The Role of Linguistic Experience in the Perception of Thai Tones

Denis Burnham and Elizabeth Francis

University of New South Wales

There are two main approaches to studying the sounds of language. The first is to describe and classify the sounds occurring in the world's languages. From such analyses historical and comparative information about languages and, to some extent, language-users can be gained. The second approach is to study the way in which people produce and perceive the sounds of language. From these analyses information can be gained about the perceptual and cognitive processes involved in language use and how these might develop. Traditionally the first approach has been used by linguists and the second by psychologists. In this study the two approaches are combined to investigate Thai speakers' and English speakers' perception of Bangkok Thai tones and its development over age.

Linguists have long studied phonetic segments—consonants and vowels. Recently psychologists have also begun to study these in order to understand language processing and particularly the age-related development of the processes involved in language learning (Burnham, Earnshaw, & Clark, 1991; Werker & Tees, 1983). In addition, the majority of the world's languages use tones—pitch variations at the level of lexical units, to convey semantic information (Goldsmith, 1994). Linguists have conducted many studies on the tonal attributes of various languages (e.g., Abramson, 1978); however, there have been relatively fewer studies on the *psychological* aspects of tone. This is unfortunate because the processes involved in perceiving pitch variations over time may differ from those involved in perceiving spectral composition, and also in light of the increasing attention now being given to pitch variations in other contexts: the distinctive pitch modulations in infant-directed speech (Fernald, 1989) and affective parent-infant interactions (Fernald, 1993); and music perception by infants and children (Lynch, Eilers, & Bornstein, 1992).

In the study reported here the focus is the tonal system of Bangkok Thai with its five lexical tones: mid, low, high, rising, and falling. In the descriptive part of the study various acoustic parameters were extracted from productions of these tones. In the perceptual part, subjects were given minimal pairs of the tones and asked to make discrimination judgments. Three variables were manipulated: language background (Thai and English speakers), age (adults, 8-, 6-, and 4-year-olds), and processing mode (phonological, involving a 1500 msec interval between the two tones to be discriminated, and phonetic , involving a 500 msec interval between the two tones to be discriminated). Each of these is discussed below in relation to the relevant literature.

Salient Dimensions in the Perception of Tones

Bangkok Thai has five tones, mid, high, low, rising, and falling (Luksaneeyanawin, forthcoming). Mid, high, and low are often labelled static tones and rising and falling, dynamic tones (Abramson, 1978). Research has been conducted on the relative perceptual salience of these. Gandour (1979) conducted a multi-dimensional scaling analysis of Thai listeners' perceptual discrimination of various pitch contours and revealed three important dimensions. The first and most important factor in subjects' perceptual judgments was the relatively low-level acoustic dimension of average pitch. This is also an important cue for speakers of other tonal languages and for speakers of non-tonal languages (Gandour, 1983), and so it appears to be a relatively universal basis for the distinction of tonal contrasts. Saravari and Imai (1983) found a similar dimension, initial F_0 value, to be important in the identification of synthetic versions of the five Thai tones. In addition to these relatively low level dimensions, Gandour (1979) found two rather more linguistically-based dimensions to be important. The direction dimension served to distinguish between rising and falling contours on the basis of pitch movement rather than absolute start and end points. The slope dimension served to distinguish between static and dynamic tones. These results are consistent with those of Abramson (1978, 1996) who found that differences in pitch levels are sufficient for the identification of static tones, although this is enhanced by slow F_0 movement; and that fairly rapid F_0 movement is required for the identification of dynamic tones.

Burnham, Kirkwood, Luksaneeyanawin, and Pansottee (1992), studying Thai- and English-speaking adults' tone and consonant perception found, not surprisingly, that Thai speakers discriminated tonal contrasts better than did English speakers, but that the *pattern* of results was similar for the two language groups: the dynamic-dynamic contrast (rising vs. falling) was most easily discriminated, followed by static-static (e.g., low vs. high) then static-dynamic (e.g., low vs. rising). Burnham et al. (1992) suggested that the most relevant acoustic dimension appeared to be initial F_0 level although they conducted no formal acoustic analyses. It will be of interest in the study reported here to investigate whether differences found by Burnham et al. (1992) between Thai- and English-speaking adults are also obtained for children with these language backgrounds; whether acoustic analyses support the relative salience of initial F_0 as a perceptual cue; and most importantly whether the same acoustic dimensions are important perceptual cues for native Thai speakers, and English speakers with no experience of tone languages.

The Development of Tone Perception

Recently event-related brain potential studies using the mismatch negativity (MMN) paradigm have been conducted in order to investigate the neural representations of auditory events. These have shown that adult humans are sensitive to *relative* frequency patterns of sine wave tones irrespective of the absolute frequency of tones (Saarinen, Paavilainen, Schröger, Tervaniemi, & Näätänen, 1992), and that change in frequency over time and spectral composition share equal status at the neural level (Schröger, 1994). Insofar as this basic neural representation may reflect patterns of ontogenetic priority, it would seem that the acquisition of linguistic tone should at least parallel that of linguistic segments. However, studies of both production and perception suggest that tone acquisition may actually *precede* the acquisition of segments.

