Becoming a Native Listener

Janet F. Werker

The syllables, words, and sentences used in all human languages are formed from a set of speech sounds called phones. Only a subset of the phones is used in any particular language. Adults can easily perceive the differences among the phones used to contrast meaning in their own language, but young infants go much farther: they are able to discriminate nearly every phonetic contrast on which they have been tested, including those they have never before heard. Our research has shown that this broad-based sensitivity declines by the time a baby is one year old. This phenomenon provides a way to describe basic abilities in the young infant and explore the effects of experience on human speech perception.

To put infants' abilities in perspective, adult speech perception must be understood. The phones that distinguish meaning in a particular language are called phonemes. There is considerable acoustic variability in the way each individual phoneme is realized in speech. For example, the phoneme /b/ is very different before the vowel /e/ in "beat" from the way it is before the vowel /o/ in "boot." How do adults handle this variability? As first demonstrated in a classic study by Liberman and his colleagues (1967), they treat these acoustically distinct instances of a single phoneme as equivalent. This equivalency is demonstrated in the laboratory by presenting listeners with a series of pairs of computer-synthesized speech stimuli that differ by only one acoustic step along a physical continuum and asking them first to label and then to try to discriminate between the stimuli. Adult listeners are able to discriminate reliably only stimuli that they have labeled as different—that is, they cannot easily discriminate between two acoustically different stimuli that they labeled /pa/, but they can discriminate between two similar stimuli if one is from their /ba/ category and one from their /pa/ category.

The phenomenon by which labeling limits discrimination is referred to as categorical perception. This has obvious advantages for language processing. It allows a listener to segment the words he hears immediately according to the phonemic categories of his language and to ignore unessential variations within a category.

Given that adults perceive speech categorically, when do such perceptual capabilities appear? To find out, Eimas and his colleagues (1971) adapted the so-called high-amplitude sucking procedure for use in a speech discrimination task. This procedure involves teaching infants to suck on a pacifier attached to a pressure transducer in order to receive a visual or auditory stimulus. After repeated presentations of the same sight or sound, the sucking rate declines, indicating that the infants are becoming bored. The infants are then presented with a new stimulus. Presumably, if they can discriminate the new sight or sound from the old, they will increase their sucking rate.

In Eimas' experiments, infants one and four months old heard speech sounds that varied in equal steps from /ba/ to /pa/. Like adults, they discriminated between differences in the vicinity of the /ba/-/pa/ boundary but were unable to discriminate equal acoustic changes from within the /ba/ category. Rather than having to learn about phonemic categories, then, infants seem capable of grouping speech stimuli soon after birth.

Experiments in the 17 years since Eimas' original study have shown that infants can discriminate nearly every phonetic contrast on which they are tested but are generally unable to discriminate differences within a single phonemic category (for a review, see Kuhl 1987). That is, like adults, infants perceive acoustically distinct instances of a single phoneme as equivalent but easily discriminate speech sounds from two different categories that are not more acoustically distinct.

Of special interest are demonstrations that young infants are even able to discriminate phonetic contrasts not used in their native language. In an early study, Streeter (1976) used the high-amplitude sucking procedure to test Kikuyu infants on their ability to discriminate the English /ba/-/pa/ distinction; which is not used in Kikuyu. She found that the infants could discriminate these two syllable types. Similar results have been obtained from a variety of laboratories using other nonnative phonetic contrasts (Lasky et al. 1975; Trothub 1976; Adlin et al. 1989; Gillers et al. 1982). This pattern of results indicates that the ability to discriminate phones from the universal phonetic inventory may be present at birth.

Developmental changes

Given these broad-based infant abilities, one might expect that adults would also be able to discriminate nearly all phonetic contrasts. However, research suggests that adults often have difficulty discriminating phones that do not contrast meaning in their own language. An
English-speaking adult, for example, has difficulty perceiving the difference between the two /p/ phones that are used in Thai (Lisker and Abramson 1970). So too, a Japanese-speaking adult initially cannot distinguish between the English /a/ and /i/, because Japanese uses a single phoneme intermediate between the two English phonemes (Miyawaki et al. 1975; MacKain et al. 1981). This pattern of extensive infant capabilities and more limited capabilities in the adult led to the suggestion that infants may have a biological predisposition to perceive all possible phonetic contrasts and that there is a decline in this universal phonetic sensitivity by adulthood as a function of acquiring a particular language (Eimas 1975; Trehub 1976).

