 39/40 presentations.) Pitch falls and then levels out on the first syllable. On the second syllable, pitch falls more gradually during the first mora, then rises steeply during the second, reaching a peak at the end of the vowel, and then returns to mid (210 Hz) in the middle of the following syllable. There is a break during the voiceless affricate, and then the natural pitch contour is used for the final word. Note that on the penultimate syllable [cɕɔ̂ŋ], which is lexically-specified as having a falling tone, pitch is actually realized as a high plateau.

Presentation of the stimuli for Experiment 3

The stimuli were played to 10 listeners, 4 men and 6 women between the ages of 24 and 32, all native speakers of the Bangkok dialect. (Three of these listeners also participated in Experiment 2). Two different randomizations were used, for five subjects each. Tokens of /na:/ and /lau/ were interspersed, and each utterance was heard twice, for a total of 324 stimuli (81 patterns x 2 syllables x 2 repetitions). Participants were seated in a quiet room and heard the stimuli over headphones as they were played from a laptop computer. There was a two-second delay between the presentation of each stimulus.

Listeners were given an answer sheet on which the five words corresponding to the syllable /na:/ or /lau/ were written in numbered rows, in Thai script. (Each row consisted of only variants of /na:/ or /lau/, corresponding to the syllable that would be presented.) Listeners were instructed to circle on the answer sheet the word they heard.

Results of Experiment 3

Results are presented in Figure 6 The format follows that of Figure 4. Each grid represents identifications for a different tone. Each row corresponds to a particular midpoint, each column to a particular endpoint. Two differences from Figure 4 are that the step size is smaller (thus there are more cells in each grid), and the possible total for each cell is 40 rather than 100.

FIGURE 6 ABOUT HERE

Falling tones (6D) have shifted from upper right (high midpoint/low endpoint) to upper left (high midpoint/non-low endpoint). Rising tones (6E) have shifted from lower left (low midpoint/high endpoint) to lower right (low midpoint/low endpoint). Correspondingly, high tones (6A) have shifted down in midpoint value, and low tones (6B) have shifted up. Thus, surprisingly, the contours that have the greatest change in pitch are identified as high or low (6A, B), while the tones identified as falling and rising (6D, E) have fairly flat contours.

Only the mid tones (6C) appear to be identified with the same pitch contours in both citation form and connected speech. In order to be identified as a mid tone, the pitch contour had to remain fairly flat and close to the midline, with neither an extreme midpoint nor extreme endpoint. As in citation form, no contour that deviated more than 20 Hz from the midline was identified predominantly as a mid tone.

In order to have been identified as falling, a contour must have a midpoint in the top half of the pitch range, and the endpoint must *not* fall below the midline. Those tones identified most strongly as falling do not fall at all. (Or at least do not fall at all during the target syllable. Pitch may fall back to the mid range on the following syllable, but the pitch drop is no greater than for those contours identified as high tones.) The contour 210-240-250, with pitch rising over the course of the target syllable, was identified as falling in 40/40 presentations. (More accurately, this connected-speech pitch pattern was identified with the lexical item that typically bears a rise-fall contour in citation form, a contour that linguists have termed "falling".) If the endpoint of a contour fell below 210 Hz, the contour was identified as a primarily as low tone, not a falling tone.

Conversely, patterns heard as rising (4E) must have a low midpoint (170 or 180 Hz), and must *not* rise above the midline at endpoint. As was the case in the citation forms, no contour with a midpoint above 180 Hz was predominantly identified as rising. Contours that in fact fell over the entire course of the target syllable, such as 210-180-170 were predominantly identified as rising tones, although contours with midpoints at 180 Hz, rather than the very lowest 170 Hz, were sometimes ambiguous between rising and low or mid. Unlike the results seen in the citation forms, however, if the endpoint of the connected speech tone rose above the midline 210 Hz, the contour was identified as high rather than rising.

In fact, in an apparent reversal of the citation form identifications, all contours with a low midpoint and high endpoint were consistently identified as high, not rising, and contours with a high midpoint and low endpoint were identified as low, not falling.

Low was chosen most often, but also with the least certainty. Any contour with endpoint below 210 was identified primarily as low, as long as the midpoint value was above 180. Midpoint values for low tones could be anywhere between 190 and 250 Hz. Consistent with the confusions of low and rising tones found in Experiment 1, many patterns in this study were ambiguous between low and rising. Tokens with a high midpoint and low endpoint are primarily identified as low, but are (surprisingly) often heard as rising as well.

Results of Experiment 3 are summarized in Table 7.