Clumeck (1980) reports that tone-language infants use the pitch of their utterances to convey affect earlier (around 8 months) than they use either pitch or segmental differences at a lexical level (Tuaycharoen, 1977; Luksaneeyanawin, 1976). Once lexical-level distinctions begin the use of pitch and segmental differences coincide, each first being evident around 11 months. Thereafter however, acquisition of tonal distinctions appears to be easier, because these are completely acquired by around 23 months, before the acquisition of segments is completed (Tuaycharoen, 1977; Clumeck, 1980). There is also anecdotal evidence from a 10- to 11-month-old child showing that pitch is more perceptually salient than segmental information (Li & Thompson, 1977). Together this evidence suggests that pitch differences may be more perceptually basic than segmental differences. However, no consideration has yet been given to the fact that in tonal languages the absolute number of segments far outweighs the number of tones (maximum of the latter is six: Goldsmith, 1994), so it may simply be the case that there are fewer tone distinctions to learn. On the other hand, Ioup and Tansomboon (1987) have suggested that tones may enjoy ontogenetic precedence as a by-product of infants' selective attention to the prosodic features of language (Fernald, 1989, 1993) and that young children may be relatively more sensitive to tonal features of a new language than are their adult counterparts. If pitch is a perceptually salient dimension in childhood then in the current study it might be expected that Australian, as well as Thai children will find tonal distinctions relatively easy to learn.

With regard to the order of acquisition of particular tones in tonal languages, static tones are generally found to be produced contrastively earlier than dynamic tones in Mandarin (Li & Thompson, 1977), Cantonese (Tse, 1978), and Thai (Tuaycharoen, 1977). With regard to dynamic tones, it has been suggested by Li and Thompson (1977) that falling tones require less physiological effort to produce than rising tones. This is reflected in the order of emergence of tones produced by infants learning Mandarin (Li & Thompson, 1977). However, in the case of Thai, Tuaycharoen (1977) found that the rising tone was produced before the falling tone. Perhaps such inconsistencies in order of production between languages reflect the degree of variation between the tonal systems of different languages. Moreover, as tones are perceived less categorically than consonantal segments (Abramson, 1979), it is more difficult to compare two tones, say two "falling" tones, across languages than it is to compare two consonants, say two voiceless bilabial stops.

Turning to the perception of tones, we find that very few developmental studies have been conducted. Clumeck (1980) reports that Mandarin-speaking children first discriminate rising and falling tones and suggests that order of acquisition of perceptual competence with particular tonal contrasts is based primarily on the phonetic distinctiveness of the tone pair, and on the degree to which the tones are distinct despite variations resulting from tone sandhi rules. In the study reported here the phonetic similarity of Thai tones can be gauged by testing English-speaking adults and children, for whom the tones are phonologically irrelevant and Thai-speaking adults and children for whom the tones are an integral part of their linguistic system. It is possible that the way in which subjects from these two different backgrounds process tonal distinctions may differ, and it is to a consideration of modes of processing that we now turn.

Modes of Processing in Tone Perception

Werker (Werker & Logan, 1985) found that a 500 msec interstimulus interval (ISI) between two to-be-discriminated sounds tends to induce a language-general phonetic mode of perception in which phones are perceived directly without any influence of linguistic experience. In such a mode even non-native listeners should be able to perceive contrasts which are not phonologically relevant in their language. In contrast, an ISI of 1500 msec was found to induce a language-specific or phonological mode of perception. In this mode perception is constrained, and perhaps enhanced, by experience with the phonological categories of a particular language. Burnham et al. (1992) investigated these two modes of perception with Thai- and English-speaking adults. They found that English speakers' perception of Thai speech contrasts (including tonal contrasts) was generally better at 500 msec than at 1500 msec ISI, while the reverse was true for Thais. Thus a method of testing which allows a phonetic mode of perception is more beneficial when the sounds to be discriminated are linguistically irrelevant to the listener, while a method which allows the filtering of irrelevant phonetic and acoustic information and subsequent classification into phonologically relevant categories is more beneficial for distinctions which are linguistically relevant.

Many studies have shown that linguistic experience systematically biases speech perception abilities towards the phonological distinctions present in the ambient language (Burnham et al., 1991; Werker & Tees, 1983), such that adults have more restricted perceptual abilities than infants in the same linguistic environment. In this regard it will be of interest to determine whether the phonological (1500 msec) advantage for Thais and the phonetic (500 msec) advantage for English speakers (Burnham et al., 1992) is accentuated as age increases. By investigating Thai- and English-speaking children's and adults' perception of Thai tonal contrasts at two ISI levels, and by examining the acoustic data, the perceptual data and the relationship between these, such issues in the development of tone perception should be elucidated.

METHOD

Subjects and Design

A total of 192 subjects were tested. There were 96 native Australian English speakers, each with little or no experience of other languages (e.g., one or two years of high school), and no experience with Thai or any tonal language. There were 96 native Thai speakers, who could also speak English with varying degrees of proficiency. In each language group, there were 24 adults, 24 8-year-olds, 24 6-year-olds, and 24 4-year-olds. A native language (Thai/English) x Age (Adult, 8, 6, 4 years) x ISI (500/1500 msec) design was employed. Half the subjects in each language group were tested with an ISI of 500 msec and half with a 1500 msec ISI. In each language x ISI subgroup, half the subjects were males and half females.

Stimulus Materials and Apparatus

Tonal speech tokens were produced by a native Thai female linguist. The prevoiced syllable, [ba:], was used to carry the five tones (high, mid, low, rising, falling), and five exemplars of each were used. These were digitised and stored on computer. All possible pairings of the five tones were used, resulting in a total of 10 tone contrasts. One of the five exemplars of each sound was chosen at random for presentation in each trial. This exemplar variation over trials was employed to encourage linguistic processing, in which the phonetic features are important for perceptual discrimination, as opposed to acoustic processing, in which idiosyncratic features of particular tokens may become important cues for discrimination.