My work has been designed to explore this intriguing possibility. In particular, I wanted to trace how speech perception changes during development. Are infants actually able to discriminate some pairs of speech sounds better than adults, or have they simply been tested with more sensitive procedures? If infants do have greater discriminative capacities than adults, when does the decline occur and why?

The first problem that my colleagues and I faced was to find a testing procedure which could be used with infants, children of all ages, and adults. We could then begin a program of studies comparing their relative abilities to perceive the differences between phonetic contrasts of both native and nonnative languages.

The testing routine we chose is a variation of the so-called infant head turn procedure (for a complete description, see Kuhl 1987). Subjects are presented with several slightly different versions of the same phoneme (e.g., /ba/) repeated continuously at 2-sec intervals. On a random basis every four to twenty repetitions, a new phoneme is introduced. For example, a subject will hear ‘ba,’ ‘ba,’ ‘ba,’ ‘ba,’ ‘ba,’ ‘da,’ ‘da.’ Babies are conditioned to turn their heads toward the source of the sound when they detect the change from one phoneme to another (e.g., from ‘ba’ to ‘da’). Correct head turns are reinforced with the activation of a little toy animal and with clapping and praise from the experimental assistant. Figure 1 shows a baby being tested. Adults and children are tested the same way, except that they press a button instead of turning their heads when they detect a change in the phoneme, and the reinforcement is age-appropriate.

In the first series of experiments, we compared English-speaking adults, infants from English-speaking families, and Hindi-speaking adults on their ability to discriminate the /ba/-/da/ distinction, which is used in both Hindi and English, as well as two pairs of syllables that are used in Hindi but not in English (Werker et al. 1981). The two pairs of Hindi syllables were chosen on the basis of their relative difficulty. The first pair contrasts two ‘t’ sounds that are not used in English. In English, we articulate ‘t’ sounds by placing the tongue a bit behind the teeth at the alveolar ridge. In Hindi, there are...
two different "t" phonemes. One is produced by placing the tongue on the teeth (a dental t—written /t/). The other is produced by curling the tip of the tongue back and placing it against the roof of the mouth (a retroflex t—written /r/). This contrast is not used in English, and is in fact very rare among the world's languages.

The second pair of Hindi syllables involves different categories of voicing—the timing of the release of a consonant and the amount of air released with the consonant. Although these phonemes, called /p/ and /d/, are not used in English, we had reason to believe that they might be easier for English-speaking adults to discriminate than the /n-/ /t/ distinction. The timing difference between /p/ and /d/ spans the English /t/ /d/ boundary. Moreover, this contrast is more common among the world's languages.

The results of this study, which are presented in Figure 2, were consistent with the hypothesis of universal phonetic sensitivity in the young infant and a decline by adulthood. As expected, all subjects could discriminate /ba/ from /da/. Of more interest, the infants aged six to eight months performed like the Hindi adults and were able to discriminate both pairs of Hindi speech contrasts. The English-speaking adults, on the other hand, were considerably less able to make the Hindi distinctions, especially the difficult dental-retroflex one.

Timing of developmental changes

The next series of experiments was aimed at determining when the decline in nonnative sensitivity occurs. It was originally believed that this decline would coincide with puberty, when, as Lenneberg (1967) claims, language flexibility decreases. However, our work showed that twelve-year-old English-speaking children were no more able to discriminate non-English syllables than were English-speaking adults (Werker and Tees 1983). We then tested eight- and four-year-old English-speaking children, and, to our surprise, even the four-year-olds could not discriminate the Hindi contrasts. Hindi-speaking four-year-olds, of course, showed no trouble with this discrimination.

Before testing children even younger than age four, we felt it was necessary to determine that the phenomenon of developmental loss extended to other languages. To this end, we chose a phonemic contrast from a North American Indian language of the Interior Salish family, called Nthlakapmx by native speakers in British Columbia but also referred to as Thompson.

North American Indian languages include many consonants produced in the back of the vocal tract, behind our English /k/ and /g/. The pair of sounds we chose contrasts a "k" sound produced at the velum with another "k" sound (written /q/) produced by raising the back of the tongue against the uvula. Both are glottalized—that is, there is an ejective portion (similar to a click) at the beginning of the release of the consonants.

