The Development of Tone Perception: Cross-Linguistic Aspects and the Effect of Linguistic Context

Denis BURNHAM, Elizabeth FRANCIS, and Di WEBSTER

School of Psychology, University of NSW, Sydney, 2052, Australia

A considerable amount is now known about the development of speech perception with respect to consonants, and more recently vowels and prosody. However, much less is known about the development of tone perception. In the studies reported here the development of tone perception in both tonal and non-tonal language speakers is investigated. The results bear on both the development of tone perception specifically, and the development of speech perception generally. Three experiments were conducted, all testing perceptual discrimination with Consonant-Vowel (CV) syllables. In the first it was found, predictably, that Thai subjects of all the ages tested (4, 6, 8 years and adults) were much better at discriminating contrasts phonemic only in Thai than English-speaking subjects of corresponding ages. Most importantly it was found that English-speaking adults were better at discriminating tone contrasts than a non-native voicing contrast, while English-speaking children were better at the consonant contrast than the tones. In the second study English-speaking adults were tested for their discrimination of Thai tones carried on the syllable /ba/, or either sine-wave analogue, or musical (violin) equivalents of these. It was found that discrimination was better for the musical and the sine-wave analogues of tones than for the linguistic tones. A subsequent study with 5-, 6-, and 8-year-old English speakers yielded similar results. Together the studies show that learning a non-tonal language does not attenuate the perceptual ability to perceive pitch differences. Rather, a perceptual bias is set up such that in a linguistic (phonological) context, attention to pitch differences is attenuated. Results are discussed in terms of Cutler and Mehler's periodicity bias, Best's PAM, and Burnham's RAF model.

1. Introduction

A considerable amount is now known about the development of speech perception with respect to consonants, and more recently vowels and prosody. However, much less is known about the development of tone perception. In the studies reported here the development of tone perception in both tonal and non-tonal language speakers is investigated.

Twenty-five years of speech perception research shows that young often neonatal, infants can perceive just about any consonantal contrast, both native and non-native in their language environment, on which experimenters wish to test them (eg, Eimas, Siqueland, Jusczyk, & Vigorito; Streeter, 1976; Best, McRoberts, Sithole, 1988; for *reviews, see Burnham, 1986; Wode, 1992). Until recently far less work was conducted with vowel contrasts (Trehub, 1976), but it is now evident that infants also perceive and discriminate various vowel contrasts (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994; Polka & Bohn, submitted).* Thus young infants' *segmental* speech perception (of consonants and vowels) is present very early in life and appears to be based on universal principles - abilities are initially unconstrained by the language environment and extend to both native and non-native contrasts.

What of the *suprasegmental* aspects of speech? All languages involve prosodic variations - of fundamental frequency (F_0), amplitude, and duration - across sentences or utterances. At the word level there are two main methods by which languages use prosodic features to distinguish words: in lexical stress languages, such as English, there are patterns of stress across the syllables of a word; in lexical tone languages and grammatical tone languages the level or contour of F_0 serves a lexically distinctive function (Luksaneeyanawin, forthcoming; Cutler, 1989; Cutler & Chen, submitted; Demuth, 1993). Are infants' sensitive to the *suprasegmental* aspects of speech?

Newborn infants prefer their mother's voice to that of somebody else's mother (DeCasper & Fifer, 1980), prefer familiar passages (read consistently by the mother before birth) to unfamiliar passages (DeCasper & Spence, 1986), prefer the distinct contours of infant-directed speech to those of adult-directed speech (Panneton Cooper & Aslin, 1989), and prefer their native language to other languages (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, and Amiel-Tison, 1988). These preferences appear to be based on the relatively early maturity of the human fetal auditory system (Bredberg, 1985) and the availability of the low frequency components of human voices (especially the mother's) inside the womb (Querleu & Renard, 1981; Querleu, Renard, Versyp, Paris-Delrue & Crepin, 1988), because such preferences are not evident when this information is removed from the speech signal (Mehler, Bertoncini, Barriere, & Jassik-Gerschenfeld, 1978; Mehler et al., 1988). Moreover, newborn infants attend to syllabic (eg, *pat*) in preference to non-syllabic (eg, *pst*) utterances (Moon, Bever, & Fifer, 1992); to the number of syllables in an utterance, but not the number of phonemes (Bijeljac-Babic, Bertoncini & Mehler, 1993); and discriminate bisyllables containing a word boundary from segmentally-equivalent bisyllables which do not (Christophe, Dupoux, Bertoncini, & Mehler, 1994). Thus infants perceive and prefer certain aspects of speech prosody from birth.

Not too long after birth infants' speech perception abilities begin to be shaped by the ambient language environment. The specific language in the ambient speech environment appears first to affect prosody, then vowels, and finally consonants. With regard to prosody, infants become selectively responsive to native-language prosodic patterns in words (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993), and those which mark clausal boundaries (Hirsch-Pasek, Kemler-Nelson, Jusczyk, Wright-Cassidy, Druss, & Kennedy, 1987) by 6 months; and to the predominate stress patterns of words in the native language (Jusczyk, Cutler, & Redanz, 1993), and to the prosodic cues marking phrasal boundaries by 9 months (Jusczyk, Hirsch-Pasek, Kemler-Nelson, Kennedy, Woodward & Piwoz, 1992). With regard to vowels, it seems that perceptual re-organisation - attenuation of the perception of non-native vowels - may begin between 4 and 6 months[1] (Polka & Werker, 1994), and that prototypes of native vowels are in place by 6 months (Kuhl et al., 1992). Kuhl (eg Kuhl, 1994) describes these prototypes as language-specific "perceptual magnets" that assimilate neighbouring representations of vowels (though see Lacerda, 1995). More recent evidence suggests the corner vowels of the F_1 - F_2 space may initially act as perceptual magnets, ie, that there may be a default language-general bias (Polka & Bohn, submitted), it is suggested that these magnets are modified by perceptual experience in particular vowel environments (Polka, 1995; Polka, in press). Finally, numerous studies by Werker and her colleagues (Werker, & Tees, 1984a; Werker & Tees, 1983; Werker, Gilbert, Humphrey, & Tees, 1981; Werker, & Lalonde, 1988) show that between about 7 and 11 months there is extensive re-organisation of infants' speech perception resulting in a decline in the ability to perceive various non-native consonant contrasts.

Selective re-organisation of the perception of speech segments on the basis of the ambient phonetic environment appears to occur first for prosody and vowels and then for consonants. This pattern also appears to be reflected over the first year in shifts in infants' productions towards the ambient language (see Werker, 1993; and Cutler, 1994 for reviews). Cutler (1994; Cutler & Mehler, 1993), in fact, argues that infants come to the task of language perception armed with a "periodicity bias", a tendency to attend more to vowels, with their longer duration and marked periodic structure, than to consonants. At a more prosodic level, Cutler, Mehler, Norris, and Segui (1992) suggest that infants will attend to the smallest level of rhythmic regularity in the ambient language. If this periodicity bias in fact occurs, then infants should be especially tuned to intonation, rhythm, stress, and tone. The task of the language learner is to discover which of these are used in a regular fashion in the language around them and to attend to these regularities and ignore other more random prosodic variation. Thus the learner of a stress-timed language such as English should learn to attend to lexical stress, and disregard lexical tone, while the learner of a tone language should do the opposite (Cutler, 1994).

The loss of speech perception abilities that occurs for segments, at least consonants, is not a sensorineural loss. Werker and her colleagues have conducted a number of ingenious experiments showing that when testing conditions favour phonetic or acoustic processing (by reducing the interstimulus interval between sounds to be discriminated in an AX discrimination task to 500 or 250 msec respectively), adults can usually perceive non-native contrasts, while with a 1500 msec ISI, which forces reliance on long-term and thus phonemic processes, they cannot (Werker & Logan, 1985; Werker & Tees, 1984b).