The experiment was conducted using a laboratory-based system at the University of NSW—an IBM AT type computer; and a portable system at Chulalongkorn University—a Toshiba 3100e AT laptop computer modified to accommodate D-A, digital I/O, and filter boards. The computers stored the sounds on disk, controlled presentation and timing of the sounds, and recorded subjects' responses and reaction times for each trial. A response panel attached to the computers contained a "same" key and a "different" key for subjects' responses, and a set of coloured lights which flashed when subjects made a correct response.

Procedure

Equal numbers of same phone pairs (AA or BB) and different phone pairs (AB or BA) were randomly presented to the listener, whose task it was to respond as quickly as possible by pressing either the "same" or "different" key. There were two phases in the experiment. The first was a task competence phase in which subjects were required to respond correctly on seven out of the eight trials on a simple auditory distinction [dʒoːn] vs [ləː]. There were four same, and four different trials in this phase. This task-competence phase was included to ensure that each subject could make simple auditory discriminations, and was acquainted with the task and the apparatus. The test phase followed. Adults were presented with 3 blocks of 16 trials and children 1 block of 16 trials. In each block there were four tone contrasts, one "static-static" contrast (mid/high, mid/low, or low/high), two "static-dynamic" contrasts (mid/rising, mid/falling, low/rising, low/falling, high/rising, or high/falling), and one "dynamic-dynamic" contrast (rising/falling). Each contrast was presented in each of the four possible combinations: AA, AB, BA, BB (i.e., two same and two different trials). Thus in each block, there were 8 same and 8 different trials presented in random order.

Dependent Variables

The number of correct and incorrect responses and the accompanying reaction times were recorded on disk. From these, two dependent variables were derived. The first was a discrimination index (DI) given by [number of correct responses on different trials minus number of incorrect responses on same trials]/number of trials. The resulting score is a measure of how well subjects were able to discriminate speech sounds on AB and BA trials. A score of 1 would indicate that all AB and BA trials were responded to as "different" and all AA and BB trials as "same." A score of zero would indicate that the subject responded "different" on an equal number of AB/BA trials and AA/BB trials, i.e., that they showed no greater ability to respond "different" on different than on same trials. The second dependent variable was reaction time (RT) for all correct responses on AB and BA trials.

RESULTS AND DISCUSSION

Four aspects of the results are considered: 1) analyses of variance of subjects' responses to the three types of tone contrasts (static-static, static-dynamic, and dynamic-dynamic); 2) tone contours and acoustic data of the five stimulus tones used in the experiment; 3) subjects' discrimination accuracy for each of the ten tone contrasts; and 4) correlations between subjects' accuracy and acoustic variables.

1. Tone Types—Analyses of Variance

The two measures of interest, discrimination indices (DIs) and reaction times on AB/BA trials (RTs) were analysed in separate but identical analyses of variance: 2 x 4 x 2 x (3), Language Background (Thai, English) x Age (Adult, 8 years, 6 years, 4 years) x ISI (500 msec, 1500 msec) x Contrast Type (Static-Static, S-S, Static-Dynamic, S-D, Dynamic-Dynamic, D-D), with repeated measures on the last factor.

Table 1. *Mean DIs for Thai- and English-speaking Adults and Children*

Age	Thai	English	Combined
4 year olds	.45	.02	.23
6 year olds	.90	.27	.58
8 year olds	.93	.31	.62
Children	.76	.20	.48
Adults	.97	.70	.84

Mean DIs for Thai and Australian adults and children are shown in Table 1. As would be expected Thai speakers performed significantly better than English speakers $F(1,176)=171.04$. Performance improved between childhood and adulthood $F(1,176)=68.96$, and this improvement was greater for English speakers ($\overline{X}_{adults}=.70$, $\overline{X}_{children}=.20$) than for Thais (.97 vs .76), $F(1,176)=11.23$. For children, irrespective of their language background, there was a significant improvement between 4 and 8 years, $F(1,176)=54.05$, and a significant levelling off between 6 and 8 years, $F(1,176)=11.61$.

With regard to the processing level, there was a significant interaction of ISI (500 msec vs 1500 msec) and language background, $F(1,176)=4.31$. As illustrated in Figure 1, there is a crossover effect: as expected all Thai speakers are slightly better at discriminating tones when forced to use a phonological level of processing at 1500 msec ISI, while all English speakers are much better at discriminating tones when allowed to use a phonetic level of processing. Understandably Thai speakers, for whom the tones are familiar linguistic elements, can better discriminate the tones when higher level processes are operating in long-term storage, while English speakers, for whom the tones are unfamiliar in a linguistic context, are better able to discriminate when short-term direct comparison processes can be engaged. These results confirm and extend similar results obtained by Burnham et al. (1992); we now know that both adults *and* children have a processing advantage at 1500 msec for linguistically familiar elements (Thai speakers) and at 500 msec for linguistically unfamiliar elements (English speakers).