TABLE 7 ABOUT HERE

Discussion of Experiment 3

There is an apparent reversal in perceptual cues between citation form and connected speech. In citation form, falling tones must end below the midline. In connected speech, they must end above the midline. In citation form, rising tones must end above the midline. In connected speech, they must end below the midline. Though these connected speech identifications may look surprising; they are consistent with the connected speech patterns reported in the literature, and as illustrated in Figure 2 above. In connected speech in Thai, falling tones do not fall, and rising tones often do not rise, and these generalizations are reflected in the identifications made by the listeners in Experiment 3.

The results of Experiment 3 are consistent with the moraic alignment hypothesis, as seen by a comparison between the cues and the representations in Table 7. The cues that are consistent between citation form and connected speech (summarized in Table 8) are just those predicted by a representation based on moraic alignment. Perceptual cues based on slope (also summarized in

Table 8) are not consistent across different contexts. Except for the mid tone, slope does not prove to be a consistent cue to tone identification. For example, an abrupt fall is required for falling tones in citation form, but is prohibited for falling tones in connected speech.

TABLE 8 ABOUT HERE

A question arises, however, about cases where the cues are conflicting. How are contours with both an extreme midpoint and extreme endpoint perceived? In citation form, contours with a high midpoint and low endpoint are identified as falling, as predicted by the moraic representations. But in connected speech such contours are usually identified as low, or as rising. Why the switch? One plausible explanation is that, in connected speech, Thai listeners expect only one pitch inflection per syllable. When there is more than one inflection, and cues conflict, the later of the two takes precedence. Thus, if there is a high midpoint and low endpoint, it is the low endpoint that determines the identification, in this case as a low tone. This end-ward bias is consistent with the findings of House (1990), that for the task of tone identification in all tone languages, listeners pay more attention to the ends of syllables than to the beginnings.

Slope and overall shape appear to be playing a role here as well. The results in Experiments 2 and 3 show that if a pitch pattern can be matched to the lexical representation of a given item, in terms of high and low pitch targets at appropriate points, the identification will be made on that basis. As was the case in Experiment 2, many of the patterns presented in this study did not match any possible tonal pattern in Thai. In that case, if a pitch pattern can not be matched to any of the target cues in Table 8, listeners have to judge the tone to which the pattern is most similar. Then, consistent with the findings of Gandour (1978 et seq.), overall contour slope and shape are salient cues on which to make similarity judgements. The findings of the present study show, however, that these cues become predominant only in the ambiguous cases, where target-based cues fail to produce an identification.

Experiment 4 now turns to the question of the timing of pitch inflections.

EXPERIMENT 4

Experiments 2 and 3 varied the height of pitch values at midpoint and endpoint. The temporal alignment of the pitch inflections was not varied, however. This experiment now investigates the question of how the alignment of pitch peaks and valleys affect tone perception in Thai.

Stimuli for Experiment 4

The basis of the stimuli for Experiment 4 are the same naturally produced utterances as were used in Experiment 3: The syllables [na:] and [lau] in the carrier phrase [tɤ:m ____ naɪ cɕɔ̂ŋwâ:ŋ], "fill in the blank with ____." Again, pitch on the first three syllables was digitally altered, and the pitch on the final two syllables, following the voiceless affricate, was left unchanged.

The duration of the target syllable was normalized to exactly 280 ms. The pitch on the target syllable was altered such that each token had one pitch peak (high inflection) or one pitch valley (low inflection). The alignment of the inflection point was varied over the course of the target syllable, in 20 ms steps from 20 ms after onset to 260 ms after onset, which was just before closure for the palatal stop. In the high inflection (peak) tokens, pitch began at the midline (210 Hz), rose linearly over the course of 140 ms to a peak of 250 Hz, fell linearly over the course of 140 ms back to 210 Hz, then remained at 210 Hz through the end of the third syllable. Thus the slope to and from the peak remained the same in all tokens. The low inflection (valley) tokens repeated the pattern, but with a fall from 210 Hz to 170 Hz instead of a rise. There were thus 13 contours with a high inflection point and 13 contours with a low inflection point. An additional four tokens, for which pitch remained flat at 210 Hz throughout the first three syllables, were also included. Figure 7 illustrates the pitch patterns for the 13 peak tokens.

FIGURE 7 ABOUT HERE

Presentation of the stimuli for Experiment 4

The experimental presentation was the same as that used for Experiment 3. The stimuli were played to 10 listeners, 8 women and 2 men between the ages of 35 and 32, all native speakers of the Bangkok dialect. (Six of these listeners also participated in Experiment 3). Two different randomizations were used, for five subjects each. Tokens of /na:/ and /lau/, with high, low, or no inflection, were interspersed. There was a total of 60 stimuli (30 contours x 2 syllables), and each stimulus was heard once. Participants were seated in a quiet room and heard the stimuli over headphones as they were played from a laptop computer. There was a two-second delay between the presentation of each stimulus.