Not all contrasts are lost in this early period, if at all. Burnham (1986) in a review and theoretical formulation, points to a number of contrasts which are "lost late", sometime after 4 years of age. These include the bilabial prevoiced vs voiceless unaspirated stops, [b]-[p] (Burnham, 1986; Burnham, Earnshaw & Clark, 1991), plosive vs implosive voiced bilabial, [b]-[ɓ], and alveolar, [d]-[ɗ] stops (Bond & Adamscu, 1979), and a suprasegmental stress cue used contrastively in some southern dialects in France (Allen, 1983). Burnham labels these "robust" contrasts and, contrasts them with what he calls "fragile" contrasts in a robust and fragile (RAF) model of speech perception development. Burnham claims that robust contrasts are lost late possibly due to their stronger psychoacoustic basis (eg, temporal cue rather than spectral cue basis), their greater universality as reflected in representation in the world's languages, and their tendency to be allophonically present in the target language even though they are not used contrastively. Moreover, Burnham (1986; see also Burnham, Earnshaw, and O'Connor, in preparation) presents evidence to show that the degree of loss of such non-native contrasts relative to comparable native contrasts, a variable he calls "phonological bias", is predicted by concurrent reading ability, and suggests that the loss may be a consequence of the controlled processes required for phone-to-grapheme mapping in early acquisition of reading. Perceptual ability with robust contrasts also tends to show some spontaneous resurgence around 8 years of age, and is more easily recovered in adulthood through retraining than is ability with fragile contrasts. In the RAF model it would seem that tonal contrasts should fall into to the robust category. First, there is the evidence that a stress contrast is lost late (Allen, 1983); second, tonal contrasts (lexical pitch variations) certainly occur allophonically even when they are not used to contrast meaning; and finally, tonal contrasts are based mainly on pitch, which bears a very close relationship with the highly salient psychoacoustic cue of the level and contour of F_0.

However this can not be the whole story, as Best has found that there are some non-native contrasts which appear never to be lost. Best, McRoberts, and Sithole (1988) found that English language environment (ELE) infants from 6 to 14 years of age, like their adult counterparts, discriminated Zulu click contrasts, which are not used in English. In addition, Best (1990) reports that 10- to 12-month-old infants and adults show good discrimination of the Ethiopian ejective [p']-[t'] contrast. Best (Best et al., 1988; Best, in press) puts forward the Perceptual Assimilation Model (PAM), in which she claims that non-native segments tend to be perceived according to their degree of similarity to native segment constellations close to them in native phonemic space. On this basis, Best has outlined a number of patterns of perceptual assimilation. For example, the Ethiopian ejectives [p'] and [t'] are perceived by ELE listeners to be highly similar to English [p^h], and [t^h], so a Two-Category (TC) assimilation pattern applies resulting in good discrimination. In the case of Zulu clicks, neither sound can be assimilated into ELE listener's native phonemic space, and adults report that they do not sound like speech. Consequently, these sounds can be discriminated by attending to acoustic differences, with discriminability depending on their psychoacoustic salience. Even under conditions designed to minimise acoustic processing (Werker & Logan, 1985; Werker & Tees, 1984b), an AXB discrimination task with 1000 msec ISI, Best et al. (1988) found excellent discrimination for these non-assimilable click contrasts. Best (in press) outlines various other assimilation patterns. For example, the consonant contrasts for which there is attenuation of ability in infancy tend to fall into one of two assimilation patterns: either the Single Category (SC) pattern, in which both non-native sounds are assimilated into the same native category, or the Category Goodness Difference (CG) pattern, in which both non-native sounds are assimilated into the same native category but differ in the degree of discrepancy from the native "ideal". Predictions based on these patterns have generally been upheld (Best, in press; Polka, 1995).[2]

According to Best's model, for speakers of non-tonal languages lexical tones should fall into the SC or, more likely the CG pattern. In the latter case the discriminability of tones will depend on how close each of the individual tones in the contrast are to the ideal tone for the segmental configuration on which they are carried in a particular linguistic context. In Best's model psychoacoustic factors should only play a role when both speech sounds are non-assimilable (NA). However, it is possible that even if tones are assimilated into a single category (SC), that tone perception by speakers of a non-tonal language depends upon the psychoacoustic salience of the difference between the F_0 height and movement of the particular tones involved (Burnham, 1986). Moreover, from the point of view of the RAF model, if tonal contrasts can be considered to be robust and similar to the stress contrast reported by Allen (1983), and if allophonic pitch variations keep tone "alive" in some sense, then tonal contrasts should be "lost late" (Burnham, 1986). Conversely, if there is a periodicity bias and young ELE infants are efficiently able to learn not to attend to prosodic variation which is irrelevant for learners of a stress timed language (Cutler, 1994), then the ability to perceive tonal contrasts should be attenuated early in development.

Clearly, theories of speech perception development and non-native speech perception have not given high priority to the position and ontogenetic status of lexical tone. The empirical studies described here concern the perception of central Thai tones by both Thai- speaking and English-speaking children and adults, and the results will be relevant to the development of tone perception, processes of speech perception development, and the psychoacoustic salience of tones and F_0 height and contour.

2. Relative Salience of Consonants and Tones over Development

2.1 Introduction

This study concerns the relative discriminability of Central Thai bilabial stop consonant contrasts, and tonal contrasts by Thai- and English-speaking adults and children at two levels of processing, phonetic, with a 500 msec ISI and phonemic, with a 1500 msec ISI (Werker & Logan, 1985; Werker & Tees, 1984b). As shown above many studies have shown that linguistic experience systematically biases speech perception abilities towards the phonological distinctions present in the ambient language. Using these same contrasts just with Thai- and English-speaking adults, Burnham, Kirkwood, Lusaneeyanawin, and Pansotte (1992) found, predictably, that both Thai and English speakers were better at discriminating the [b]-[p^h] contrast, and less predictably both were better at discriminating the tonal contrasts than the [b]-[p] contrast. English-speaking adults were generally better at 500msec than at 1500msec ISI while the reverse was true for Thais. It will be of interest to discover whether the same pattern of perceptual salience is obtained across development.

2.1.1 Thai Bilabial Stop Consonant Contrasts

Thai has three levels of voicing for bilabial stops realised as /b/, /p/, and /p^h/, the average voice onset time (VOT) of these being -97, 6, and 64 msec respectively; English has two, realised as /b/ and /p/, with VOTs of 1 and 58 msec, (Lisker & Abramson, 1964). In this experiment Thai productions of two consonantal contrasts will used: [b]-[p^h], which is phonemically relevant in both Thai and English, and [b]-[p], which is phonemically relevant only in Thai. Comparison of the perception of these by Thai and English children and adults will aid in the understanding of the role of linguistic experience in the development of speech perception.

2.1.2 Thai Tonal Contrasts

Central Thai is a lexical tone language with five tones - mid, high, low, rising, and falling (Luksaneeyanawin, forthcoming). In a multi-dimensional scaling analysis on Thai listeners' discrimination of various pitch contours Gandour (1979) found three tone-related dimensions - average pitch, which appears to be a relatively universal cue for tonal distinction (Gandour, 1983); direction, which served to distinguish between rising and falling contours; and slope, which served to distinguish between what Abramson (1978) has called static (mid, low, high) and dynamic (rising, falling) tones. Gandour's results agree to some extent with Abramson's studies of Thai subjects' identification of synthetic (Abramson, 1974, 1978, 1986) and natural (Abramson, 1975) tonal stimuli. Abramson (1978, 1986) 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. In a development study, Burnham and Francis (1995) found that English listeners rely on a complex set of acoustic variables to perceive tone differences and that this strategy persists over age. In contrast, by adulthood Thai listeners had streamlined their earlier use of a complex of acoustic variables in childhood to concentrate relatively more on the mean relative pitch and relatively less on the actual contour of the tones.

2.1.3 Relative Salience of Consonantal and Tonal Contrasts

Some theoretical predictions regarding the relative salience of tones and segments can be obtained from the literature. 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 and that young children may be relatively more sensitive to tonal features of a new language than are their adult counterparts. This notion is in part agreement with Cutler and Mehler's (1993) idea of a "periodicity bias", but the claim that children may be more sensitive to the tonal features of a new language conflicts with Cutler's (Cutler et al., 1992) notion that learning about the prosodic regularities of a language is something that one only does once.

What of the empirical evidence regarding the ontogeny of tonal and segmental perception? Some information about the relative salience of Thai consonantal and tonal contrasts can be gleaned from acquisition studies. 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)[3]. 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 lexical tone 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 less tone distinctions to learn.[4]

Turning from the acquisition literature, to experimental studies, the picture becomes less clear. Kuhl and Miller (1982) using synthetic stimuli report that 1- to 4-month-old infants can detect a vowel change in the presence of irrelevant pitch changes, while they could not do the opposite - detect pitch changes in the presence of irrelevant vowel changes. In a more naturalistic study Li and Thompson (1977) found that a 10- to 11-month-old child found tone to be more perceptually salient than segmental information in the perceptual understanding of words. So the evidence with infants is as yet unclear. In childhood, Burnham and Torstensson (1995) have found that 6-year-olds perceive non-native vowel contrasts more easily than tonal contrasts, while older children (10 and 14 years) and adults find discrimination and identification of tones easier. These results with adults conflict with those of Cutler and Chen (in press) who found that both Cantonese and Dutch adults find tonal distinctions relatively more difficult to discriminate than vowel distinctions. As can be seen, this experimental literature does not lead to a clear picture and systematic developmental studies are required.

