FACILITATION OR ATTENUATION IN THE DEVELOPMENT OF SPEECH MODE PROCESSING? TONE PERCEPTION OVER LINGUISTIC CONTEXTS

Denis Burnham[♠], Elizabeth Francis[♠], Di Webster[♠], Sudaporn Luksaneeyanawin[♦], Francisco Lacerda[♣], and Chayada Attapaiboon[♦]

[♠] School of Psychology; University of NSW; [♦] Linguistics Research Unit, Chulalongkorn University; [♣]Department of Phonetics, Stockholm University

ABSTRACT - Perceptual discrimination of Thai tones was tested in three contexts: normal speech, filtered speech, and musical sounds with speakers of Central Thai; speakers of Cantonese, a tonal language of similar complexity; and speakers of English. The inter-sound interval (ISI) was varied: half the subjects in each group were tested with 500 msec ISI (to encourage phonetic language-general processing), and half with 1500 msec ISI (to encourage phonemic language-specific processing). English speakers discriminated tonal contrasts best in the musical context, and better in filtered than in full speech. Thai speakers, however, discriminated the tonal contrasts equally well in all three contexts, although *manner* of processing across contexts differred: reaction times for speech showed a 1500 msec advantage, while for music and filtered speech there was a 500 msec advantage. Cantonese speakers' showed some similarities to their fellow tonal language speakers, and some to the non-tonal, English speakers. Preliminary results for a group of native speakers of Swedish, a pitch accented language with two tonal variants, suggest that they respond similarly to the Cantonese subjects. In Experiment 2 Thai-speaking and English-speaking children of 5, 6, and 8 years were tested in the three contexts. Their results essentially mirrored those of their adult counterparts. Thus it appears that a special mode of speech processing becomes established relatively early in life.

INTRODUCTION

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

Twenty-five years of speech perception research shows that neonatal infants can perceive just about any consonantal contrast, both relevant and irrelevant to the ambient language environment, on which experimenters wish to test them (for reviews, see Burnham, 1986; Werker, 1991). From this universal beginning children become increasingly attuned to the structure of the ambient language. There are two intensive periods of selective reorganisation. The first is the second six months of the first year, which has been extensively researched by Werker and her colleagues (Werker, 1991); the second occurs soon after the onset of reading instruction (Burnham, 1986). However, the loss of speech perception ability that occurs in these periods is not sensorineural loss. Werker and her colleagues have conducted ingenious experiments showing that when testing conditions favour phonetic or acoustic processing (reducing the interstimulus interval (ISI) between sounds to be discriminated to 500 or 250 msec respectively), adults can usually perceive non-native consonant contrasts, while with a 1500 msec ISI, forcing reliance on long-term memory and thus phonemic processes, they find great difficulty (Werker & Logan, 1985).

Recent research suggests that perceptual reorganisation favouring ambient language characteristics also occurs for vowels (Polka & Werker, 1994; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), prosody (Hirsh-Pasek, Kemler-Nelson, Jusczyk, Wright-Cassidy, Druss, & Kennedy, 1987), and the stress patterns of words (Jusczyk, Cutler & Redanz, 1993). Rather less work has been conducted on the pitch variations of individual syllables or words, ie lexical tone. Cutler, Mehler, Norris, and Segui (1992) suggest that infants have a "periodicity bias" and attend to the smallest level of rhythmic regularity in the ambient language. If so, then infants should be especially tuned to intonation, rhythm, stress, and tone; discover which of these are used regularly in the language around them; and attend to these regularities and ignore other more random variations. 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).

In two recent studies involving the tonal languages Swedish (Burnham & Torstensson, 1995) and Thai (Burnham, Francis & Webster, 1996), Burnham and colleagues found that English-speaking children perceive non-native vowels and consonant contrasts better than non-native lexical tone contrasts, while English-speaking adults perceive non-native tonal contrasts better than non-native segmental contrasts. This is thought to be due to young English speakers learning that tonal variations at the lexical level, are relatively unimportant compared with segmental distinctions. Moreover, this may suggest that children are particularly sensitive to those aspects of intonation which are functionally relevant in *their* language environment. Similar results should be expected for adults, of course, but Burnham argued that adults' superior perception for tones was due to their ability to treat the experiment as a perceptual task, in which F_0 variations in tone contrasts are more acoustically salient (Best, in press) than temporal and/or spectral variations in segmental contrasts.