Listeners were given an answer sheet on which the five words corresponding to the syllable /na:/ or /lau/ were written in numbered rows, in Thai script, and were instructed to circle on the answer sheet the word they heard.

Results of Experiment 4

Results are graphed in Figure 8. Results for the high inflection patterns are shown in 8A; for low inflection patterns in 8B. The graphs show the number of times each tone was identified at each peak location. The level tokens were identified as mid tones 100% of the time, and are not shown in the graphs.

FIGURE 8 ABOUT HERE

The top graph shows that when the peak is reached early in the syllable, from 20 – 40 ms after onset, the syllable is identified as mid. There is then a period of confusion from 60—120 ms, during which time mid identifications are decreasing and falling identifications are increasing, but the predominant identification is of a low tone. When the peak is reached near the halfway point (between 140 and 220 ms), identification of the contour as a falling tone clearly predominates. High tone identifications increase as the peak moves later, but it is only when the peak is reached at the very end of the syllable (240—260 ms) that high tone identifications predominate.

Results for the low inflection tokens show the same effects of temporal alignment. When the valley is reached 20—60 ms after onset, the syllable is identified as mid. As the valley gets later, mid identifications decrease and rising identifications increase, with a crossover occurring at 80 ms. Around 80 ms, the contour may also be identified as high. From 100 ms to 220 ms rising identifications predominate, with the least ambiguous tokens occurring when the valley is reached at or near the syllable midpoint. Low tone identifications increase as the valley gets later, and come to predominate only when the low point is reached at the very end of the syllable.

Discussion of Experiment 4

Experiment 4 shows that peak alignment is an important factor in the identification of Thai tones. Results are again consistent with the moraic alignment hypothesis. Falling tones are identified when a peak is reached near the midpoint of the syllable; rising tones are identified when a low point is reached near the midpoint of the syllable. High and low tones are identified with high and low points reached at the end of the syllable. The results also confirm the findings of Gandour et al. (1991) and Mixdorff et al. (2002) that high and falling tones are distinguished by the location of the pitch inflection, and suggest that low and rising tones are also distinguished by the same parameter.

When the peak is reached early in the syllable, there is confusion and disagreement among the subjects, though it seems that the percept of the flat trajectory in the second half of the syllable predominates, and these tokens are usually identified as mid. This is consistent with the finding in Experiment 2 that citation form contours that had more extreme movement in the first half of the syllable and then leveled off were also confusing and ambiguous. Listeners in Experiment 4 reported that the contours with early peaks "sounded weird." Several listeners in fact noted that the target syllables sounded like normal mid tones, but that the *preceding* syllable sounded high or low instead of mid. Thus is would appear that listeners are attributing an early peak or valley to a preceding syllable, and are using primarily the pitch information in the second half of the syllable to identify the target tone.

Also, consistent with the findings of Experiments 2 and 3, when peak alignment cues are conflicting or ambiguous, overall pitch direction and slope play a more important role in tone identification. Thus when the pitch inflection occurs between about 60 and 120 ms after onset, too late to be attributed to the preceding syllable and too early to be attributed to the mid-syllable inflection of a contour tone, falling trajectories may be identified as low tones and rising trajectories may be identified has high tones.

GENERAL DISCUSSION

This study sought to test three hypotheses, each of which was confirmed, with some caveats.

1. Tones are simplified in connected speech, but such tonal simplification does not result in perceptual neutralization.

Experiment 1 demonstrated that tonal contrasts are not neutralized in stressed syllables, despite the contour simplifications found in connected speech. Falling tones, in particular, were very accurately identified despite the fact that the actual pitch contours on these syllables did not fall at all. Rising and low tones were identified less accurately, however, perhaps because the particular speaker used for the experiment simplified rising tones more often or to a greater extent than is usual. Future research, with a wider variety of talkers, may tease apart effects of context and speaker idiosyncracy.

Syllables excised from connected speech were usually identified at higher than chance rates, but not with great accuracy. This may indicate that some important information is carried on adjacent syllables. It may also be the case, however, that listeners were unable to compensate for contextual effects absent the context, or that the syllables were just too short to be heard accurately.

2. The presence of pitch inflections at syllable midpoint and endpoint will provide more consistent cues to tone identification, in both citation form and connected speech, than will overall contour shape or slope.

Experiments 2 and 3 (Figures 4 and 6) showed that pitch inflections at syllable midpoint and endpoint successfully categorize the tones of Thai in the perceptual space. Falling tones are identified as having a high pitch target at syllable midpoint. A fall from that point is required in citation form, but does not take place in connected speech. Rising tones must have a low pitch target at syllable midpoint. A rise from that point is required in citation form, but is optional in connected speech. In both connected speech and citation form, high tones are cued by a high point at the end of the syllable, low tones are cued by a low point at the end of the syllable. Mid tones are consistently identified by the lack of any pitch inflection.