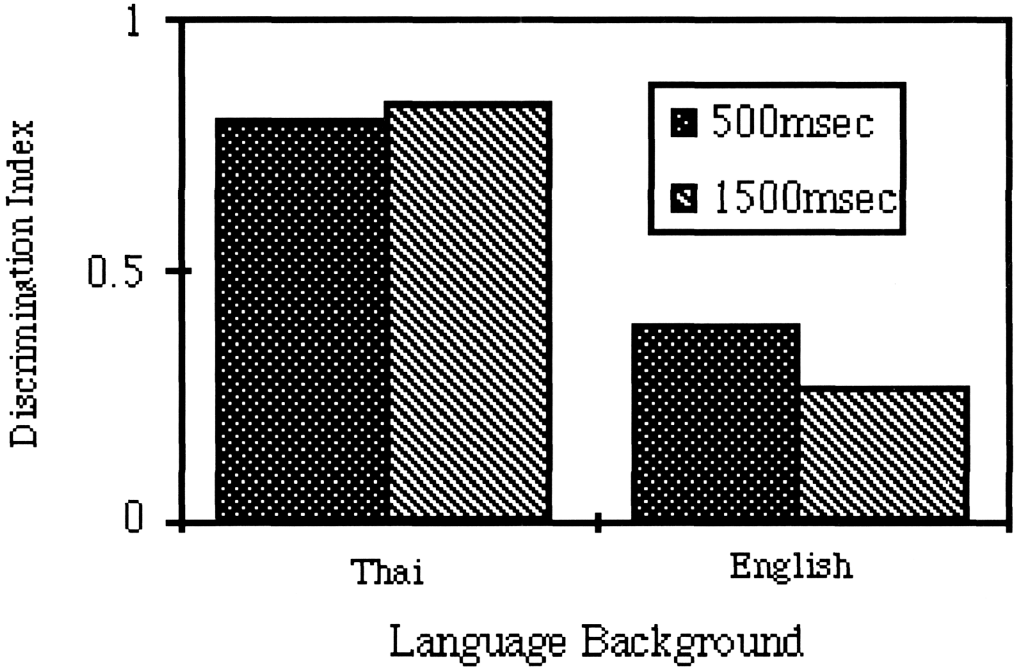

Figure 1. Interaction of Language Background and Processing Level

Performance differed for the three types of tone contrasts. All subjects were generally better at discriminating the D-D contrast ($\overline{X}$=.64) than the six S-D contrasts ($\overline{X}$=.55) or the three S-S contrasts ($\overline{X}$=.52), F(1,176)=11.20. This difference in performance is qualified by an interaction with ISI, F(1,176)=7.06. As shown in Figure 2, there are negligible differences between performance at 500 msec and 1500 msec ISI for the S-S and the S-D contrasts. However, for the D-D contrast performance is far superior at the 500 msec ISI. This suggests that the D-D (rising-falling) contrast is *acoustically* more salient than either the S-S or S-D contrasts. Presumably, this salient acoustic information is shed at higher levels of processing and tones are classified into appropriately "labelled" categories. It is interesting to note that this result is equivalent for English and Thai speakers, implying that English speakers have the propensity for classifying tone segments into phonologically-relevant categories. This propensity does not appear to be readily available for non-native *consonant* contrasts (Burnham et al., 1992), which suggests that tones may enjoy some advantage in linguistic processing. Such a conclusion would be consistent with evidence cited earlier that, in language acquisition, tonal segments are learned more readily than consonants and vowels (Clumeck, 1980).

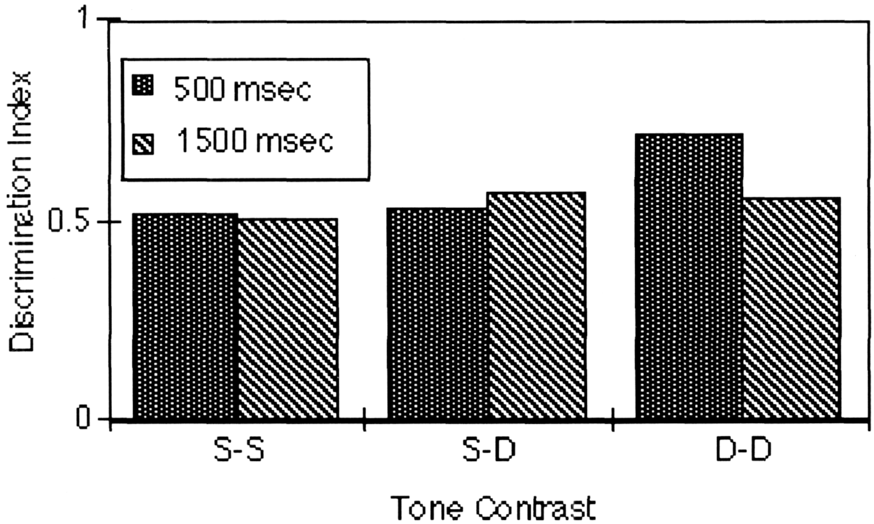

Figure 2. Interaction of Tone Contrast Type and Processing Level

Correct RTs on AB and BA trials were generally faster for the D-D contrast ($\overline{X}$=1952msec) than the S-D ($\overline{X}$=2065msec) or S-S ($\overline{X}$=2167msec) contrasts, $F(1,125)=13.57$, and this D-D superiority occurred irrespective of ISI, language background, and age of the listener. There were, however, differences between subjects when all three types of contrasts were considered together. RTs for Thai- and English-speaking subjects of each age are shown in Table 2. English-speaking subjects generally made quicker responses than Thais $F(1,125)=8.48$, and adults were faster than children $F(1,125)=79.60$.