It is possible that English-speaking adults' perception of tones bears the vestiges of perceptual reorganisation. In Best's Perceptual Assimilation Model (Best, in press), two syllables differing only on tone would be said to fall into the Same Category for English speakers, and thus would be difficult to discriminate so long as they are perceived as speech. However, if they are *not* perceived as speech, discrimination will be good to very good depending on their psychoacoustic salience.

EXPERIMENT 1

The first experiment was conducted in order to examine whether English speakers' perception of pitch variations in lexical tone is improved when they are able to process these as non-speech. To this end Thai speech tonal contrasts carried on [pa:] (Burnham & Francis, in press) were modified into two different non-speech formats: low-pass filtered speech, and musical (violin) sounds. Native English speakers, speakers of Thai (the language of the specific contrasts used here) and speakers of Cantonese (a tonal language of equivalent complexity to Thai) were tested. Two ISIs were employed, 500 msec and 1500 msec, in order to investigate the level of processing (Werker & Logan, 1985). Preliminary results for speakers of Swedish (a language with a less complex tonal system) are also presented. By comparison of perceptual ability across stimulus types, the relative roles of attenuation of due to phonological irrelevance (for English), and facilitation due to phonological relevance (for Thai, Cantonese, and Swedish speakers) can be determined.

Method

A total of 168 adults were tested: 48 English-, 48 Thai-, 48 Cantonese-, and 24 Swedish-speakers. These three groups were treated in a Language Background (English, Thai, Cantonese) x ISI (500, 1500 msec) x Tone Type (Music, Filtered Speech, Speech) x Tone Contrast (10 possible AB pairings of the 5 Thai tones) design with repeated measures on the last two factors. Half the subjects in each language group were tested at each ISI and, within these subgroups, half the subjects were males and half females. The 24 Swedish subjects were all tested in the 500 msec ISI condition.

Three stimulus sets were created, Speech, Filtered Speech, and Music, each comprising three exemplars of each of the five Thai tones. The Speech stimuli were recorded from a female native Thai speaker (SL), using the syllable [pa:] to carry the five tones: rising ☯p♋→▫☸, high ☯p♋❸▫☸, mid ☯p♋▫☸, low ☯p♋∩▫☸, falling ☯p♋←▫☸, (R, H, M, L, F, hereafter). The Filtered Speech stimuli were created by repeat low-pass filtering the speech sounds. This reduced the upper formants while leaving the fundamental frequency intact. The Music stimuli were created by a professional musician, who listened extensively to the tones in speech and then reproduced approximately 25 exemplars of each tone on the violin. From these the final three music exemplars for each tone were selected.

The experiment was conducted in parallel at the University of NSW (English and Cantonese speakers), and Chulalongkorn University (Thai speakers) on identical portable systems, each consisting of a Toshiba 3100e AT laptop computer modified to accommodate D-A, digital I/O, and filter boards. The computers stored the sounds on disk, controlled presentation and timing of sounds, and recorded subjects' responses and reaction times (RTs). An attached response panel contained a "same" key and a "different" key for subjects' responses. At Stockholm University, Swedish subjects were tested on an equivalent system.

Each subject completed three AX discrimination tasks, identical except for the stimulus type employed, Speech, Filtered Speech, or Music. In each the subject first listened to 1-minute "context" tape (a woman conversing in Thai, a concatenation of filtered speech excerpts, a violin recording of Bach's Crab Canon, respectively). In each, two 40-trial test blocks were given, each with five of the possible 10 different contrast pairs presented in the first block, and the other five in the second block. For each contrast pair (eg, L/F), each of the four possible combinations (L-L, F-F, L-F, F-L) were presented twice. The actual exemplars of each tone on any particular trial were selected randomly by the computer from the pool of three possible exemplars, in order to encourage phonetic processing, and to discourage processing based on idiosyncratic acoustic properties. Subjects were required to listen to the randomly presented pairs and respond by pressing either the "same" or "different" key within 1000 msecs. (Due to some program differences maximum response times for Swedish subjects was 1500 msecs.) Finally, subjects completed two rating scales on the similarity of the sounds to speech and music. These revealed that for all three subject groups, speech was perceived as speech, music as music, and filtered speech as neither speech nor music. The number of correct and incorrect responses were converted to discrimination indices (DIs), given by {number correct on different (AB, or BA) trials (hits)} minus {number incorrect on same (AA or BB) trials (false positives)} / (total number of responses). RTs on different, AB and BA, trials were recorded.