2.2 Method

2.2.1 Subjects and Design

A total of 192 subjects were tested, 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, and 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 (500/1500msec) x Contrast Type ([b]-[pʰ], [b]-[p], tone contrasts) design with repeated measures on the last factor was employed. Half the subjects in each language group were tested with an ISI of 500msec and half with a 1500msec ISI. In each language x ISI subgroup, half the subjects were males and half females.

2.2.2 Stimulus Materials and Apparatus

Speech tokens, bilabial stop plus [a:] vowel tokens, [ba:], [pa:], [pʰa:] were recorded by a native Thai female linguist. For the consonant contrasts these were all presented with mid tone. For the tone contrasts the prevoiced consonant was used to carry the five tones: mid [ba:], low [bà:], falling [bâ:], high [bá:], rising [bǎ:]. Plots of these can be seen in Burnham & Francis (1995). For each of the seven consonant-vowel pairs, five exemplars were produced by the speaker, a total of 35 speech sounds. These were digitised and stored on disk of IBM-compatible computers. Three different Thai contrasts were tested in the experiment: [ba:] vs [pʰa:], a contrast which is phonologically relevant in English; [ba:] vs [pa:], which is phonologically irrelevant in English; and tone contrasts, e.g., [bâ:] vs [bá:], which again are irrelevant in English. All possible pairings of the five tones were used, 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 both 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.

2.2.3 Procedure

Equal numbers of same (AA or BB) and different 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 [dʒo:n vs lə:]. 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. Three test phases followed, with the order of these counterbalanced between subjects. For the adults there were 48 trials in each phase arranged in three blocks of 16. For the children there was one block of 16 trials in each phase. In each block of 16 there were eight same and eight different trials. The actual exemplars presented on any particular trial were selected randomly by the computer from the pool of five possible exemplars for each phone. At the end of each of the three phases the three response keys and the feedback lights flashed on and off to inform the subject that the phase had finished and that a new phase was about to begin.

In the tone phase the block structure was a little more complex. In each block there were four tone contrasts, one "static-static" (S-S) contrast (mid/high, mid/low, or low/high), two "static-dynamic" (S-D) contrasts (mid/rising, mid/falling, low/rising, low/falling, high/rising, or high/falling), and one "dynamic-dynamic" (D-D) 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 of 16, there were 8 same and 8 different trials presented in random order.

2.2.4 Dependent Variable

The number of correct and incorrect responses and the accompanying reaction times were recorded on disk. Only the correct and incorrect responses were used here. They were converted to 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.

2.3 Results

Discrimination indices were analysed by planned contrasts in a 2 x 4 x 2 x (3), Language Background (Thai, English) x Age (Adult, 8 years, 6 years, 4 years) x ISI (500msec, 1500msec) x Contrast Type ([b]-[p^h], [b]-[p], tone contrasts) analysis of variance, with repeated measures on the last factor. The critical F-value was $F_{.05}$ (1,176) = 3.9.

Mean DIs for the three contrasts for Thai and English speakers at each age are shown in Figure 1. Thai speakers (X=.777) generally discriminated better than did English speakers (X=.488), F (1,176)=124.51; adults (X=.793) generally better than children (X=.579), F (1,176)=50.97; and children generally improved linearly over age, F(1,176)=78.38, with some quadratic component, F (1,176)=6.86 (see Figure 1).

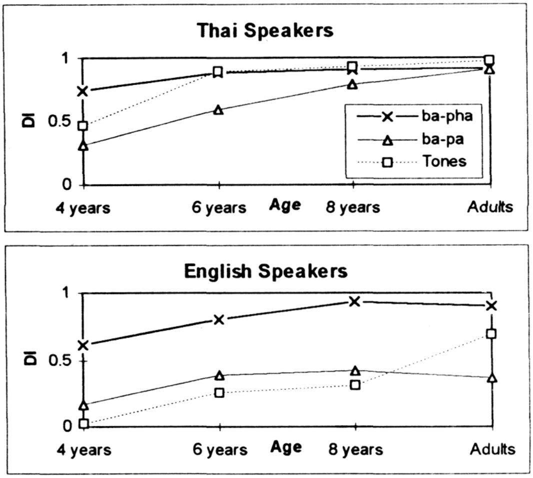

Figure 1 Mean Discrimination Indices for the three speech contrasts for Thai and English Speakers at each age

Of greatest interest are the interactions with the type of contrast. Two planned contrasts were tested, the difference between the two consonant speech contrasts, [ba]-[p^{h}a] vs [ba]-[pa], and the difference between the two contrasts that were non-native for the English speakers, [ba]-[pa] vs the tonal contrasts.

All speakers generally discriminated [ba]-[p^{h}a] (X=.839) better than [ba]-[pa] (X=.491), F (1,176)=167.44. The fact that this relative deficit for [ba]-[pa] was obtained, albeit reduced, even for adult Thais, suggests the involvement of some language-independent factor. Despite this overall effect, predictably, the difference in discriminability of the two consonantal contrasts was greater for the English than the Thai speakers, F (1,176)=25.06. Moreover, as can be seen in Figure 1, both across childhood, F (1,176)=6.23, and from childhood to adulthood, F (1,176)=7.57, performance on the two contrasts converged for Thai speakers, showing the effect of phonological experience, but remained parallel for the English speakers, for whom the second is phonologically-irrelevant.

Listeners generally discriminated tones (X=.568) better than the [ba]-[pa] (X=.491) contrast, F (1,176)=6.20. In itself this difference is not of great interest, as it is difficult to compare segmental and pitch-based contrasts, due to the lack of a common metric. This difference is of interest, however, when it interacts with other factors. There was a significant interaction with language background, F(1,176)=8.78, the difference being generally attenuated for English speakers, and with age, F (1,176)=5.39, the difference being generally greater for adults. Of greatest interest was a significant interaction of [ba]-[pa] with age (children vs adults) and language background. This can be seen in Figure 1, and is more clearly represented in Figure 2.

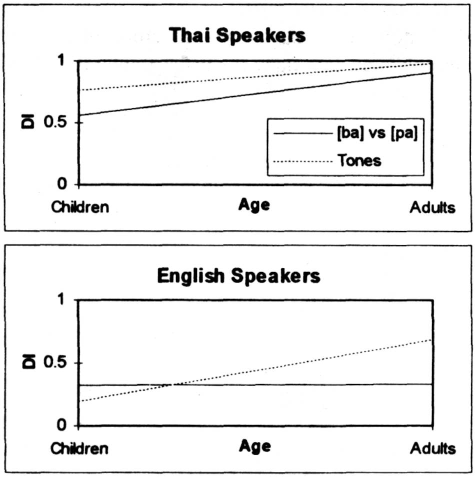

Figure 2. Interaction of performance on [ba]-[pa] vs tone, with age and language background.

As can be seen, for the Thais there is a relatively parallel development for [ba]-[pa] and tones between childhood and adulthood. However, for the English listeners, the children are better at discriminating the relatively difficult [ba]-[pa] contrast than the tones, while the reverse is true for adults. This can be taken to show that children have relatively more phonological bias (Burnham, 1986) for a feature which is not used to distinguish meaning in their language, lexical tone, than one that is, albeit not at the same levels, consonantal voicing. Presumably, the adults are able to treat the task as one of perceptual discrimination with lesser involvement of linguistic processes than are the children, which results in their relatively better performance on the salient pitch differences.

There was a marginally significant interaction of processing level (as reflected in performance at the different ISIs) and language background, $F(1,176)=3.74$ $p = .0547$. As would be expected, Thais generally discriminated better when processing phonemically (.786 vs .768), while English speakers generally discriminated better when able to access phonetic processing strategies (.510 vs .466). There was also an interaction of [ba]-[pa] vs tones x the linear age effect across childhood, language background, and ISI, which is shown in Figure 3. If the difference between DIs at 500 msec ISI and at 1500 msec ISI are taken to reflect the degree to which processing is more phonetic or language-general than phonemic or language-specific (Werker & Logan, 1985; Werker & Tees, 1984b), then it can be seen that, except for

[ba] vs [pa] at 4 years, Thais are generally processing both contrasts more phonemically. On the other hand, English children are generally better at discriminating these two non-native contrasts when phonetic processes are available. For tones, this becomes especially strong at 8 years, which may reflect an attentionally-based phonological bias against a feature which is not used to contrast meaning in the phonology and especially the orthography of the child's language.

