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 (F₀), 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 F₀ 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, Bertocnini, 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, Verspy, Paris-Delrue & Crepin, 1988), because such preferences are not evident when this information is removed from the speech signal (Mehler, Bertocnini, 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, Bertocnini & 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 months¹ (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 F₁ - F₂ 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 leaner 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 prevocal vs voiceless unaspirated stops, [b]-[p] (Burnham, 1986; Burnham, Earnshaw & Clark, 1991), plosive vs implosive voiced bilabial, [b]-[l], and alveolar, [d]-[l] stops (Bond & Adams, 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