Table 2. *Mean RTs for Thai- and English-speaking Adults and Children (msec)*

Age	Thai	English	Combined
4-year-olds	1961.7	1417.1	1836.0
6-year-olds	1823.7	1902.1	1846.8
8-year-olds	1744.4	1602.7	1699.9
Children	1836.3	1672.3	1619.2
Adults	1491.3	1267.3	1384.2

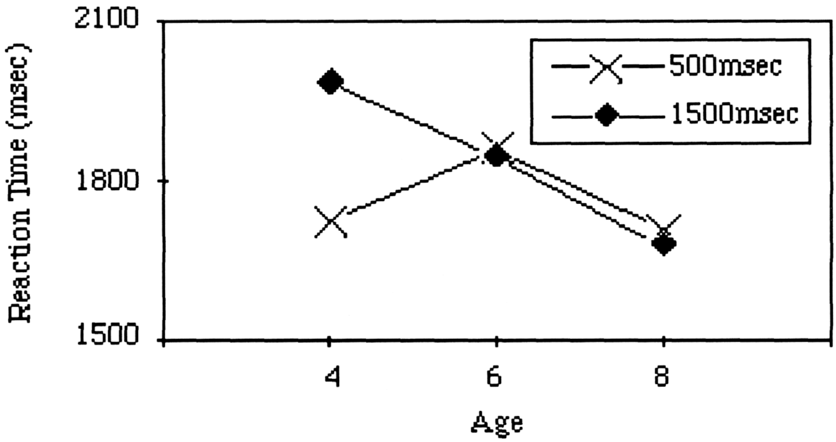

Figure 3. Interaction between Improvement over Age in Children and Processing Level

For the children there was a significant interaction between the linear improvement of RTs over age and performance at the two ISI levels, $F(1,125)=9.87$. The basis of the interaction is shown in Figure 3. The 4-year-olds were faster to respond to tone differences at the 500 msec than at 1500 msec, suggesting that they process tonal differences more efficiently when allowed access to phonetic information. Conversely, older children do not show this 500 msec advantage in reaction time: they are equally fast in processing tonal differences when they have access to phonetic information (at 500 msec ISI) and when using phonological (classificatory) information (at 1500 msec ISI). Moreover, across the three child age groups there is essentially no improvement in RT for discrimination at 500 msec ISI but a linear improvement in RT for discrimination at 1500 msec ISI. This suggests that for both Thai- and English-speaking children, age-related development of tone perception involves the increasing use of phonologically-relevant dimensions and *no* increase in attention to phonetically or acoustically related dimensions. The phonological mode of processing in which tone contours are classified in terms of phonologically relevant dimensions appears to gain precedence over the phonetic mode sometime between the ages of 4 and 6 years. This age-related improvement in the ability to classify tones in phonologically relevant ways will be noted again when correlations are examined in section 4.

2. Acoustic Statistics for the Five Tones

The five exemplars of each of the five tones were analysed using the Kay CSL system which employs a pitch-synchronous analysis to extract F_0. From the resultant values mean frequency plots of the five tones were produced, and these are shown in Figure 4. Relevant statistics for each of the tones are shown in Table 3, and in Table 4 differences between these statistics for each of the ten tone contrasts are shown.

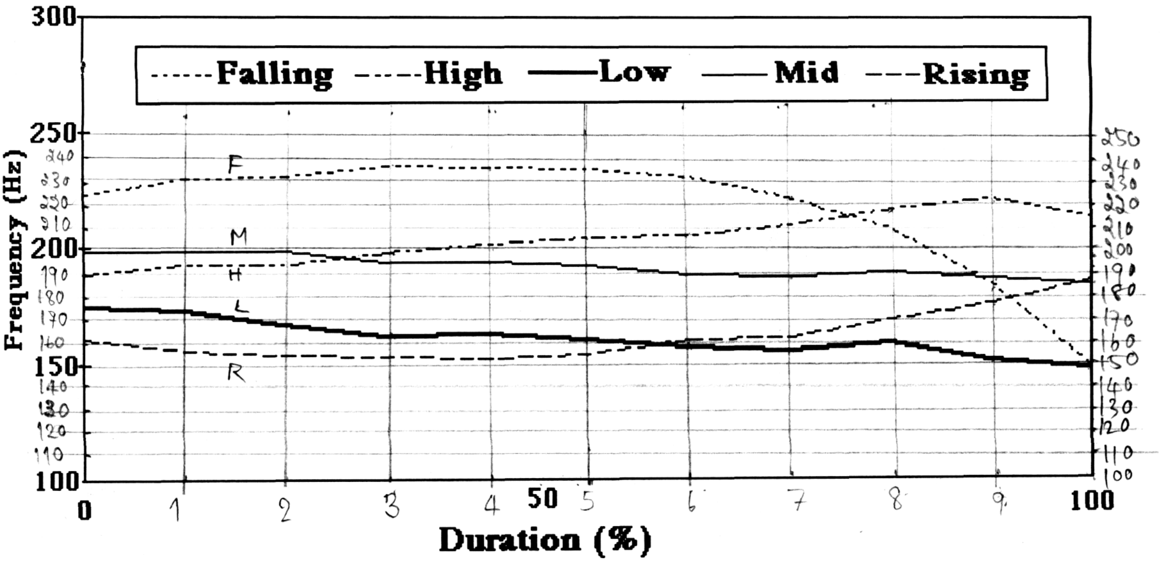

Figure 4. Mean Frequency Plots for the Five Exemplars of each of the Five Thai Tones

The two dynamic tones, falling and rising, show the greatest F_0 movement from onset to offset (-75.4 Hz and +24.6 Hz respectively). The static tones show less contour but there is still some movement especially for the low (-26.8 Hz) and high (+23.2 Hz) tones. While such onset-offset differences may appear large for supposedly 'static' tones, there is, for example, much less F_0 movement for the static low tone than for the dynamic falling tone. Thus there is support here for Abramson's (1978, 1986) dichotomous classification of Thai tones as static and dynamic.