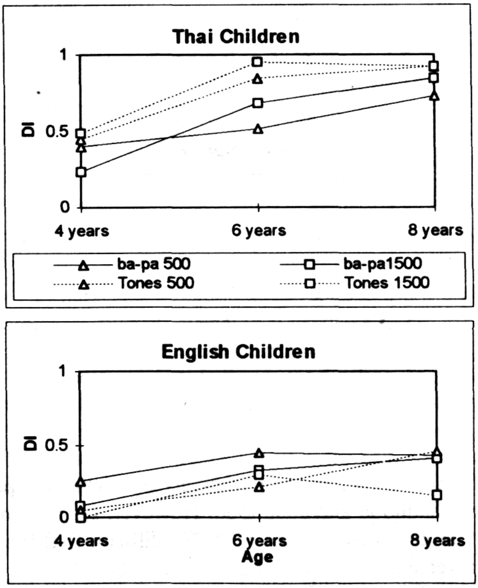

Figure 3. Interaction of performance on [ba]-[pa] vs tones, with improvement over age, language background, and ISI.

An additional analysis of variance, 2 x 4 x 2 x (3), Language Background (Thai, English) x Age (Adult, 8 years, 6 years, 4 years) x ISI (500msec, 1500msec) x Tone Contrast Type (Static-Static, Static-Dynamic, Dynamic-Dynamic) was conducted separately on the tones, to investigate the relative discriminability of tone types. 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), $F(1,176)=11.20$. These differences are qualified by an interaction with ISI, $F(1,176)=7.06$. As shown in Figure 4 there are negligible differences between performance at 500msec and 1500msec ISI for the S-S and the S-D contrasts, but for the D-D contrast performance is far superior at 500msec.This suggests that the D-D (rising-falling) contrast is acoustically more salient than either the S-S or S-D contrasts.

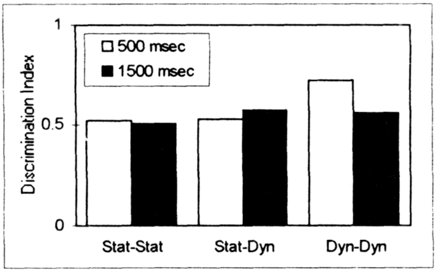

Figure 4. Interaction of tone contrast type and processing level

2.4 Discussion

All subjects discriminated the [ba]-[p^{h}a] contrast better than the [ba]-[pa] contrast. This is understandable for the English speakers, for whom the latter is phonemically-irrelevant, but for the Thais this result suggests that [ba]-[pa] is difficult for psychoacoustic reasons. Pisoni (1977) suggests that there is more effective masking of a high frequency component (such as the release) by a preceding low frequency component (such as prevoicing), as in [b], than of a low frequency component by a preceding high frequency component, as in [p^h]. This non-linguistic interpretation is supported by the fact that infants with limited linguistic experience show discrimination (Aslin et al., 1981) and identification (Burnham, Earnshaw & Clark, 1991) superiority for [pa]-[p^{h}a] over [ba]-[pa]. Despite this general superiority of [ba]-[p^{h}a] over [ba]-[pa], there was an effect of linguistic experience on the Thais' perception: while for the English speakers the curves for phonemic [ba]-[p^{h}a] and non-phonemic [ba]-[pa] remain parallel over age, for the Thais the two curves converge over age. This convergence for Thais shows the attunement effect of linguistic experience, despite the relative perceptual difficulty of [ba] vs [pa].

Predictably, the Thais discriminated the phonemically-relevant [ba]-[pa] and the tones contrasts generally better than the English speakers for whom these contrasts are irrelevant. The most striking and interesting aspect of the results concerns the difference for these two contrasts for the English speaking children and adults. Results for the adults and children will be considered in turn. Like the Thai adults, the English-speaking adults better discriminated the tonal contrasts than the consonantal contrasts. As both contrasts are relevant for the Thais and both irrelevant for the English-speakers, this suggests that for some reason the tonal contrasts are more perceptually salient than the prevoiced vs voiced unaspirated consonantal contrast. It is interesting that this is exactly the opposite result to that obtained by Cutler and Chen (in press), who found that both Cantonese and Dutch speakers discriminated Cantonese vowel contrasts better than they did Cantonese tonal contrasts. There are at least four possible causes for these findings. First, in their same-different discrimination studies Cutler and Chen used a restricted range of two tone contrasts while all 10 possible tone contrasts were used here. This may have made adult subjects here, both Thais and English speakers alike, more sensitive to tonal distinctions in this experimental procedure. Secondly, Cutler and Chen provide subjects with a 250 msec ISI, while subjects had either 500 or 1500 ISI here. As Cutler and Chen point out in an explanation of their results, information from lexical tone is extended over time and so arrives relatively late. Thus it is possible that a certain minimum period of time is required to process the time-extended tonal information and that the longer ISIs provided in our study allowed this. Thirdly, the scores which we report are discrimination indices which take into account both hits and false positives on different trials. Cutler and Chen report raw error rates. It is not clear without more detailed investigation how these two different approaches may affect the results. Finally, the Cutler and Chen results concern tones and *vowels* while those here concerns tones and *consonants*. However, if indeed the relative depression of tone discrimination is due to the relative time of arrival of information, given that both studies used CV syllables, performance should actually be much better here for consonants. Clearly, more work is required on this issue.

For the English-speaking children, the *opposite* was found - better perception of a phonologically-irrelevant consonantal contrast than tone contrasts. It is striking that elsewhere (Burnham & Torstensson, 1995) we have found essentially the same effect with a different set of contrasts, Swedish tones and Swedish vowels. There it was found that English-speaking 6-year-old children discriminated Swedish vowels better than Swedish tones, and in a procedure which allowed identification on the basis of vowels or tones or both, 6-year-olds preferred to use vowels, even when very difficult vowels contrasts were incorporated, while older children and adults preferred to use tones. One interpretation of these results is in terms of Cutler and Mehler's (1993) periodicity bias: children very early learn to pick out the smallest unit of regularity in prosody in the ambient language and attend to these in preference to variable prosodic information. The fact that tone perception in the current study is at a relatively low level for children at all three ages is consistent with this interpretation, for by 4 years children may already have learned to disregard lexical tone. We agree. However, this can not constitute the whole story, because there is a resurgence of ability with tones by adulthood. If this were a resurgence based on linguistic principles then it would be inconsistent with Cutler's (Cutler, 1994; Cutler et al., 1992) claim that the learning of the rhythmic regularities occurs early and once. No, we think that this resurgence is simply due to adults being released from restrictions which bind the children, so that they can perceive the tones in all (or most) of their perceptual salience. Moreover, the processing level interaction (see Figure 3) suggests there may be some phonemic suppression of tones especially around 8 years. Note that this actually supports in an *ad hoc* way, an amalgamation of Cutler's view and the RAF model. For if tones are indeed lost early by virtue of their prosodic nature, but still robust, in RAF terms, they may still be subject to phonological bias at this late stage, just like other robust lost late contrasts.

Tones may also be difficult to discriminate because they fall into the single category (SC) assimilation pattern (Best, in press). This may well be so, but would not explain why adults are better than children. It is possible that in Best's terms the adults can perceive the tones in terms of category goodness (CG) and that some of these tones are better representatives of the [ba] category. This would need to be investigated in further studies. It may also be the case that adults are more able, for whatever reason, to perceive these tone contrasts as non-speech (NA pattern) and therefore rely on psychoacoustic rather than linguistic information. The fact that the most salient of the tone contrasts, D-D, is better discriminated when a less linguistic (500 msec ISI) mode of processing is allowed, shows that there are indeed psychoacoustic factors operating here even for Thais and even for children.

3. The Role of Linguistic Context in Adults' Perception of Tones

3.1 Introduction

In Best's model (Best et al., 1988; Best, in press) if two non-native sounds fall into a single native category (SC pattern) then they will be difficult to discriminate. However, if they are not perceived as speech (NA), then discrimination will be good to very good depending on the relative psychoacoustic salience of the sounds. Polka (1992; footnote 1) also suggests that if sounds are perceived as non-speech then subjects will perceive the dimensions of the stimulus differently (Repp, 1981; Best, Morrongiello & Robson, 1981). We suggested that the relatively poorer discrimination of tones by children in Experiment 1, is due to their phonological bias towards sound categories which have phonemic and graphemic representation in their language, and which have no analogous representation in their language such as the non-native voicing contrast does. There is some evidence from this experiment that presentation of speech contrasts, especially the tones (see Figures 3 & 4), under conditions favouring more phonetic/acoustic perception (500 msec ISI), results in better discrimination.