Pitch slopes and contour shapes also played a role in tone identification in this experiment, particularly when cues to peak location were ambiguous, or conflicting. However, slopes and contour shapes were not consistent between citation form and connected speech. Because slopes can always be derived from endpoints, the results here do not contradict the finding of earlier studies that Thai listeners pay attention to pitch slope. However, previous perceptual experiments (specifically Gandour 1978 et seq.) that emphasize the importance of pitch slope have generally used similarity judgments, not an identification task. In the task of lexical tone identification, height and location of pitch peaks provided the stable perceptual cues.

3. Pitch targets are aligned to the right edges of moras. Tonal identifications will change if peak alignment is varied, even if overall shape and slope remain constant.

Experiment 4 showed that tone identification does change based on peak alignment. Falling tones are associated with a peak at or near syllable midpoint, high tones are associated with a peak at or near syllable endpoint. Rising tones have a low inflection at or near syllable midpoint; low tones have a low pitch target at or near syllable endpoint. Exact alignment of inflection points with the right edge of the mora was not necessary for tone identification: a pattern with a low inflection, for example, was primarily identified as rising if the inflection point occurred within about an 80 ms window centered at syllable midpoint. However, the window of identification is centered on, and peaks at, the syllable midpoint, and patterns become more ambiguous the further from syllable midpoint, the right edge of the first mora, the peak moves.

One interesting finding that was apparent across the three experiments has to do with the realization and perception of the rising tone. In Experiment 1, it was found that low and rising tones were the most confusable (Tables 2 and 3). In Experiment 2, it was found that a pattern was required to have a very low midpoint, but not necessarily a high endpoint, in order to be identified as rising (Figure 4 and Table 5). This was a difference from the finding of Abramson (1978), who found that a straight-line rising trajectory, with high endpoint but mid-range midpoint, was predominantly identified as a rising tone. In the present study, such a pattern was identified as high. In Experiment 3 (Figure 6), it was found that in connected speech, a pattern that had both a low midpoint and low endpoint was ambiguous between rising and low. Ambiguity between low and rising was also seen in Experiment 4 (Figure 8), where it was found that while a pattern with a low inflection point very late in the syllable was predominantly identified as low (60% of tokens), it could also be identified as rising (40% of tokens). Finally, in Experiments 1, 2, and 3, rising was the identification chosen least frequently. These results are consistent with acoustic findings (Figures 1 and 2) that show that in the productions of younger speakers the rise of the rising tone is realized variably in connected speech and weakly in citation form. Taken together, these results suggest that the rising tone is undergoing diachronic change, becoming more strongly associated with a low pitch target and less strongly associated with a high pitch target, even in citation form. A tendency for low tones to reach the bottom of the pitch range earlier than the right edge of the syllable (Figures 2 and 4b) may also be contributing to confusability between low and rising tones.

Overall, the perceptual data confirm the predictions of the moraic alignment hypothesis. While overall pitch slope and contour shape are not irrelevant, especially in ambiguous cases, pitch inflections aligned with moras provide the most stable acoustic cues to the contrastive tones of Thai, in both citation form and connected speech. Alignment of tones with moras accounts for phonological patterning, as argued in Morén and Zsiga (2006). Phonological tone specifications may then be interpreted as pitch targets aligned with the right edge of their specified prosodic domain, the mora, either as part of a Thai-specific phonology-to-phonetics mapping as proposed by Morén and Zsiga (2006) or perhaps as a reflex of a universal trend toward rightward alignment (as proposed by Xu 1998, 2004). The present study has shown that these pitch peaks, aligned at syllable midpoints and endpoints, provide the needed perceptual cues to the tonal contrasts. Thus Thai provides an example of one Asian tone system where a straightforward mapping from phonological autosegments to perceptual cues can be achieved.

REFERENCES

ABRAMSON, A. (1962). The vowels and tones of Standard Thai: Acoustical measurements and experiments. *International Journal of American Linguistics*, **28;2**, part II. (Also published by the Indiana Research Center in Anthropology, Folklore, and Linguistics: Bloomington, Indiana).

ABRAMSON, A. (1975). The tones of Central Thai: some perceptual experiments. In J. G. Harris and J. Chamberlain, (Eds.), *Studies in Tai Linguistics* (pp. 1—16). Bangkok: Central Institute of English Language.

ABRAMSON, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, **23**, 19—325.

ABRAMSON, A. (1979). The coarticulation of tones: an acoustic study of Thai. In T.L. Thongkum, V. Panupong, P. Kullavanijaya, and M.R.K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer Phonetics and Phonology in honor of Eugénie J. A. Henderson*. (pp. 1—9). (Also published in *Haskins Laboratory: Status Report on Speech Research* SR-44: 119-125 (1975)).