Again, we compared English-speaking adults, infants from English-speaking families, and Nthlakapmx-speaking adults in their abilities to discriminate this pair of sounds (Werker and Tees 1984a). As was the case with the Hindi syllables, both the Nthlakapmx-speaking adults and the infants could discriminate the non-English phonemes, but the English-speaking adults could not.

If infants do have greater discriminative capacities than adults, when does the decline occur and why?

We were now satisfied that there is at least some generality to the notion that young infants can discriminate across the whole phonetic inventory but that there is a developmental decline in this universal sensitivity. Our next series of experiments involved testing children between eight months and four years of age to try to determine just when the decline in sensitivity might start. It quickly became apparent that something important was happening within the first year of life. We accordingly compared three groups of infants aged six to eight, eight to ten, and ten to twelve months. Half of each group were tested with the Hindi (/tai-/ /Tai/) and half with the Nthlakapmx (/k/-/q/) contrast.

As shown in Figure 3, the majority of the six-to-eight-month-old infants from English-speaking families could discriminate the two non-English contrasts, whereas only about one-half of the eight-to-ten-month-olds could do so. Only two out of ten ten-to-twelve-month-olds could discriminate the Hindi contrast, and only one out of ten the Nthlakapmx. This provided strong evidence that the decline in universal phonetic sensitivity was occurring between six and twelve months of age. As a further test to see if this developmental change would be apparent within the same individuals, six infants from English-speaking families were tested at two-month intervals beginning when they were about six to eight months old. All six infants could discriminate both the Hindi and Nthlakapmx contrasts at the first testing, but by the third testing session, when they were ten to twelve months old, they were not able to discriminate either contrast.

To verify that the decline in nonnative sensitivity around ten to twelve months was a function of language experience, we tested a few infants from Hindi- and Nthlakapmx-speaking families when they reached eleven to twelve months old. As predicted, these infants were still able to discriminate their native contrasts, showing quite clearly that the decline observed in the infants from English-speaking families was a function of specific language experience. Since doing these studies, we have charted the decline between six and twelve months old using a computer-generated set of synthetic syllables which model another pair of Hindi sounds not used in English (Werker and Lalonde 1988).

Janet F. Werker is an assistant professor of developmental psychology at the University of British Columbia. She received a B.A. from Harvard University and a Ph.D. in psychology from the University of British Columbia in 1982. She was an assistant professor of psychology at Dalhousie University from 1982 to 1986, and joined the faculty at the University of British Columbia in 1986. The research reported here was supported by the Natural Sciences and Engineering Research Council of Canada and the Social Science and Humanities Research Council. In addition, portions of this work were made possible by an NICHD grant to Hadámá Laboratories. Address: Department of Psychology, University of British Columbia, Vancouver, British Columbia, V6T 1Y7, Canada.
How does experience affect development?

A theoretical model for considering the possible effects of experience on perceptual development was suggested by Gottlieb in 1976. As expanded by both Gottlieb (1981) and Aslin (1981), the model includes several roles experience might—or might not—play, as shown in Figure 4.

Induction refers to cases in which the emergence and form of a perceptual capability depend entirely on environmental input. In this case, an infant would not show categorical perception of speech sounds without prior experience. Attunement refers to a situation in which experience influences the full development of a capability, enhancing the level of performance; for example, categorical boundaries between phonetic contrasts might be sharper with experience than without. In facilitation, experience affects the rate of development of a capability, but it does not affect the end point. If this role were valid, speech perception would improve even without listening experience, but hearing specific sounds would accelerate the rate of improvement. Maintenance/loss refers to the case in which a perceptual ability is fully developed prior to the onset of specific experience, which is required to maintain that capability. Without adequate exposure an initial capability is lost. Finally, maturation refers to the unfolding of a perceptual capability independent of environmental exposure. According to this hypothetical possibility, the ability to discriminate speech sounds would mature regardless of amount or timing of exposure.

Our work is often interpreted as an illustration of maintenance/loss, since it suggests that young infants can discriminate phonetic contrasts before they have gained experience listening but that experience hearing the sounds used in their own language is necessary to maintain the ability to discriminate at least some pairs of phones.