Table 3. *Mean Duration (msec) and Acoustic Variables (Hz) for the Five Thai Tones*

	Mid	Low	High	Falling	Rising
Duration	519.2	480.0	460.8	489.4	453.2
Frequency	190.9	159.4	203.8	219.6	160.3
Standard Deviation	5.7	9.0	10.4	23.7	10.1
Min. Frequency	179.6	140.4	185.0	145.8	145.0
Max. Frequency	204.0	180.6	227.0	244.8	189.4
Range	24.4	40.2	42.0	99.0	44.4
F_0 Onset	198.4	173.8	188.6	223.0	160.2
F_0 Offset	182.8	147.0	211.8	147.6	184.8
Onset to Offset	-15.6	-26.8	23.2	-75.4	24.6

The mean F_0 levels of the tones are relatively distinct (Table 3) with falling showing the highest mean F_0 ($\overline{X}$=219.6 Hz) followed by high, mid, and then rising and low. The difference in mean F_0 between rising ($\overline{X}$=160.3 Hz) and low ($\overline{X}$=159.4 Hz) is negligible. Presumably the opposite direction of their contours and the difference in onset frequency provide sufficient cues for their discrimination. Onset values are generally quite different for each tone (Table 3), but there is some similarity of onset for mid and high tones. In this case the slow downward movement of the mid tone and the relatively quicker upward movement of the high tone may provide distinctive perceptual cues.

Table 4. *Acoustic Differences (in Hz) for the Ten Tone Contrasts*

		Average F0	Standard Deviation	F0 Range	F0 Onset	F0 Offset	On-Off Change	Absolute Change
S-S	High-Low	44.4	1.4	1.8	14.8	64.8	50.0	3.6
	High-Mid	12.9	4.7	17.6	9.8	29.0	38.8	7.6
	Mid-Low	31.5	3.3	15.8	24.6	35.8	11.2	11.2
S-D	High-Rising	43.5	0.3	2.4	28.4	27.0	1.4	1.4
	High-Falling	15.8	13.3	57.0	34.4	64.2	98.6	52.2
	Mid-Rising	30.6	4.4	20.0	38.2	2.0	40.2	9.0
	Mid-Falling	28.7	18.0	74.6	24.6	35.2	59.8	59.8
	Low-Rising	0.9	1.1	4.2	13.6	37.8	51.4	2.2
	Low-Falling	60.0	14.7	58.8	49.2	0.6	48.6	48.6
D-D	Rising-Falling	59.3	13.6	54.6	62.8	37.2	100.0	50.8

3. Perceptual Salience of the Ten Tone Contrasts

Table 5. *Accuracy (Percentage Correct and Rank Orders) for the Ten Tone Contrasts - (a) Thai and (b) English Speakers*

(a) Thai	Adults	8-year-olds	6-year-olds	4-year-olds
S-S High-Low	100.0 (1)	100.0 (1)	91.7 (8)	75.0 (4)
High-Mid	100.0 (1)	100.0 (1)	100.0 (1)	68.8 (7)
Mid-Low	97.9 (8)	100.0 (1)	95.0 (7)	58.3 (9)
S-D High-Rising	100.0 (1)	95.0 (8)	100.0 (1)	75.0 (4)
High-Falling	100.0 (1)	100.0 (1)	100.0 (1)	50.0 (10)
Mid-Rising	100.0 (1)	100.0 (1)	85.7 (9)	64.3 (8)
Mid-Falling	97.9 (8)	100.0 (1)	95.5 (6)	69.2 (6)
Low-Rising	97.9 (8)	80.0 (10)	83.3 (10)	78.6 (2)
Low-Falling	100.0 (1)	95.0 (8)	100.0 (1)	77.8 (3)
D-D Rising-Falling	100.0 (1)	100.0 (1)	95.8 (5)	81.3 (1)
(b) English				
S-S High-Low	93.8 (1)	12.5 (10)	16.7 (10)	25.0 (10)
High-Mid	81.3 (6)	25.0 (8)	54.2 (4)	37.5 (7)
Mid-Low	79.2 (7)	62.5 (4)	41.7 (6)	43.8 (5)
S-D High-Rising	87.5 (4)	75.0 (1)	60.0 (3)	55.0 (3)
High-Falling	75.0 (9)	16.7 (9)	37.5 (8)	50.0 (4)
Mid-Rising	79.2 (7)	50.0 (6)	66.7 (1)	70.0 (1)
Mid-Falling	87.5 (4)	43.3 (7)	38.9 (7)	27.3 (9)
Low-Rising	54.2 (10)	55.6 (5)	50.0 (5)	38.9 (6)
Low-Falling	93.8 (1)	66.7 (2)	28.6 (9)	57.1 (2)
D-D Rising-Falling	91.0 (3)	64.6 (3)	66.7 (1)	37.5 (7)

While the comparative statistics in Tables 3 and 4 show interesting physical differences between the ten tone contrasts, these need to be supplemented by perceptual data. In Table 5 the accuracy of Thai- and English-speaking adults and children on each of the ten contrasts is shown. To facilitate comparison, the rank order from easiest to most difficult is provided for each age group. Note that where two or more scores tie, the subsequent scores are ranked from the next available rank. For both language groups there are particular tonal contrasts which produce strikingly different performances across age. Consideration of some of these may help to elucidate variables which are important in the development of tone perception.

Mid-Low Contrast.