ANDERSON, S. (1978). Tone features. In V. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 133—161). New York: Academic Press.

ARVANITI, A., LADD, D. R., & MENNEN, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics,* **36**, 3—25.

BENNETT, J. F. (1994). Iambicity in Thai, *Studies in the Linguistic Sciences* 24(1/2):39-57.

BOERSMA, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise natio of a sampled sound, *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, **17**, 97 – 110.

BOERSMA, P. & WEENINK, D. (2003). Praat 4.19: Doing phonetics by computer. (www.praat.org.)

BROSELOW, E., CHEN, S.-I., & HUFFMAN, M. (1997). Syllable weight: Convergence of phonology and phonetics, *Phonology*, **14**, 47—82.

BRUCE, G. & GARDING, E. (1978). A prosodic typology for Swedish dialects. In E. Garding, G. Bruce and R. Bannert (Eds.), *Nordic Prosody (Travaux de L'institut Linguistique de Lund)* (pp. 219—228). Lund: L'institut Linguistique de Lund.

COHN, A. (2003). Phonological structure and phonetic duration: The role of the mora. *Working Papers of the Cornell Phonetics Laboratory,* **15**, 69 – 100.

DUANMU, S. (1994). Against contour tone units. *Linguistic Inquiry,* **25**, 555-608.

ERICKSON, D. (1974). Fundamental frequency contours of the tones of Standard Thai. *Pasaa,* **4**, 1—25.

ERICKSON, D. (1976). *A Physiological Analysis of the Tones of Thai*. Ph.D. dissertation, University of Connecticut.

ERICKSON, D. (1994). Laryngeal muscle activity in connection with Thai tones. *Festschrift in Honor of Professor Hajima Hirose, RILP, University of Tokyo*, **27**, 135—149.

GANDOUR, J. (1974a). On the representation of tone in Siamese. In J. G. Harris and J. R. Chamberlain, (Eds.), *Studies in Tai Linguistics in Honor of William J. Gedney* (pp. 170—195). Bangkok: Central Institute of English Language. (Also published in *UCLA Working Papers in Phonetics*, **27**, 118-146.)

GANDOUR, J. (1974b). The glottal stop in Siamese: Predictability and phonological description. *UCLA Working Papers in Phonetics*, **27**, 84—91.

GANDOUR, J. (1978). The perception of tone. In V. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 41—76). New York: Academic Press.

GANDOUR, J. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics,* **9**, 20—36.

GANDOUR, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, **11**, 149—175.

GANDOUR, J. (1984). Tone dissimilarity judgements by Chinese listeners. *Journal of Chinese Linguistics,* **12**, 235—261.

GANDOUR, J, & HARSHMAN, R. (1978). Cross-language differences in tone perception: A multi-dimensional scaling investigation. *Language and Speech*, **21;1**, 1—33.

GANDOUR, J, POTISUK, S., PONGLORPISIT, S. & DECHONGKIT, S. (1991). Inter- and intraspeaker variability in fundamental frequency of Thai tones. *Speech Communication,* **10**, 355—372.

GANDOUR, J., POTISUK, S. & DECHONGKIT, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics,* **22**, 474-492.

GANDOUR, J., TUMTAVITIKUL, A.& SATTHAMNUWONG, N. (1999). Effects of speaking rate on Thai tones, *Phonetica,* **56**, 123—134.

GANDOUR, J., WONG, D., HSIEH, L., WEINZAPFEL, B., VAN LANCKER, D. & HUTCHINS, G. (2000). A cross-linguistic PET study of tone perception. *Journal of Cognitive Neuroscience,* **12 (1)**, 207—22.

GARDING, E., KRATOCHVIL, P., SVANTESSON, J. O., & ZHANG, J. (1986). Tone 4 and tone 3 discrimination in Modern Standard Chinese. *Language and Speech*, **29**, 281—293.

GOW, D. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics*, **65(4)**, 575—590.

GUSSENHOVEN, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

HERMES, D. J. (1997). Timing of pitch movements and accentuation of syllables in Dutch. *Journal of the Acoustical Society of America*, **102**, 2390—2402.

HIRANBURANA, S. (1971). *The Role of Accent in Thai grammar*. Ph.D. dissertation, University of London.

HOUSE, D. (1990). *Tonal Perception in Speech*. Lund, Sweden: Lund University Press.

KALLAYANAMIT, S. (2004). *Thai Intonation: Contours, Registers, and Boundary Tones*. Ph.D. dissertation, Georgetown University.