Results

The results for the English, Thai and Cantonese subjects are presented first, followed by the in progress results for the Swedish speakers. DIs and RTs on AB trials were analysed in separate analyses of variance (ANOVAs): Language (English, Thai, Cantonese) x ISI (500, 1500 msec) x Tone Type (Speech, Filtered Speech, Music) x Tone Contrast (10 AB combinations), the last two with repeated measures. As results for the tone contrasts differed only slightly and as they are not of central concern here, specific results for this factor are not presented. The Swedish subjects' DIs and RTs were analysed in separate single factor 3-level (Speech, Filtered Speech, Music) ANOVAs.

Mean DIs are shown in Figure 1. Non-tonal language (English) speakers responded differently to the three stimulus types than did the tone language speakers, $F(1,138) = 25.48$. Simple main effects showed that tonal speakers, Thai and Cantonese, discriminated the three stimulus types equally well, and no differently from each other, while the English listeners discriminated Music better than Filtered Speech, $F(1,46) = 12.42$, and both of these better than Speech, $F(1,46) = 20.15$. Thus English listeners' pitch discrimination improves significantly once the sounds are not perceived as speech. The fact that the non-speech sounds contain exactly the same F_0 information as the speech suggests that the English listeners' attenuation for speech is attentional, in nature. While the lack of difference between the Thai and Cantonese speakers might suggest that the language of the tones is unimportant, inspection of RTs suggests otherwise.

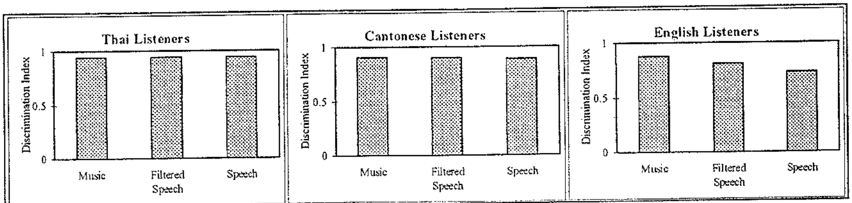

Figure 1. Thai, Cantonese, and English speakers' mean discrimination indices for music, filtered speech and speech

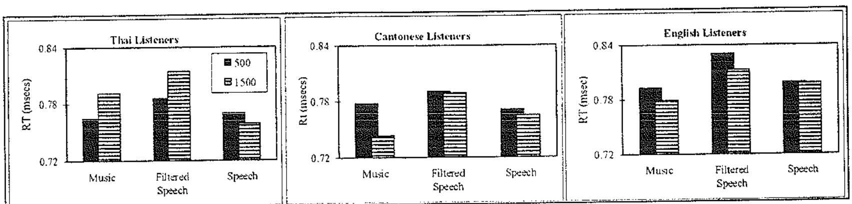

Figure 2: Thai, Cantonese, and English speakers' mean "different" RTs for music, filtered speech and speech

Figure 2 shows mean RTs on AB trials. There was a significant interaction of ISI x language (Thai/Cantonese) x type of stimulus (speech/non-speech), $F(1,121) = 4.46$. Simple main effects showed that for Thai speakers there was a significant ISI x speech/non-speech effect, $F(1,46) = 5.85$: Thais made speech discriminations better at 1500 msecs ISI (favouring phonemic level processing), but non-speech discriminations better at 500 msecs ISI (favouring phonetic/acoustic processing). On the other hand, neither Cantonese nor English speakers' RTs differed due to the speech/non-speech dimension - both actually had a significantly greater 1500 msec advantage for music than for filtered speech, $F(1,41) = 5.06$, and $F(1, 34) = 13.24$.

The results for the Swedish subjects are shown in Figure 3. The ANOVA revealed no significant difference between the three contexts, ie, their results were similar to those of the Thai and Cantonese rather than to the English speakers. ANOVA of their 500 msec condition RTs revealed generally faster RTs for music, a result more similar to English and Cantonese, than Thai speakers.