Abramson (1986) found that of the five tones, Thai adults most easily confused the mid and low tones. The results obtained here for Thai adults (accuracy=97.9%, equal lowest rank) and English-speaking adults (accuracy=79.2%, equal second lowest rank) attest to the relative difficulty of this contrast. However, especially for the English listeners, the children's performance on this contrast is relatively better (ranks of 4, 6, & 5 for 8-, 6- & 4-year-olds), which suggests that the factors important in adults' tone perception may not be the same as those important in children's tone perception. In the examples which follow some attempt to identify these factors will be made.

Low-Rising Contrast

The low-rising contrast is difficult for Thai adults, 8-year-olds and 6-year-olds (ranks of 8, 10, & 10) but relatively easy for Thai 4-year-olds (rank=2). For the English speakers low-rising is the most difficult contrast for the adults but of only moderate difficulty for children. As the mean frequency of the low and rising tones differs only by 0.9 Hz (see Table 4) it might appear that older listeners rely more on mean frequency and younger listeners more on tone contour.

High-Falling

Contrast. Thai adults, 8-year-olds, and 6-year-olds discriminate the high and falling tones perfectly but the 4-year-olds only discriminate 50% correctly, this being their worst performance. As both the mean F_0 difference ($\overline{X}$=15.8 Hz) and the onset difference (($\overline{X}$=34.4 Hz) are relatively large for this contrast (see Table 4), it may be that older subjects respond more to these two cues than do the 4-year-olds.

High-Low Contrast

English-speaking adults discriminate the high and low tones very well: at 93.8% correct it has the equal highest rank. Conversely, it is the most difficult (rank=10) for 8-, 6-, and 4-year-olds, (percent correct=12.5, 16.7, and 25.0 respectively). The high and low tones have a relatively large difference in mean frequency, but their differences in onset frequency and absolute movement are less distinct. As for the Thai subjects' perception of the high-falling contrast, this may indicate greater reliance on mean frequency differences by older listeners, and more reliance on a configuration of cues including tone contour by younger listeners. This possibility will be pursued further in the next section.

4. Correlations between Acoustic Differences and Perceptual Salience

In order to tease out which acoustic variables are perceptually salient at each age, Pearson correlation coefficients were calculated between accuracy on each of the ten tone contrasts (Table 5) and differences between four acoustic variables for each contrast (Table 4). Thus ten pairs of values for each tone contrast contributed to each of the correlation coefficients. This is not a large number of pairs for a correlation coefficient and the critical r value for significance is quite high (r_{crit}=.63). Nevertheless, comparison of these correlation coefficients across age may prove enlightening.

The correlations are shown in Table 6. Thai speakers have a relatively high correlation for mean frequency at all ages, while the correlation for onset steadily increases from 4 years to adulthood. For Thai children, other cues such as offset and standard deviation show moderate correlations which are not present in the adult data. Clearly, Thai children rely on a configuration of cues, while adults rely heavily on just two: mean onset and frequency. This limited cue usage is presumably very efficient, as Thai adults discriminate tones at near perfect levels.

Table 6. *Correlation Coefficients of Four Acoustic Differences with Accuracy for Thai and English Speakers*

	Adults	8-year-olds	6-year-olds	4-year-olds
Thai				
Frequency	.44	.41	.36	.43
Onset	.38	.25	.22	.17
Offset	-.09	.09	.07	-.27
Standard Deviation	-.00	.33	.41	-.09
English				
Frequency	.85	.40	-.09	.12
Onset	.42	.47	.30	.41
Offset	-.12	-.64	-.42	-.66
Standard Deviation	.35	-.02	-.14	-.09

English speakers show a different pattern of development. Most striking is the pattern of correlations for mean frequency. The 4- and 6-year-olds show virtually no correlation for mean frequency, by 8 years there is a moderate correlation, and by adulthood the correlation is extremely high. There is a consistent correlation for onset in all English-speaking subjects, and while children use offset but not standard deviation, adults appear to consider standard deviation but not offset.

Overall it appears that Thai speakers use mean frequency at all ages, steadily increase their use of onset, and when young consider cues that adults have learned to disregard. In contrast, English speakers use a complex set of acoustic cues at all ages. They seem unable to use mean frequency while young, but by the time they are adults, they rely perhaps too heavily on mean frequency. English-speaking adults continue to use a variety of cues rather than the optimum configuration achieved by Thai adults. Perhaps this can only be achieved when surrounded by a language in which lexical tone is phonologically-relevant. It appears that for those whose ambient language does not use tone to signal semantic differences at the lexical level (English speakers), the necessary developmental exposure and subsequent fine tuning does not occur.

These data support Gandour's (1983) findings that mean pitch or pitch height is an important variable in tone discrimination by adults, both tone language speakers and non-tone language speakers. The data also provide some support for the perceptual salience of onset pitch found by Saravari and Imai (1983) for Thai adults. However, neither Gandour (1983) nor Saravari and Imai (1983) studied children's tone perception. Here we have found that young listeners, irrespective of their language background, rely less on either of these two single dimensions and more on a complex set of acoustic variables. It appears that the development of tone perception involves the gradual identification of a particular subset of cues which are the most useful and that this must occur in a developmental context: English-speaking adults identified the most useful variable (mean F_0) but could not learn the optimal combination of cues over the period of the experiment.

The results of these studies can only bear on lexical pitch: English speakers do, of course, understand and produce pitch variations across sentences, so in future studies it would be of interest to investigate Thai and English speakers' relative ability to process sentential pitch variation. Moreover, English speakers continue to peceive pitch information in non-linguistic settings, so it would be of interest to investigate the influece of context, e.g., linguistic vs. musical, on the perception of pitch.