In order to investigate whether tone discrimination improves when the linguistic nature of the stimuli are removed, in Experiments 2 and 3 the tone stimulus sounds from Experiment 1 are explicitly manipulated so that they are perceived as non-speech. This is done by transforming the pitch contours of the speech tonal contrasts into two different non-speech formats: sine-wave analogues, and musical sounds (played on a violin). If perception of tones by the English speakers in Experiment 1 was poor because they perceived the phonologically-irrelevant sounds in a speech mode, then both children and adults should discriminate the tones better in the two non-speech formats than in the speech format. In Experiment 2 English-speaking adults' are tested on all 10 possible tone contrasts in the three formats at both 500 and 1500 msec ISI. In Experiment 3, English-speaking children are tested on a restricted set of five tone contrasts (chosen of the basis of the results of adults in Experiment 2) at just the lower ISI, 500 msec.

3. 2 Method

3.2.1 Subjects and Design

A total of 48 native English-speaking subjects were tested, 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. Subjects' musical experiences varied, however, no subjects were practising ensemble musicians. An ISI (500/1500msec) x Tone Type (Music, (M), SineWave Analogue, (SWA), Speech, (S)) x Tone Contrast (Static-Static, Static-Dynamic, Dynamic-Dynamic) design with repeated measures on the last 2 factors was employed. In each ISI group, half the subjects were males and half females.

3.2.2 Stimulus Materials and Apparatus

Three stimuli sets were created, Speech, SineWave Analogue and Music, each comprising 3 exemplars of each of the 5 Thai tones. The Speech stimuli were recorded by a female Thai native speaker. The syllable [pa] was used to carry the five tones: mid [pa:], low [pà:], falling [pâ:], high [pá:], rising [pǎ:]. Three of the five exemplars recorded were chosen as stimuli on the basis of their similarity and consistency.

These fifteen speech stimuli were then manipulated to create the Sine Wave Analogue stimuli and the Music stimuli. The SineWave Analogues were created by repeat filtering the speech sounds with a low pass digital filter. This removed the upper formants from the speech whilst leaving the fundamental frequency intact. Filtering reduced the intensity of the sounds and as a result the filtered sinewaves were not quite as loud as the speech stimuli.

The Music tones were more difficult to create. Analysis of the speech tokens showed that fluctuations in frequency occur throughout any particular exemplar. It would have been possible to use a music synthesiser to construct a copy comprised of a series of linear segments based on averaged frequencies of the speech sample, or to alter the fundamental frequency of an existing musical sound, but this would disturb the relationship between the upper harmonics and the fundamental frequency and lead to unsatisfactory results. The best solution was to create the Music sounds with a musical instrument. The violin was chosen because we required an instrument which could both maintain a continuous sound and reproduce rapid pitch changes, in order to deal with the pitch dynamics of the Thai falling tone, for example, which covers approximately one and half octaves in a short space of time. A professional violinist listened extensively to the speech recordings and then reproduced the sounds on his violin. Approximately 25 exemplars of each tone were recorded and digitised. Selection of the final 3 exemplars of each tone was based on their similarity to the original speech tokens.

All the sounds were digitised and stored on disk of IBM-compatible computers. In addition to the three stimuli sets, 3 context tapes were constructed. For each of the three stimuli types, a "context-setting" tape recording was played prior to experimental trials.

The Speech context tape was part of a recording of a woman coversing in Thai. The Sine Wave tape was created by concatenating a one minute selection of the sinewave analogues and recording repeats of the sequence. The context tape for the Music sounds was a violin recording of Bach's Crab Canon.

For each stimulus type, all possible pairings of the five tones were used, a total of 10 tone contrasts. One of the three exemplars of each sound was chosen at random for presentation in each trial. This exemplar variation over trials was employed to encourage linguistic processing, and to discourage 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 both 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.

3.2.3 Procedure

Each subject completed three tasks, in counterbalanced order. The 3 tasks were identical, differing only in the stimulus type employed, Speech, SineWave Analogue or Music. Each task comprised four distinct phases. The first simply required the subject to listen to the appropriate context tape. The second was a task competence phase in which subjects were required to respond correctly on four simple auditory distinction [ræg vs rʌg]. There were two same, and two 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. Two test blocks followed, with 40 trials in each block. Five of the possible 10 contrasts were presented in the first block, and the other five in the second block. The order of presentation of blocks was counterbalanced between subjects. For each contrast pair, the four possible combinations (AA, AB, BA, BB) were presented twice. Thus in each block of 40, there were twenty same and twenty different trials. The actual exemplars presented on any particular trial were selected randomly by the computer from the pool of three possible exemplars for each sound. The subject was required to listen to the randomly presented contrast pairs and respond as quickly as possible by pressing either the "same" or "different" key. The final phase of the task was the completion of two rating scales. Subjects were asked to rate the similarity of the sounds to speech and music on 2 separate scales. The scales ranged from 1 to 7 where 1 is “not at all like speech” or “not at all like music”, and 7 is “exactly like speech” or “exactly like music”.

3.2.4 Dependent Variables

The number of correct and incorrect responses and the accompanying reaction times were recorded on disk. Only the correct and incorrect responses were used here, and these were converted to a discrimination index as described in Experiment 1. In addition, the rating scale scores were used, where a score of 1 indicated an exact likeness to speech or music and a score of 7 indicated no similarity to speech or music.

3.3 Results

Discrimination indices (DIs) were analysed by planned contrasts in an analysis of variance: 2 x (3 x 3), ISI (500msec, 1500msec) x Tone Type (Music - M, SineWave Analogue - SWA, Speech - S) x Tone Contrast (Static-Static, Static-Dynamic, Dynamic-Dynamic), with repeated measures on the second two factors. The critical F-value was $F_{.05}$ (1, 46) = 4.05. There were no significant effects of ISI or its interactions. This may be due to the reduced power in this experiment compared to Experiment 1.

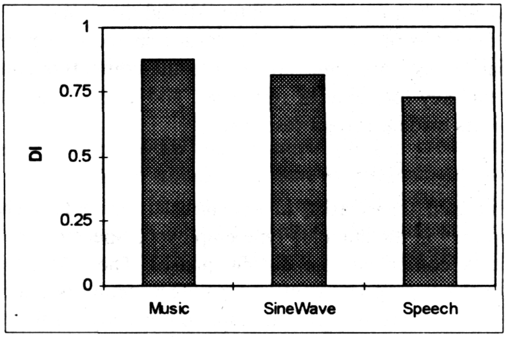

Figure 5. Mean discrimination indices of the three stimulus types.

The most striking finding is that shown in Figure 5. English-speaking adults better discriminated the non-speech tones (Music and Sine-Wave Analogues) than the Speech tones, F (1,46) = 20.15, and within the non-speech, the Music better than the SineWave Analogues, F (1,46) = 12.42, (X_M =.87, X_{SWA} =.81 , X_S =.73). Over all three stimulus types, adults were better at the D-D than either the S-D or S-S contrasts, F (1,46) = 12.08, and showed statistically equivalent performance for S-D, and S-S contrasts, F (1,46) = 0.01 ($X_{D\text{-}D}$ =.85, $X_{S\text{-}D}$ =.80, $X_{S\text{-}S}$ = .80).

The Music > Sine Wave Analogue > Speech pattern of results was generally obtained across all three S-S contrasts, and all six S-D contrasts, however, as shown in Figure 6, a slightly different pattern, Sine Wave Analogue > Music > Speech, was obtained for the D-D contrast, F (1,46) = 14.91. So, as in Experiment 1, the most perceptually salient of the contrasts, the rise-fall contrast, behaves differently. Nevertheless, even for this contrast, discrimination for each of the two non-speech tone types was significantly better than for the speech tones.

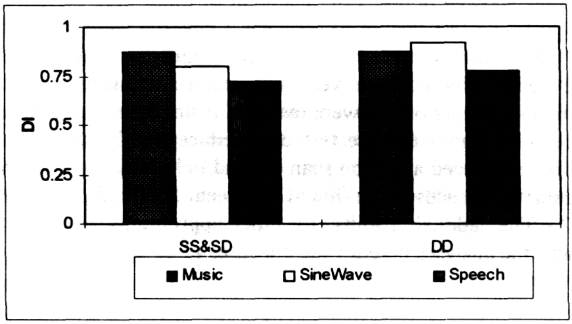

Figure 6. Mean discrimination indices of SS & SD contrasts compared to DD contrasts

	Music		SineWave		Speech	
	like speech?	like music?	like speech?	like music?	like speech?	like music?
Mean	2.06	5.54	3.10	3.46	5.15	2.38

Table 1. Mean ratings of Music, SineWave and Speech tones

Subjects' ratings of the three tone types as music or speech on a 7-point scale is shown in Table 1. These results confirm that adults were perceiving the stimuli as intended: music and sinewave analogue stimuli were rated to be more music-like than were the speech stimuli; and speech stimuli were rated to be more speech-like than were the music and sine-wave analogue stimuli, $F(1, 46) = 145.04$.

