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