CONCLUSIONS

We conclude that in early childhood tone perception is acoustically-based, young listeners taking into account a variety of acoustic cues. This is shown in the developmental differences in correlation coefficients (Table 6) and in 4-year-olds' (both English and Thai) faster RTs for tone discrimination when the shorter ISI allows the more acoustically oriented phonetic mode of processing. In contrast, it appears that the perception of tones at a more phonological level involves the selection of one or two critical acoustic variables and almost exclusive reliance on these. This more phonological mode is evident in older Thai speakers, adults and 6- and 8-year-old children (see Table 6 & Figure 3). Perhaps for Thais the onset of schooling and formal language training at 5 years facilitates this shift to a more phonological mode of processing (Burnham, 1986; Burnham et al., 1991). The optimal phonological mode which emerges over age appears to involve greater concentration on the mean relative pitch (mean F_0 and onset values) and less concentration on the actual contour of the tones (onset, offset and standard deviation values). It will be of interest to investigate whether this developmental trend is the same in other tone languages.

The difference between speakers of tone and non-tone languages appears to support the importance of linguistic experience in the shift from phonetic to phonological processing. Thai adults and children perceive tones better in a phonological mode (1500 msec ISI), while English-speaking adults and children perceive tones better in a phonetic mode (500 msec ISI). English listeners rely on a complex set of acoustic variables to perceive tone differences and this strategy persists over age despite English-speaking adults' increasing use of mean frequency. In contrast, by adulthood Thai listeners have streamlined their use of acoustic variables such that only those which allow clear classification of tones into distinct phonological categories are used. Thus two aspects of linguistic experience are important in tone perception , age and language background. Over age, tone perception becomes more phonologically-based irrespective of language background. Moreover, even non-tone language adults are able to identify the most salient acoustic cue. However, it is only with appropriate linguistic experience that the optimal perceptual strategy develops, and only with appropriate linguistic experience that a truly phonological mode of tone perception is achieved.

REFERENCES

Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech, 21*, 319-325.

Abramson, A. (1979). The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Ohman (Eds.), *Frontiers of Speech Communication* (pp. 127-134). London: Academic Press.

Abramson, A.S. (1997). The Thai tonal space. In A.S. Abramson (Ed.), *Southeast Asian linguistic studies in honour of Vichin Panupong* (pp. 1–10). Bangkok: Chulalongkorn University Press. [This volume]

Burnham, D. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, 7, 206-240.

Burnham, D., Earnshaw, L., & Clark, J. (1991). Development of categorical identification of native and non-native bilabial stops: infants, children and adults. *Journal of Child Language, 18*, 231-260.

Burnham, D., Kirkwood, K., Luksaneeyanawin, S., & Pansottee, S. (1992). Perception of Central Thai tones and segments by Thai and Australian adults. *Pan-Asiatic Linguistics: Proceedings of the Third International Symposium on Language and Linguistics* (pp. 546-560). Bangkok: Chulalongkorn University Press.

Clumeck, H. (1980). The acquisition of tone. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.) *Child Phonology Vol 1: Production* (pp. 257-275). New York: Academic.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development, 60,* 1497-1510.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64,* 657-674.

Gandour, J. (1979). Perceptual dimensions of tone: Thai. In G.I. Liem (Ed.), *South-East Asian Linguistic Studies Vol 3* (pp. 277-300). ANU, Canberra: Dept. of Linguistics (Pacific Linguistics, Series C, No. 45).

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics, 11*, 49-175.

Goldsmith, M. (1994). Tone languages. In R. Asher & J. Simpson (Eds.), *The Encyclopedia of Language and Linguistics, vol 9* (pp. 4626-4628). Oxford: Pergamon Press.

Ioup, G., & Tansomboon, A. (1987). The acquisition of tone: A maturational perspective. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 333-349). Cambridge, MA: Newbury.

Li, C., & Thompson, S. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4*, 185-199.

Luksaneeyanawin, S. (1976). Some phonetic observations on the speech of Thai children. *Pasaa, 6*, 120-125.

Luksaneeyanawin, S. (forthcoming) Intonation in Thai. In D. Hirst & A. Di Cristo (Eds.), *A Survey of Intonation Systems.* Cambridge: Cambridge University Press.

Lynch, M., Eilers, R., & Bornstein, M. (1992). Speech, vision, and music perception: Windows on the ontogeny of mind. *Psychology of Music, 20,* 3-14.

Saarinen, J., Paavilainen, P., Schröger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the brain. *Neuroreport, 3,* 1149-1151.

Saravari, C., & Imai, S. (1983). Perception of tone and short-long judgements of vowel of variants of a Thai monosyllabic sound. *Journal of Phonetics, 11*, 231-242.

Schröger, E. (1994). An event-related potential study of sensory representations of unfamiliar tonal patterns. *Psychophysiology, 31,* 175-181.

Tse, J.K. (1978). Tone acquisition in Cantonese: A longitudinal case study. *Journal of Child Language, 5,* 191-204.

Tuaycharoen, P. (1977). *The phonetic and phonological development of a Thai baby: From early communicative interaction to speech.* Unpublished doctoral dissertation, University of London.

Werker, J., & Logan, J. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37,* 35-44.

Werker, J., & Tees, R. (1983). Developmental changes across children in the perception of nonnative speech sounds. *Canadian Journal of Psychology, 37,* 278-286.