3.4 Discussion

The results of this experiment clearly show that adult speakers of a non-tonal language (English) discriminate the pitch variations in tonal contrasts of a natural language (Thai), better when these are presented as non-speech than when they are presented as speech. The implication is that in this particular case, perception in a speech mode actually attenuates an underlying psychoacoustic ability. This occurs even though adults' perception of speech tones is quite good, and certainly better than their perception of another non-native contrast, [ba]-[pa] (see Experiment 1). It remains to be seen whether the same effect is evident for children.

4. The Role of Linguistic Context in Children's and Adults' Perception of Tones

4.1 Introduction

In Experiment 2 it was found that adults' perception of the pitch differences in tonal stimuli was facilitated when these were presented in non-speech stimuli. It would be expected that the same effect should be found for children, especially given that in

Experiment 1 they showed poorer discrimination of tones than the other non-native contrast, [ba]-[pa]. To test this, in Experiment 3 children were also tested on their perception of tones with the three types of stimuli, music, sine-wave analogues, and speech. Three main differences between this and Experiment 2 were that here, children of 5, 6, and 8 years of age were tested; a restricted range of five rather than the possible 10 tone contrasts were tested to reduce the duration of the test and allow for children's reduced attention span by; and only the 500 msec ISI was used, as no significant differences were found between 500 and 1500 msec ISIs in Experiment 2. The adults' results for the appropriate five contrasts were incorporated into the analyses here for the purposes of comparison.

4.1 Method

4.1.1 Subjects and Design

A total of 72 native English-speaking subjects were tested, each with little or no experience of other languages, and with no experience of Thai or any tonal language. There were 24 8-year-olds, 24 6-year-olds, and 24 5-year-olds. An Age (8, 6, 5 years) x Tone Type (Music, (M), SineWave Analogue, (SWA), Speech, (S)) x Tone Contrast (Static-Static, Static-Dynamic, Dynamic-Dynamic) design with repeated measures on the last 2 factors was employed. In each ISI group, half the subjects were males and half females.

4.1.2 Stimulus Materials and Apparatus

The same three stimuli sets which were created for Experiment 2 were used here: Speech, SineWave Analogue and Music, each set comprising 3 exemplars of each of the 5 Thai tones. The Speech stimuli were recordings of the five Thai tones carried on the syllable [pa] and recorded by a female Thai native speaker. These fifteen speech sounds were manipulated to create the Sine Wave Analogue stimuli and the Music stimuli, (see Experiment 2 Method). In addition to the three stimuli sets, the 3 context tapes used in Experiment 2 were used for this study.

For each stimulus type, only five tone contrasts were used. The contrasts selected were: Rising-Falling, High-Mid, High-Low, Low-Rising, and Low-Falling. These five contrasts were chosen on the basis of a detailed examination of the adults' results, taking into account the desirability of having all three tone contrast types (SS, SD, and DD), represented and represented in roughly equal numbers, and taking care to include contrasts which covered the range of adults' discrimination ability from poorest to best.

The experiment was conducted using the same computer-based systems as those listed in Experiment 2. Apart from increasing the maximum response time allowed to 2 sec (from 1 sec in Experiment 2) there were no differences in the apparatus used for children. As for the adult experiment, one of the three exemplars of each sound was chosen at random for presentation in each trial, in order to encourage linguistic processing, and to discourage acoustic processing.

4.1.3 Procedure

As for adults, each child completed three tasks, in counterbalanced order. The three tasks differed only in stimulus type employed, Speech, SineWave Analogue or Music. Each task was composed of the same four phases. The first required the subject to listen to the appropriate context tape, and the second was the task competence phase in which subjects were required to respond correctly on four simple auditory distinction [ræg vs rʌg]. The task competence phase was important to ensure that each child could make simple auditory discriminations, and to help them familiarise to the task and the apparatus. One test block of 40 trials followed in which the selected five contrasts were presented. For each contrast pair, the four possible combinations (AA, AB, BA, BB) were presented twice. Thus in the block of 40, there were twenty same and twenty different trials. The exemplars presented on any particular trial were selected randomly by the computer from the three possible exemplars for each sound. The subject was required to listen to the randomly presented contrast pairs and respond as quickly as possible by pressing either the "same" or "different" key. As for adults, the final phase of the task required subjects to rate the similarity of the sounds to speech and music using the scales described in Experiment 2 Method.

4.1.4 Dependent Variables

Only correct and incorrect responses were used, and as in Experiments 1 and 2, these were converted to a discrimination index (see description in Experiment 2).

4.2 Results

The relevant results for the adults, DIs at 500 msec ISI for the two S-S contrasts, high-low, and high-mid, the two S-D contrasts, low-rising, and low-falling, and the D-D contrast, rising-falling, were incorporated and the resulting amalgamation analysed by planned contrasts in a 4 x (3 x 3), Age (5, 6, 8, adults) x Tone Type (Music - M, SineWave Analogue - SWA, Speech - S) x Tone Contrast (Static-Static, Static-Dynamic, Dynamic-Dynamic), with repeated measures on the second two factors. The critical F-value was $F_{.05}$ (1, 46) = 3.95.

As can be seen in Figure 7, the results of the children at all three ages mirror those of the adults. This was confirmed by the analysis of variance which showed that combined all subjects better discriminated the non-speech tones (Music and Sine-Wave Analogues) than the Speech tones, F (1,92) = 56.76, and within the non-speech, the Music better than the Sine-Wave Analogues, F (1,92) = 90.558, (X_{M} = .61, X_{SWA} = .42, X_{S} = .36). Over all three stimulus types, subjects were better at the D-D than either the S-D or S-S contrasts, F (1,92) = 16.19, and showed statistically equivalent performance for S-D and S-S contrasts, F (1,92) = 0.77 ($X_{D\text{-}D}$ = .53 , $X_{S\text{-}D}$ = .457, $X_{S\text{-}S}$ = .44).

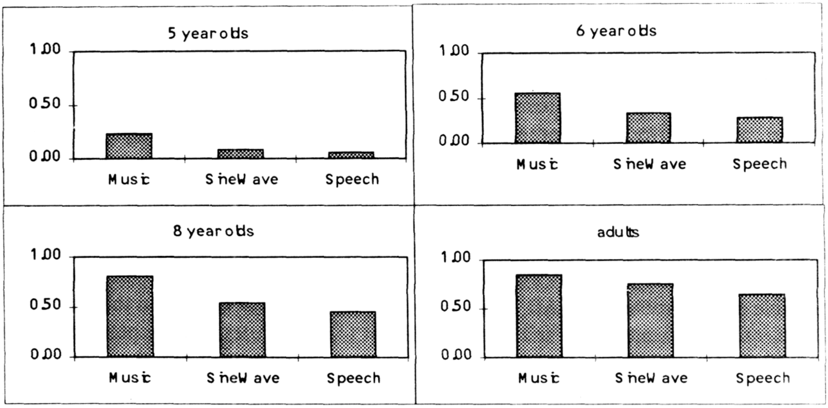

Figure 7. Discrimination indices on the three stimulus types for the four age groups

The Music > Sine Wave Analogue > Speech pattern of results was obtained across the five contrasts in the three tone contrast types, S-S, S-D, D-D, ie, there were no significant Tone Type x Tone Contrast effects. There was however, an S-S vs S-D x 6 vs 8 year-olds interaction, F(92) = 4.25, which, reveals that the improvement between 6 and 8 years was more dramatic for S-D contrasts (.36 to .60), than for S-S contrasts (.40 to .56).

There were also two interactions of age with Tone Type. As can be seen in Figure 7, the difference between Music and SineWave Analogues was generally greater for the children (.53 vs .32) than the adults (.85 vs .75), F (1,92) = 7.55; and was generally greater for the 6- and 8-year-old children combined (.56 vs .32) than the 5 year-olds (.23 vs .08), F (1,92) = 5.43. So in general the superiority of discrimination of Music over SineWave Analogues gradually increased over childhood but then levelled out in adults. This may be due a gradual improvement in an ability to hear music in a "musical mode" over development, although no more can be said about this on the basis of the current results. Despite these age interactions, discrimination for the two non-speech tone types was significantly better than for the speech tones, irrespective of age.