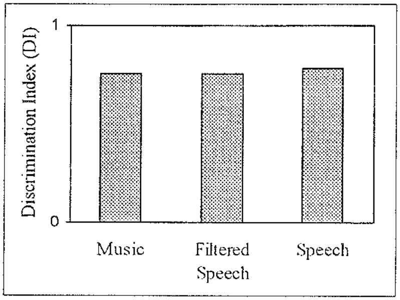

Figure 3. Swedish speakers' mean DIs and "different" RTs for music, filtered speech and speech (500 msec condition only)

Discussion

English speakers perceive pitch variations in filtered speech and music better than the same pitch variations presented as lexical tone in speech. Thus the perception of a basic psychoacoustic dimension, pitch, is compromised in linguistic contexts in which it is irrelevant. Tonal language speakers (Thai, Cantonese, and Swedish) performed equally well in all three contexts showing that tone perception development does not involve augmentation of underlying psychoacoustic abilities. However, the RT results for Thai speakers show that, within these limits, linguistic experience with a particular dimension, lexical tone, results in a subtle change in the *manner* in which phonologically relevant variations are perceived. This appears to be specific to the actual variations in the ambient language, as the differential RT results for Thai and Cantonese speakers demonstrate. It will be interesting to ascertain whether the same pattern holds when the Swedish 1500 msec condition results become available.

EXPERIMENT 2

Presumably the attenuation for tone perception in English speakers allows concentration on the important features of the ambient phonology, thus facilitating efficient linguistic processing. It is of interest to investigate whether such attenuation (and consequent efficiency) also occurs in childhood. Experiment 2 was conducted with English-speaking, and Thai-speaking children of 5, 6, and 8 years. Only one ISI condition, 500 msecs, was used.

Method

A total of 144 subjects were tested: 72 native English-speaking children, with little or no experience of other languages, and no experience of Thai or any tonal language; and 72 native Thai speaking children were. In each group 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, Filtered Speech, Speech) x Tone Contrast design with repeated measures on the last two factors was employed, with equal numbers of males and females in each sub-group. The apparatus and procedure were identical to that used in Experiment 1, with three exceptions. Firstly, to overcome attentional problems, the number of trials was halved, by testing children on only five (R-F, H-M, H-L, L-R, and L-F) of the 10 possible

AB contrasts. These five were chosen on the basis of the adults' data, taking care to include contrasts covering the range of discrimination ability. Secondly the maximum response time was increased to 2000 msec to allow the children more time for a response. Finally, only the 500 msec ISI condition was included.

Results

As only one ISI condition was tested, only the DI results are presented. For purposes of comparison, the Thai and English adults' results for just the five appropriate speech contrasts were incorporated and the DIs scores ananlysed in an Age (5, 6, 8, adults) x Tone Type (Music, Filtered Speech, Speech) x Tone Contrast ANOVA with repeated measures on the second two factors. As in Experiment 1, results for tone contrast are not reported here.

The results are presented in Figure 4. As can be seen all children improve on this task over age $F(1,184) = 75.01$ and children's data from the two language groups mirror those of their adult counterparts. There were significant interactions of Thai/English x Speech/Non-Speech $F(1,184) = 27.64$ and Thai/English x Music/FIltered Speech, showing that English subjects discriminated Music better than Filtered Speech, and this in turn, better than Speech. For Thai subjects however, discrimination ability was equivalent across all three stimulus types. There was a hint that children in general were more sensitive to stimulus differences in Non-Speech than adults (Children/Adults x Music/Filtered Speech, $F(1,184) = 8.78$), but that the Thai children soon showed similar results to their adult counterparts (Thai/English x Linear Age x Music/Filtered Speech, $F(1,184) = 6.24$).

