

PERCEPTION OF LEXICAL TONE ACROSS LANGUAGES: EVIDENCE FOR A LINGUISTIC MODE OF PROCESSING

Denis Burnham, Elizabeth Francis, Di Webster, Sudaporn Luksaneeyanawin, Chayada Attapaiboon, Francisco Lacerda, and Peter Keller

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

ABSTRACT

Pairs of Thai tones were presented for perceptual discrimination in three linguistic contexts (normal speech, low-pass filtered speech, and as musical (violin) sounds) to tonal language speakers, Thai and Cantonese, and non-tonal (English) language speakers. English speakers discriminated the tonal contrasts significantly better in the musical context than in filtered speech, and in filtered speech better than in full speech. On the other hand, both Thai and Cantonese speakers perceived the tonal contrasts equally well in all three contexts. Thus developmental absence of exposure to lexical tone results in a linguistic mode of processing which involves the attenuation of a basic psychoacoustic ability, pitch discrimination.

1. INTRODUCTION

Early in infancy most consonant contrasts, even those phonologically irrelevant in the ambient language, can be discriminated (1). From this universal beginning children become increasingly attuned to the structure of the ambient language. There appears to be two intensive periods of such selective reorganisation. The first is the second six months of the first year, which has been extensively researched by Werker and her colleagues (2); the second occurs soon after the onset of reading instruction (1). However, the attendant loss of speech perception ability that occurs in these periods 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 (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 (3).

More recently research has been conducted suggesting that perceptual reorganisation favouring ambient language characteristics also occurs for vowels (4,5), prosody (6), and the stress patterns of words (7). Rather less work has been conducted on the pitch variations of individual syllables or words, ie lexical tone. Cutler et al. (8) 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 in a regular fashion 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 (9).

In two recent studies involving the tonal languages Swedish (10) and Thai (11), Burnham and colleagues found that English-speaking children perceive non-native segmental contrasts (vowels and consonants) 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, specifically 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 (12) than temporal and/or spectral variations in segmental contrasts.

Despite these findings, it is possible that English-speaking adults' perception of tones still bears the vestiges of perceptual reorganisation. In Best's Perceptual Assimilation Model (12), two syllables differing only on lexical tone would be said to fall into the Same Category for English speakers. Such sounds, Best says, will be difficult to discriminate so long as they are perceived as speech. If, however, they are *not* perceived as speech, discrimination will be good to very good depending on their psychoacoustic salience.

The current study 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:] (13) were modified into two different non-speech formats: low-pass filtered speech, and musical (violin) sounds. By comparison of perceptual ability across stimulus types, the relative roles of attenuation of speech perception due to phonological irrelevance (for English speakers), and facilitation due to phonological relevance (for Thai and Cantonese speakers) can be determined.

2. METHOD

Subjects and Design. A total of 144 adults were tested: 48 native English-speakers, 48 native Thai speakers, and 48 native Cantonese speakers. A Language Background (English, Thai, Cantonese) x ISI (500, 1500 msec) x Tone Type (Music, Filtered Speech, Speech) x Tone Contrast (10) counterbalanced design

with repeated measures on the last two factors was employed. Half the subjects in each language group were tested at each ISI and, within these subgroups, half the subjects were males and half females.

Stimulus Materials and Apparatus. 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 [pá:], high [pā:], mid [pa:], low [pà:], falling [pâ:], (R, H, M, L, F, hereafter) [SOUND A995S01.WAV]. 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 [SOUND A995S02]. 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 [SOUND A995S03].

The experiment was conducted in parallel at the University of NSW in Australia, and Chulalongkorn University in Thailand 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.

Procedure. 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 (L/M, L/H, M/H, F/L, F/M, F/H, R/L, R/M, R/H, R/F) 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 msec. Finally, subjects completed two rating scales on the similarity of the sounds in each phase 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.

Dependent Variables. The number of correct and incorrect responses were converted to discrimination indices (DIs), given by $\{ \text{the number correct on different (AB, or BA) trials (hits)} \} - \{ \text{the number incorrect on same (AA or BB) trials (false positives)} \} / 4$ (total number of responses). RTs were recorded in msec.

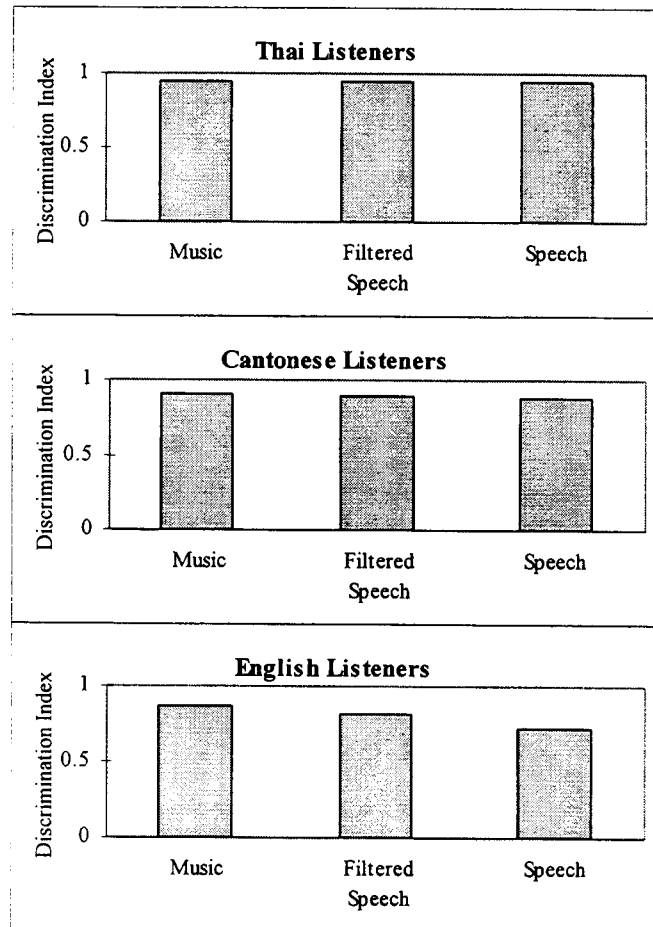


Figure 1: Thai, Cantonese, and English speakers' mean discrimination indices for music, filtered speech and speech.

3. RESULTS

DIs and RTs on AB trials were analysed in separate analyses of variance: 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.

Discrimination Indices. 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



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

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.

Reaction Times. 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 msec ISI (favouring phonemic level processing), but non-speech discriminations better at 500 msec 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 $F(1, 34) = 13.24$.

4. CONCLUSIONS

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. Presumably such attenuation is attentional and allows the listener to concentrate on the important features in their linguistic environment. Tonal language speakers performed equally well in all three contexts showing that tone perception development does not involve augmentation of underlying psychoacoustic abilities, which presumably have biological limits. 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.

Further studies could clarify these findings. Firstly, the influence of the tone variations of a specific language could be further tested with speakers of Swedish for example, which has a tonal inventory very disparate from the Thai target sounds. Secondly, if the attenuation of pitch perception 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.

5. REFERENCES

1. 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.
2. Burnham, D. Developmental loss of speech perception: Exposure to and experience with a first language. *App. Psycholing.* 7, 206-40, 1986.
3. Werker, J.F., & Logan, J.S. Cross-language evidence for three factors in speech perception. *Percep. & Psychophys.* 37, 35-44, 1985.
4. Polka, L. & Werker, J.F. Developmental changes in perception of nonnative vowel contrasts. *J. Exp. Psych: HPP.* 19, 421-35, 1994.
5. 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.
6. Hirsch-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.