LADD , D. R., FAULKNER, D., FAULKNER, H. & SCHEPMAN, A. (1999). Constant "segmental anchoring" of F0 movments under changes in speech rate, *Journal of the Acoustical Society of America,* **106**, 1543—1554.

LADD, D. R., MENNEN, I. & SCHEPMAN, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch, *Journal of the Acoustical Society of America*, **107**, 2685—2696.

LEBEN, W. 1971. On the segmental nature of tone in Thai. *Quarterly Progress Report, Research Laboratory of Electronics,* **101**, 221—224.

LEBEN, W. 1973. *Suprasegmental Phonology.* Ph.D. dissertation, MIT.

LIN, H.-B. & REPP, B. (1989). Cues to the perception of Taiwanese tones. *Language and Speech*, **32**, 25—44.

LIU, S. & SAMUEL, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, **47**, 109—138.

LUKSANEEYANAWIN, S. (1983). *Intonation in Thai*. Ph.D. dissertation, University of Edinburgh.

MASSARO, D., COHEN, M. M., & TSENG, C.-Y. (1982). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics*, **13;2**, 266—289.

MIXDORFF, H., LUKSANEEYANAWIN, S., FUJISAKI, H. & CHARNAVIT, P. (2002). Perception of tone and vowel quality in Thai. Paper presented at the 7th International Conference on Spoken Language Processing, Denver Colorado, September 2002.

MORÉN, B. & ZSIGA, E. (2006). The lexical and post-lexical phonology of Thai tones. *Natural Language and Linguistic Theory*, **24(1)**, 113—178.

NITISAROJ, R. (2006). *Effects of Stress and Speaking Rate on Duration and Tone in Thai*. Ph.D. dissertation, Georgetown University.

ODDEN, D. (1995). Tone: African languages. In J. Goldsmith (Ed.), *The Handbook of Phonology* (pp. 444—475). Oxford: Blackwell.

PIKE, K. (1948). *Tone Languages*. Ann Arbor: University of Michigan Press.

PRIETO, P., VAN SANTEN, J., & HIRSCHBERG, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, **23**, 492—451.

POTISUK, S., GANDOUR, J., & HARPER, M. (1994). F0 correlates of stress in Thai. *Linguistics of the Tibeto-Burman Area*, **17.2**, 1-25.

POTISUK, S., GANDOUR, J., & HARPER, M. (1996). Acoustic correlates of stress in Thai. *Phonetica*, **53**, 200—220.

POTISUK, S., GANDOUR, J., & HARPER, M. (1997). Contextural variations in trisyllablic sequences of Thai tones, *Phonetica*, **54**, 22—42.

SHEN, X. S. & LIN, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech*, **34**, 145—156.

SILVERMAN, K. & PIERREHUMBERT, J. (1990). The timing of prenuclear accents in English. In J. Kingston and M. Beckman, (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press.

SUNDBERG, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics,* **7**, 71—79.

WAYLAND, R. & GUION, S. (2003). Perceptual discrimination of Thai tones by naïve and experienced learners of Thai. *Applied Psycholinguistics*, **24(1)**, 113—129.

XU, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, **55**, 179—203.

XU, Y. (1999a). F0 peak delay: When, where and why it occurs. In J. Ohala (Ed.), *Proceedings of the International Congress of Phonetic Sciences 1999*. pp. 1881—1884.

XU, Y. (1999b). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, **27**, 55—105.

XU, Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics* **5**: 757-797.

YIP, M. (1982). Against a segmental analysis of Zahao and Thai: A laryngeal tier proposal. *Linguistic Analysis*, **9**, 79-94.

YIP, M. (1989). Contour tones, *Phonology*, **6**, 149—174.

YIP, M. (1995). Tone in East Asian languages. In J. Goldsmith, (Ed.), *Handbook of Phonological Theory* (pp. 476—494). Oxford: Blackwell.

YIP, M. (2002). *Tone*. Cambridge: Cambridge University Press.

ZHANG, J. (2002). *The Effects of Duration and Sonority on Contour Tone Distribution: A Typological Survey and Formal Analysis*. NY: Routledge.

Table 1

Five-way tonal contrast in Thai

mid	na: rice field	k^ha: to be stuck	lau classifier (wood instr.)
high	ná: aunt	k^há: to trade	láu coop
low	nà: custard apple	k^hà: galangal (a spice)	làu group
falling	nâ: face	k^hâ: value	lâu alchohol
rising	nǎ: thick	k^hǎ: leg	lǎu to sharpen

Table 2. Confusion matrices for Experiment 1, condition b: natural speech with tones heard in a sentence context.