4.3 Discussion

The results of this experiment show that, just as for adults, child speakers of a non-tonal language (English) discriminate the pitch variations in tonal contrasts of a natural language (Thai), better when these are presented as non-speech than when they are presented as speech. This facilitation was equivalent for all the three child age groups and adults, so for non-tonal language listeners of all the ages tested, the underlying psychoacoustic discriminabilty of differences in pitch height and contour is attenuated in a linguistic context. It seems that non-tonal language speakers, by at least the age of 5 years have learned to ignore what are for them, irrelevant changes in lexical tone. However, this conclusion must be drawn with caution until studies which parallel Experiments 2 and 3 are conducted with tone language speakers. Such studies, using the same Thai stimuli, are currently being conducted with Thai speakers, and also with speakers of other lexical tone languages.

5. General Discussion

In Experiment 1 the perception of one Thai contrast which can be considered native to English speakers, the voicing contrast, [ba]-[pʰa], and two contrasts which are non-native to English speakers, the [ba]-[pa] voicing contrast and various lexical tone contrasts was investigated. Apart from the expected language background differences between Thai and English speakers, the most interesting finding was that while English-speaking adults perceived the tones better than the [ba]-[pa] contrast, their younger counterparts did the opposite, they perceived the [ba]-[pa] contrast better than the tones. This confirms similar results we have found with Swedish tone vs vowel contrasts - English-speaking adults perceived the tones better than vowels, while children perceived vowels better than the tones (Burnham & Torstensson, 1995). Both of these contrasts fall into either Best's SC or CG pattern of assimilation (Best et al., 1988; Best, in press). (If the latter, then it would be of interest in future studies to investigate the degree to which various tonal variations may depart from the most typical in non-tone language speakers' phonetic space.) However, while Best's model has proven powerful in accounting for the results of a number of cross-language studies with both adults and infants, it is not yet clear how it could be used or adapted to account for different results in development. Perhaps studies involving category goodness ratings for tones with both adults and children are required to resolve this issue.

The age by consonant/tone contrast results in Experiment 1 may be a special case of what Burnham (1986) in the RAF model has called phonological bias. In that model Burnham claimed that contrasts which are robust in a psychoacoustic sense, and in their representation the world's languages, and which are present allophonically in the child's ambient language should be "lost late" around the onset of reading, when children begin to map perceived and produced phones onto the available graphemic structure of their language. Tones certainly seem to be psychoacoustically salient with their relative differences and contours of F_0, although the studies here and elsewhere (Abramson, 1974, 1978, 1986; Burnham & Francis, 1995; Gandour, 1979, 1983) show that salience differs between particular tones. Tones are also predominant linguistically, occurring in over half of the world's languages (Goldsmith, 1994). Finally, lexical pitch variations are indeed allophonically present in non-tone language speakers' language environment. It appears that tone perception in the children tested in Experiment 1 was already at a low rate and that there was not the characteristic dip in performance for robust contrasts found in other studies (Burnham, 1986; Burnham et al, 1991). Possible reasons for this are that a discrimination task rather than a categorical identification task was used here (but see Burnham et al., in preparation), and that both consonant and tone contrasts were tested in the same session, which may have masked any such dip. However, there is also the strong possibility that by 4 years children had already learned to disregard tones because of the special place attention to prosodic information may play in the child's early attempts to segment the ambient speech stream (Cutler, 1994; Cutler & Mehler, 1993). The results here appear to bear this out: at 4 years children's perception of tones is already at a low level and lower than that of the non-native consonant contrast (Figure 1). Nevertheless, if the difference between phonetic and phonemic processing is inspected (Figure 3), it can be seen that there is an abrupt rise in the relative degree of phonetic processing of tones at 8 years, at which stage such a phonetic superiority for [ba]-[pa] is waning. This *could* signal a late phonological bias beginning to operate on tones, however we must acknowledge that this argument is *ad hoc*. Clearly, more precise predictions need to be made and tested in future studies. Studies of the development of tone perception also need to be extended to younger children, so that changes in the relative degree of attention to prosodic (both tonal and intonational) and segmental distinctions over age can be ascertained.

In Experiments 2 and 3 it was found for adults and children respectively that translation of lexical tone contours onto non-linguistic carrier stimuli resulted in improved discrimination of pitch differences by English speakers. This suggests that at all ages, 5, 6, 8 years and adulthood, linguistic experience with a non-tone language had served to attenuate differences which are not important in the language. Thus linguistic experience appears to have a detrimental effect on pitch discrimination, an ability which appears to have a strong psychoacoustic basis. This attenuation presumably has an advantage for the processing of non-tone languages, as it would help listeners to suppress irrelevant and distracting information. However, it is yet to be determined whether tone language speakers show the opposite effect - facilitation, or at least lack of attenuation, for pitch variations in a linguistic context. These studies are currently being conducted.

If tone and intonation can be thought of as two aspects of prosody (Cutler, 1994, 1995), then it should also follow that the attenuation shown here for prosodic features should be specific to lexical tone. Tones were presented in isolation on single syllables here. Speakers of English should perhaps be more sensitive to lexical, or supra-lexical, pitch differences presented at the end of a short sentence than in the middle, and there should be predictable differences in such experiments between speakers of lexical tone and non-tone languages. When young children begin to produce language they appear to approach the task at the level of words (Wode, 1992, 1994; Piske, 1995). Extensive experiments of this nature are yet to be conducted (though see preliminary results in Lacerda, 1995). If such experiments were conducted across infancy, childhood, and adulthood then, with the appropriate manipulations, some specification of when the young infant begins to process the ambient speech in terms of words rather than syllables, phonemes, or intonation patterns may be made.

References

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

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

Abramson, A.S. (1986). The Thai tonal space. *Haskins Laboratory Status Report on Speech Research*, SR-85, 105-114.

Allen, G.D. (1983) Linguistic experience modifies lexical stress perception. *J. of Child Lang., 10*, 535-549.

Aslin, R.N., Pisoni, D.B., Hennessy, B.L., & Peters, A.J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Dev., 52*, 1135-1145.

Best, C. T. (1990) Adult perception of nonnative contrasts differing in assimilation to native phonological categories. *J. of the Acoustical Soc. of Amer, 88:* S177.

Best, C. T. Faber, A., & Levitt, A. (In preparation) Association between adults perceptual assimilation and discrimination of diverse non-native vowel contrasts.

Best, C.T. (in press) A direct realist view of cross-language speech perception. In W. Strange (Ed) *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* Timonium, MD, USA: York Press.

Best, C.T., McRoberts, G.W., Sithole, N.M. (1988) Examination of perceptual reorganization for nonnative speech contrasts: Zulu Click Discrimination by English speaking adults and infants. *Journal of Experimental Psychology:HPP, 14*, 345-360.

Best, C.T., Morrongiello, B., & Robson, R. (1981) Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics, 29*, 191-211.

Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993) How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychol., 29*, 711-721.

Bond, Z.A., & Adamscu, L. (1979) Idenification of novel phonetic segments by children, adolescents and adults. *Phonetica, 36*, 182-186.

Bredberg, G. (1985) The anatomy of the developing ear. In S.E. Trehub & B. Schneider (eds) *Auditory development in infancy*. New York: Plenum Press.

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. *J. of Child Lang., 18*, 231-260.

Burnham, D., Earnshaw, L., O'Connor, G., & Clark, J. (In preparation) Native-language phonological bias in speech perception and the onset of reading.

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 3rd International Symposium on Language and Linguistics* Bangkok: Chulalongkorn Univ. Press, 546-560.

Burnham, D.K., & Torstensson, C. (1995) Phonological bias and foreign accents: Perception and production of Swedish vowels tones by English speakers. *Proceedings of the ICPhS, Stockholm*, *4*, 558-561.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994) Do infants perceive word boundaries? An empirical studyof the bootstrapping of lexical acquisition. *J. of the Acoustical Soc. of Amer*, *95*, 1570-1580.

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

Cutler, A. (1989) Auditory lexical access: Where do we start? In W. Marslen-Wilson (Ed.) *Lexical representation and process.* MIT Press: Cambridge Massachussets.

Cutler, A. (1994) Segmentation problems, rhythmic solutions. *Lingua*, *92*, 81-104.

Cutler, A. (1995) Universal and language-specific in the development of speech. Symposium on "Uniqueness and Universality in a Biological World", UNESCO, Paris.

Cutler, A., & Chen, H-C. (Submitted) Lexical tone in Cantonese spoken-word recognition.

Cutler, A., & Mehler, J. (1993) The periodicity bias. *J. of Phon.*, *21*, 103-108.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992) The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, *24*, 381-410.