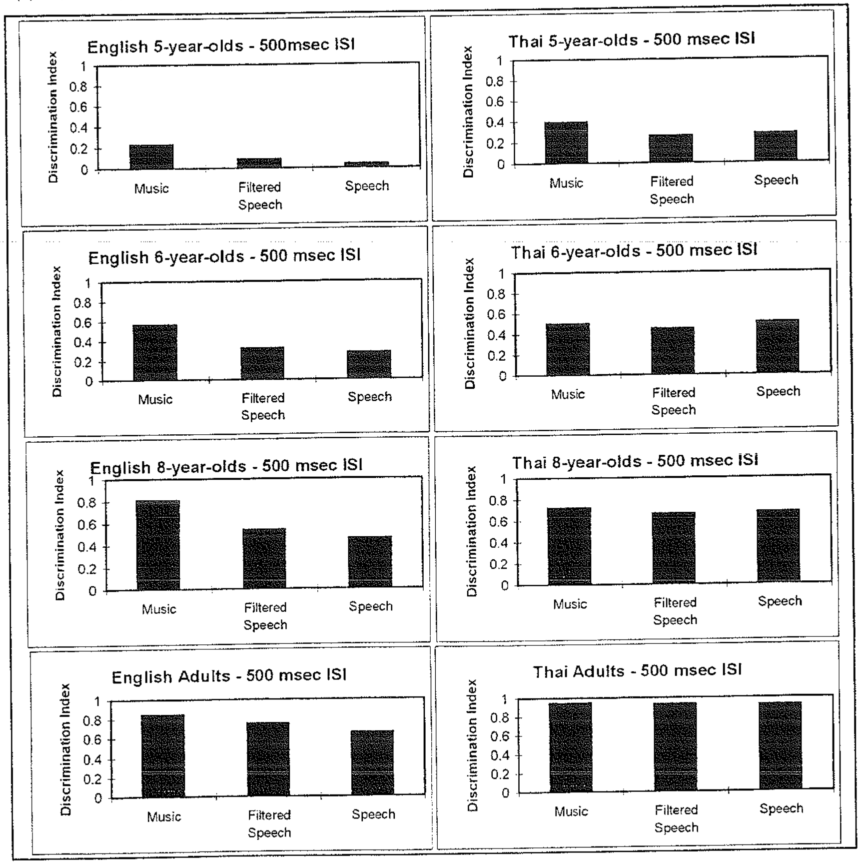

Figure 4. Discrimination indices on the three stimulus types for the four age groups

Discussion

The results of this second experiment mirror those of the first: speakers of a tone language have equivalent ability to discriminate pitch differences in speech, filtered speech, and music contexts, whereas speakers of a non-tonal language have attenuated ability for the discrimination of pitch in speech. Thus this attenuation appears to be established quite early, at least by reading age.

GENERAL DISCUSSION

The perception of lexical tone by tonal language speakers is equivalent in speech and non-speech environments, suggesting that tone perception occurs as a natural application of psychoacoustic pitch perception. However, in non-tonal language speakers, perhaps due to selective attention to periodic information (Cutler, 1994; Cutler at al., 1992) the perception of pitch in speech is attenuated below its normal level, as represented in filtered speech and music. This presumably allows non-tonal language speakers to focus on pitch (and segmental) features that *are* important in their language, and so further studies should reveal that non-tonal language speakers have no attenuation for more global pitch patterns (intonation) in sentences. Moreover, this attenuation is established early in life (at least by reading age) though exactly how early is yet to be determined. If the attenuation of pitch perception in non-tonal language speakers is due specifically to the linguistic environment, then it would be expected that even trained musicians who speak a non-tonal language should also show impaired pitch perception in a linguistic context. Again, this is yet to be tested.

REFERENCES

Best, C.T. 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.* York Press, Timonium, MD, in press.

Burnham, D. & Francis, E. The role of linguistic experience in the perception of Thai tones. In T. Thongkum (Ed) *SouthEast Asian Linguistic Studies in Honour of Vichin Panupong (Science of Language Vol. 8)* Bangkok, Chulalongkorn UP, 1995, in press.

Burnham, D. Developmental loss of speech perception: Exposure to and experience with a first language. *App. Psycholing, 7*, 206-40, 1986.

Burnham, D., Francis, E., & Webster, D. The development of tone perception: Cross-linguistic aspects and the effect of linguistic context. *In Pan-Asiatic Linguistics: Proceedings of the 4th International Symposium*, ILCRDMUS, Mahidol University, 1996.

Burnham, D.K., & Torstensson, C. Phonological bias and foreign accents: Perception and production of Swedish vowels and tones by English speakers. *ICPhS Proceedings*, Stockholm, *4*, 558-61, 1995.

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

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

Hirsh-Pasek, K., Kemler-Nelson, D.G., Jusczyk, P.W., Wright-Cassidy, K., Druss, B., & Kennedy, L. Clauses are perceptual units for young infants. *Cognition*, *26*, 269-86, 1987.

Jusczyk, P.W., Cutler, A., & Redanz, N.J. Infants' preference for the predominant stress patterns of English words. *Child Development*, *23*, 648-54, 1993.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606-8, 1992.

Polka, L. & Werker, J.F. Developmental changes in perception of nonnative vowel contrasts. *J. Exp. Psych: HPP*, *19*, 421-35, 1994.

Werker, J.F. The ontogeny of speech perception. In I.G. Mattingly & M. Studdert-Kennedy (Eds) Modularity and the motor theory of speech perception. Erlbaum, Hillsdale, N.J., 1991.

Werker, J.F., & Logan, J.S. Cross-language evidence for three factors in speech perception. *Percep.& Psychophys*, *37*, 35-44, 1985.