Overall						
heard: spoken:	H	L	M	R	F	% correct
H	**71**	4	9	2	4	78.89%
L	3	**53**	4	29	1	58.89%
M	17	1	**68**	0	4	75.56%
R	4	15	0	**70**	1	77.78%
F	2	0	1	1	**86**	95.56%
all tones						**77.33%**
Between mid tones						
heard: spoken:	H	L	M	R	F	% correct
H	**22**	1	5	0	2	73.33%
L	0	**16**	0	14	0	53.33%
M	7	0	**21**	0	2	70.00%
R	0	4	0	**26**	0	86.67%
F	0	0	1	0	**29**	96.67%
all tones						**76.00%**
Between high tones						
heard: spoken:	H	L	M	R	F	% correct
H	**22**	2	3	1	2	73.33%
L	2	**13**	1	13	1	43.33%
M	10	0	**18**	0	2	60.00%
R	3	10	0	**16**	1	53.33%
F	1	0	0	1	**28**	93.33%
all tones						**64.67%**
Between low tones						
heard: spoken:	H	L	M	R	F	% correct
H	**27**	1	1	1	0	90.00%
L	1	**24**	3	2	0	80.00%
M	0	1	**29**	0	0	96.67%
R	1	1	0	**28**	0	93.33%
F	1	0	0	0	**29**	96.67%
all tones						**91.33%**

Table 3. Confusion matrices for Experiment 1, condition c: natural speech with tones heard in excised syllables.

Overall						
heard: spoken:	H	L	M	R	F	% correct
H	**48**	13	3	13	13	53.33%
L	1	**55**	1	24	9	61.11%
M	17	17	**35**	6	15	38.89%
R	10	31	5	**35**	9	38.89%
F	16	5	2	5	**62**	68.89%
all tones						52.22%
Between mid tones						
heard: spoken:	H	L	M	R	F	% correct
H	**19**	3	1	4	3	63.33%
L	0	**17**	0	10	3	56.67%
M	12	3	**5**	4	6	16.67%
R	0	14	1	**13**	2	43.33%
F	10	3	0	4	**13**	43.33%
all tones						44.67%
Between high tones						
heard: spoken:	H	L	M	R	F	% correct
H	**11**	7	1	4	7	36.67%
L	0	**21**	1	7	1	70.00%
M	5	7	**7**	2	9	23.33%
R	2	14	1	**9**	4	30.00%
F	3	1	1	1	**24**	80.00%
all tones						48.00%
Between low tones						
heard: spoken:	H	L	M	R	F	% correct
H	**18**	3	1	5	3	60.00%
L	1	**17**	0	7	5	56.67%
M	0	7	**23**	0	0	76.67%
R	8	3	3	**13**	3	43.33%
F	3	1	1	0	**25**	83.33%
all tones						64.00%

Table 4. Results for Experiment 2: steeply sloped trajectories.

Pitch			Heard as					
onset	**midpoint**	**end**	**H**	**L**	**F**	**R**	**M**	
A. Steep fall over whole syllable								
250	210	170	0	**20**	0	0	0	100% L
B. Steep fall in first half								
250	170	170	0	**20**	0	0	0	100% L
250	170	190	0	**18**	0	1	1	90% L
250	170	210	1	**10**	1	8	0	50% L
250	170	230	3	1	1	**15**	0	75% R
250	170	250	3	0	0	**17**	0	85% R
C. Steep fall in second half								
170	250	170	0	1	**18**	0	1	90% F
190	250	170	1	4	**13**	0	2	65% F
210	250	170	2	0	**16**	1	1	80% F
230	250	170	2	3	**15**	0	0	75% F
250	250	170	0	0	**20**	0	0	100% F
D. Steep rise over whole syllable								
170	210	250	**17**	0	2	1	0	85% H
E. Steep rise in first half								
170	250	170	0	1	**18**	0	1	90% F
170	250	190	3	3	**12**	0	2	60% F
170	250	210	5	2	**7**	1	5	35% F
170	250	230	**8**	1	7	1	3	40% H
170	250	250	**16**	0	2	0	2	80% H
F. Steep rise in second half								
170	170	250	3	1	0	**16**	0	80% R
190	170	250	6	0	1	**13**	0	65% R
210	170	250	3	0	0	**17**	0	85% R
230	170	250	2	0	0	**18**	0	90% R
250	170	250	3	0	0	**17**	0	85% R

Table 5. Results for Experiment 2: straight line trajectories.