DeCasper, A.J. & Fifer, W.P. (1980) Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*, 1174-1176.

DeCasper, A.J. & Spence, M. (1986) Newborns prefer a familiar story over an unfamiliar one. *Inf. Behav. & Development*, *9*, 133-150.

Demuth, K. (1993) Issues in the acquisition of the Sesotho tonal system. *J. of Child Lang.*, *20*, 275-301.

Eilers, R.E., Gavin, W.J., & Oller, D.K. (1982) Cross-linguistic perception in infancy: Early effects of linguistic experience. *J. of Child Lang.*, *9*, 289-302.

Eilers, R.E., Oller, D.K., Bull, D.H, & Gavin, W.J. (1984) Linguistic experience and infant speech perception: A reply to Jusczyk, Shea, & Aslin (1984). *J. of Child Lang.*, *11*, 467-475.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303-306.

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

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

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

Hirsch-Pasek, K., Kemler-Nelson, D.G., Jusczyk, P.W., Wright-Cassidy, K., Druss, B., & Kennedy, L. (1987) Clauses are perceptual units for young infants. *Cognition*, *26*, 269-286.

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.* Cambridge, MA: Newbury.

Jusczyk, P.W., Shea, S.L., & Aslin, R.N. (1984)) Linguistic experience and infant speech perception: A re-examination of Eilers, Gavin, & Oller (1982). *J. of Child Lang.*, *11*, 453-466.

Jusczyk, P.W., Cutler, A., & Redanz, N.J. (1993) Infants' preference for the predominant stress patterns of English words. *Child Dev.*, *23*, 648-654.

Jusczyk, P.W., Friederici, A.D., Wessells, J.M.I., Svenkerud, V., & Jusczyk, A.M. (1993) Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Lai uage*, *32*, 402-420.

Jusczyk, P.W., Hirsch-Pasek, K., Kemler-Nelson, D.G., Kennedy, L.J., Woodward, A., & Piwoz, J. (1992) Percption of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*, 252-292.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experience alters phonaetic perception in infants by 6 months of age. *Science*, *255*, 606-608.

Lacerda, F. (1995a) The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory. *Proceedings of the ICPhS, Stockholm*, *2*, 140-143.

Lacerda, F. Sundberg, U., Andersson, C. & Rex, A. (1995b) Young infants' perception of segmental and suprasegmental information: Preliminary results. *Proceedings of the ICPhS, Stockholm*, *3*, 408-411.

Li, C. & Thompson, S. (1977). The acquisition of tone in Mandarin-speakingchildren. *J. of Child Lang.*, *4*, 185-199.

Lisker, L. & Abramson, A. (1964) A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20,* 384-422.

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: CUP.

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

Mehler, J., Bertoncini, J., Barriere, M. & Jassik-Gerschenfeld, D. (1978) Infant recognition of the mother's voice. *Nature*, *7*, 491-497.

Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, N., Bertoncini, J. & Amiel-Tison, C. (1988) A precursor of language acquisition in young infants. *Cognition*, *29*, 143-178.

Melhuish, E.C. (1982) Visual attention to mother's and stranger's faces and facial contrast in 1-month-old infants. *Developmental Psychology*, *18*, 229-231.

Mills, M. & Melhuish, E. (1974) Recognition of mother's voice in early infancy. *Nature*, *252*, 123-124.

Moon, C. & Fifer, W.P. (1986) Newborn infants prefer the sound of mother's voice as experienced in the womb. Paper presented at the International Society for Developmental Psychobiology, Annapolis, Maryland.

Moon, C., Bever, T.G., & Fifer, W.P. (1992) Canonical and non-canocical syllable discrimination by two-day-old infants. *J. of Child Lang.*, *19*, 1-17.

Panneton Cooper, R.K., & Aslin, R.N. (1989) The Language environment of the young infant: Implications for early perceptual development. *Canadian Journal of Psychology, 43*, 247-265.

Panneton, R.K. & DeCasper, A.J. (1984) Newborns prefer an intrauterine heartbeat sound to a male voice. Poster presented at the International Conference on Infant Studies, New York.

Piske, T. (1995) Articulatory patterns in early speech production *Proceedings of the ICPhS, Stockholm*, *2*, 698-701.

Polka, L. & Werker, J.F. (1994) Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: HPP*, *19*, 421-435.

Polka, L., & Bohn, O-S. (submitted) A cross-language comparison of vowel perception in English-learning and German-learning infants.

Querleu, D. & Renard, X. (1981) Les perceptions auditives du fetus humain. *Medicine & Hygiene*, *39*, 2101-2110.

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L. & Crepin, G. (1988) Fetal hearing. *European Journal of Obstetrics & Gynaecology & Reproductive Biology*, *29*, 191-212.

Repp, B.H. (1981) Two strategies in fricative discrimination. *Perception & Psychophysics*, *30*, 217-227.

Streeter, L.A. (1976) Language perception of 2 month-old infants shows effects of both innate mechanisms and experience. *Nature*, *259*, 39-41.

Trehub, S. (1976) The discrimination of foreign speech contrasts by infants and adults. *Child Dev.*, *47*, 466-472.

Tse, J.K. (1978). Tone acquisition in Cantonese: A longitudinal case study. *J. of Child Lang.*, *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.F. (1993) The contribution of the relation between vocal production and perception to a developing phonological system. *J. of Phon.*, *21*, 177-180.

Werker, J.F., & Lalonde, C.E: (1988) Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychol.*, *24*, 672-683.

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

Werker, J.F., & Tees (1984a) Cross language speech perception: Evidence for perceptual reorganization during the first year of life. *Inf. Behav. & Development*, *7*, 49-63.

Werker, J.F., & Tees (1984b) Phonemic and phonetic factors in adult cross-language speech perception. *J. of the Acoustical Soc. of Amer*, *75*, 1866-1878.

Werker, J.F., & Tees, R.C. (1983). Developmental changes across childhood in the perception of non-native speech sounds *Canadian Journal of Psychology, 37*, 278-286.

Werker, J.F., Gilbert, J.H.V. Humphrey, K., & Tees, R.C. (1981). Development aspects of cross-language speech perception. *Child Dev., 52*, 349-355.

Wode, H., (1992) Categorical perception and segmental coding in the ontogeny of sound systems: A universal approach. In C.A. Ferguson, C.A., C. Menn & C. Stoel-Gammon (Eds) Phonological development: Models, research and implications. Timonium, Maryland: York Press.

Wode, H., (1994) Speech perception and the learnability of languages. *International Journal of Applied Linguistics, 4*, 143-168.

Footnotes

---

[1] Note also that in an unsympathetic language environment the voiced alveolar flap vs trill contrast, [ɾ]-[r], appears to be lost early, around the time vowel perception is re-organised (before 6 months) rather than when consonant perception is re-organised (8 - 12 months). This contrast is phonemic in Spanish but not English, and Eilers, Gavin & Oller (1982) found that both 6-month-old English language environment infants could not perceive this contrast, while Spanish language environment infants could. (However, criticisms of the Eilers et al., 1982, study have been made by Jusczyk, Shea, & Aslin, 1984, but see reply by Eilers, Oller, Bull & Gavin, 1984. See also Burnham, 1986, pp.227, 232,236.)

2 In the fledging literature on non-native vowel perception, there is also evidence that perception of some vowel contrasts is not attenuated in infancy and is maintained in adulthood. Thus Werker & Polka (1993) found attenuation of discrimination of the Southern dialect German [u]-[y] contrast by ELE infants, but Polka & Bohn (in press) found no such attenuation for the Northern dialect equivalent. Both fall into Best's CG pattern, but adults report a smaller difference in category goodness for the Southern dialect version (Polka & Bohn, in press). Similar lack of attenuation has been found for German infants' and adults' perception of the English [æ]-[ɛ] contrast, which German adults assimilate as Uncategorised vs Categorised (UC) vowels (Polka & Bohn, in press, Polka, 1995). Best, Faber, & Levitt (in preparation, see Best, in press) have tested ELE adults on a range of vowel contrasts, finding a range of assimilation patterns. For example, a Norwegian outrounded/unrounded contrast for high front vowels resulted in a SC assimilation pattern and poor discrimination. This and the other vowel contrasts tested by Best et al (in preparation) have yet to be tested on infants.

3 This is not the case in the acquisition of more complex grammatical tone systems, in which problems in mapping underlying tonal representations onto surface structure may persist until 3 years (Demuth, 1993).

4 This is supported by the fact that grammatical tone systems (Demuth, 1993) appear to be more difficult to learn than lexical tone systems.