Support for this view was provided by another study in which we tested English-speaking adults who had been exposed to Hindi during the first couple of years of life and had learned their first words in Hindi but had little or no subsequent exposure. These subjects could discriminate the Hindi syllables much more easily than other English-speaking adults, and performed virtually as well as native Hindi speakers on the discrimination task (Tees and Werker 1984). This is consistent with the view that early experience functions to maintain perceptual abilities, suggesting that no further experience is necessary to maintain them into adulthood.

Recovery of sensitivity

Our early work led us to believe that the loss of nonnative sensitivity is difficult to reverse in adults. In one study, we tested English-speaking adults who had studied Hindi for various lengths of time. Adults who had studied Hindi for five years or more were able to discriminate the non-English Hindi syllables, but those who had studied Hindi for one year at the university level could not do so. In fact, even several hundred trials were insufficient to teach English-speaking adults to discriminate the more difficult Hindi contrasts (Tees and Werker 1984). This implies that while the ability is recoverable,
considerable experience is required. Similar conclusions can be drawn from a study by MacKain and her colleagues (1981), who tested Japanese speakers learning English. Only after one year of intensive English training in the United States could they discriminate /ra/ from /la/.

The question still remained whether recovery of normative sensitivity results from new learning in adulthood or from a latent sensitivity. To explore this question, we asked English-speaking adults to discriminate both the full syllables of the difficult Hindi and Nthlaprox phonemes and shortened portions of the syllables which do not sound like speech at all but contain the critical acoustic information specifying the difference between the phonemes (Werker and Tees 1984b). Subjects were first tested on the shortened stimuli and then on the full syllables. To our surprise, they were able to discriminate the shortened stimuli easily but were still not able to discriminate the full syllables, even immediately after hearing the relevant acoustic information in shortened form. This finding reveals that the auditory capacity for discriminating the acoustic components of these stimuli has not been lost but that it is difficult to apply when processing language-like sounds.

In a further set of experiments, we attempted to make English-speaking adults discriminate the full-syllable normative stimuli (Werker and Logan 1985). One task involved presenting adults with pairs of stimuli and asking them to decide simply if the stimuli were the same or different, a test that proved to be much more sensitive than the head turn procedure. In this "same/different" task, listeners have to compare only two stimuli at a time. Moreover, if the interval between the two stimuli is short enough, listeners can hold the first stimulus in auditory memory while comparing it to the second. In the head turn task, on the other hand, listeners have to compare each stimulus to a whole set of variable stimuli and judge whether it is a member of the same category.

We found that English-speaking adults could discriminate the Hindi syllables when tested in the same/different procedure, particularly after practice. Thus there was evidence that adults can discriminate normative contrasts if tested in a more sensitive procedure. Similar results have been reported by other researchers (Pisoni et al. 1987). This suggests that the developmental changes between infancy and adulthood should be considered a language-based reorganization of the categories of communicative sounds rather than an absolute loss of auditory sensitivity. The increasing reliance on language-specific categories accounts for the age-related decline, implying that maintenance has its effect at the level of linguistic categories rather than simple peripheral auditory sensitivity (see Best et al. 1988).

Parallels in speech production

It is interesting to compare our findings of developmental changes in speech perception to recent work on speech production. Although it is impossible to survey this substantial literature here, there appear to be systematic regularities in the repertoire of sounds produced at different stages of babbling. These regularities may reflect vocal tract and neuromuscular maturation, with phones appearing as a child develops the ability to articulate them (Locke 1983). In contrast to early work suggesting that the sounds produced during babbling gradually narrow to those that are used in the language-learning environment, recent research shows very little influence from the native language on vocal develop-
ment during the babbling stage. This conclusion is particularly strong for consonants. However, it is clear that after the acquisition of the first word children’s vocal productions start becoming differentiated on the basis of language experience. That is, once a child begins to talk, the sounds produced conform more and more closely to the subset of phones used in his native language. The stage at which these changes occur is consistent with our work showing universal sensitivity in early infancy followed by only language-specific sensitivity beginning around ten to twelve months.

This leads us to believe that just as a reorganization of language production is related to the emergence of the first spoken word, so too the reorganization of perceptual abilities may be related to the emergence of the ability to understand words. By the time he is one year old, a child understands a fair amount of spoken language, even though he may produce only a few words. We are currently conducting experiments to see if the reorganization of speech perception is related to the emerging ability to understand words. This work will add another piece to the solution of the puzzle of how early sensitivity to all language sounds becomes limited to the functional categories that are necessary for communicating in one’s own language.

References