Present study									Comparison to Abramson 78
Pitch			Heard as						
onset	midpoint	end	H	L	F	R	M		
A. Level trajectories									
170	170	170	0	**16**	0	2	2	80% L	90% L
190	190	190	0	6	0	0	**14**	70% M	50% M, 50% L
210	210	210	0	2	0	1	**17**	85% M	75% M
230	230	230	1	0	2	0	**17**	85% M	70% H, 30% M
250	250	250	**11**	0	3	0	6	55% H	85% H
B. Falling trajectories									
250	230	210	0	4	1	0	**15**	75% M	
250	210	170	0	**20**	0	0	0	100% L	
230	210	190	0	**13**	0	1	6	65% L	
210	190	170	0	**20**	0	0	0	100% L	95% L
C. Rising trajectories									
170	190	210	**12**	3	0	0	5	60% H	70% R, 20% H
170	210	250	**17**	0	2	1	0	85% H	90% R
190	210	230	**17**	1	0	0	2	85% H	
210	230	250	**18**	0	1	1	0	90% H	85% H

Table 6. Representations and cues in citation form.

<table>
<tr><th>Tone</th><th>Moraic representation</th><th>Target-based cue</th><th>Slope-based cue</th></tr>
<tr><td>Mid</td><td>μ μ</td><td>flat, not more than 20Hz above or below midline</td><td>level</td></tr>
<tr><td>High</td><td>H
|
μ μ</td><td>non-low midpoint and high endpoint</td><td>gradual rise on 2nd mora</td></tr>
<tr><td>Low</td><td>L
|
μ μ</td><td>non-high midpoint and low endpoint</td><td>gradual fall on 2nd mora</td></tr>
<tr><td>Fall</td><td>H L
| |
μ μ</td><td>high midpoint and low endpoint</td><td>abrupt fall on 2nd mora</td></tr>
<tr><td>Rise</td><td>L H
| |
μ μ</td><td>low midpoint and high endpoint</td><td>abrupt rise on 2nd mora</td></tr>
</table>

Table 7. Representations and cues in connected speech.

Tone	Moraic representation	Target-based cue	Slope-based cue
Mid	$\mu\ \ \mu$	no inflection	level
High	H | $\mu\ \ \mu$	non-high midpoint and high endpoint	abrupt rise on 2nd mora
Low	L | $\mu\ \ \mu$	non-low midpoint and low endpoint	abrupt fall on 2nd mora
Fall	H | $\mu\ \ \mu$	high midpoint and non-low endpoint	rise or gradual fall on 2nd mora
Rise	L | $\mu\ \ \mu$	low midpoint and non-high endpoint	fall or gradual rise on 2nd mora

Table 8. Consistency of cues.

tone	target-based		slope-based	
	citation form	connected speech	citation form	connected speech
Mid	**no inflection**	**no inflection**	**level**	**level**
High	non-low midpoint and **high endpoint**	non-high midpoint and **high endpoint**	gradual rise on 2^{nd} mora	abrupt rise on 2^{nd} mora
Low	non-high midpoint and **low endpoint**	non-low midpoint and **low endpoint**	gradual fall on 2^{nd} mora	abrupt fall on 2^{nd} mora
Fall	**high midpoint** and low endpoint	**high midpoint** and non-low endpoint	abrupt fall on 2^{nd} mora	rise or gradual fall on 2^{nd} mora
Rise	**low midpoint** and high endpoint	**low midpoint** and non-high endpoint	abrupt rise on 2^{nd} mora	fall or gradual rise on 2^{nd} mora

Figure legends.

Figure 1. Contour shapes of Thai tones in citation form. Representative examples from one speaker.

Figure 2. Contour shapes of Thai tones in connected speech. Representative examples from two speakers, from Morén & Zsiga (2006).

Figure 3. Effect of pitch target variation at syllable onset, midpoint, and endpoint on percentage of tone identification.

Figure 4. Tone identification in Experiment 2: citation forms. Shaded cells show the number (out of 100) of each midpoint-endpoint combination identified as the given tone. Darkness of the shading is proportional to the strength of the identification. Data is pooled over onsets, syllables, and subjects.
A. Number heard as a high tone.
B. Number heard as a low tone.
C. Number heard as a mid tone.
D. Number heard as a falling tone.
E. Number heard as a rising tone.

Figure 5. The phrase /tɤ:m láu naɪ cç̂ɔ̂ŋwâ:ŋ/, with target syllable (/lau/) midpoint at 190 Hz and target syllable endpoint at 230 Hz. The pitch contour on the first three syllables is synthesized, the pitch contour on the last two syllables is natural.

Figure 6. Tone identification in Experiment 3: sentence context. Shaded cells show the number (out of 40) of each midpoint-endpoint combination identified as the given tone. Darkness of the shading is proportional to the strength of the identification. Data is pooled over syllables and subjects. All onsets are 210 Hz.
A. Number heard as a high tone.
B. Number heard as a low tone.
C. Number heard as a mid tone.
D. Number heard as a falling tone.
E. Number heard as a rising tone.

Figure 7. Stepwise variation in alignment of the high pitch peak. Vertical lines indicate boundaries of the target syllable.

Figure 8. Tone identification in Experiment 4, as a function of peak location.