Developmental Science

Developmental Science 11:1 (2008), pp 122–134 DOI: 10.1111/j.1467-7687.2007.00653.x

PAPER

Statistical phonetic learning in infants: facilitation and feature generalization

Jessica Maye,[1] Daniel J. Weiss[2] and Richard N. Aslin[3]

1. Department of Communication Sciences & Disorders and the Northwestern Institute on Complex Systems, Northwestern University, USA
2. Department of Psychology, The Pennsylvania State University, USA
3. Department of Brain & Cognitive Sciences, University of Rochester, USA

Abstract

Over the course of the first year of life, infants develop from being generalized listeners, capable of discriminating both native and non-native speech contrasts, into specialized listeners whose discrimination patterns closely reflect the phonetic system of the native language(s). Recent work by Maye, Werker and Gerken (2002) has proposed a statistical account for this phenomenon, showing that infants may lose the ability to discriminate some foreign language contrasts on the basis of their sensitivity to the statistical distribution of sounds in the input language. In this paper we examine the process of enhancement in infant speech perception, whereby initially difficult phonetic contrasts become better discriminated when they define two categories that serve a functional role in the native language. In particular, we demonstrate that exposure to a bimodal statistical distribution in 8-month-old infants' phonetic input can lead to increased discrimination of difficult contrasts. In addition, this exposure also facilitates discrimination of an unfamiliar contrast sharing the same phonetic feature as the contrast presented during familiarization, suggesting that infants extract acoustic/phonetic information that is invariant across an abstract featural representation.

Introduction

The perception of speech sounds changes drastically between birth and adulthood. In particular, the phonetic and phonological system of the native language influences adults' perception of speech. When adults listen to foreign languages they often have trouble hearing the difference between certain pairs of sounds in the foreign language. Specifically, when two foreign sounds are perceived as sounding like two different sounds of the native language (e.g. the ejective sounds [p'] and [t'] in Ethiopian are heard as /p/ and /t/ by English speakers: Best, McRoberts & Goodell, 1990), listeners find it easy to discriminate the sounds (Best, McRoberts & Sithole, 1988; Abramson & Lisker, 1970). However, when two foreign sounds are both perceived as sounding like a single native language sound category (e.g. the dental [d] and retroflex [d̢] of Hindi are both heard as /d/ by English speakers: Werker, Gilbert, Humphrey & Tees, 1981), discrimination is often very poor (Best *et al.*, 1988; Abramson & Lisker, 1970). Language-specific perceptual difficulties of this sort are common and well documented; to give a few more examples (of which there are many), English [r] vs. [l] is difficult for native Japanese speakers to discriminate (Goto, 1971), Zulu plosive [b] vs. implosive [ɓ] is difficult for native English speakers (Best *et al.*, 1990), and Catalan [e] vs. [ɛ] is difficult for native Spanish speakers (Pallier, Bosch & Sebastián-Gallés, 1997). The fact that speakers of different languages exhibit different discrimination patterns reflects differences in the sound systems of the languages.

In the first months of life, before infants have gained sufficient experience with the native language, discrimination of speech sounds appears to be universal, showing little or no effects of the native language sound system. For example, at 6 months English infants can discriminate Hindi dental [d] vs. retroflex [d̢] (Werker *et al.*, 1981) and Zulu plosive [b] vs. implosive [ɓ] (Best, 1990), and Japanese infants can discriminate English [r]–[l] (Kuhl, Stevens, Hayashi, Deguchi, Kiritani & Iverson, 2006); and at 4 months Spanish infants can discriminate the [e]–[ɛ] vowel contrast that is used in Catalan but not in Spanish (Bosch & Sebastián-Gallés, 2003).[1] But by the

---

[1] Werker *et al.* (1981), Best *et al.* (1990), and Kuhl *et al.* (2006) did not test infants at 4 months, and Bosch and Sebastián-Gallés (2003) did not test infants at 6 months. However, previous research suggests that development of language-specific perception of vowels precedes that of consonants by approximately 2 months (Polka & Werker, 1994).

Address for correspondence: Jessica Maye, Department of Communication Sciences & Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208-2952, USA; e-mail: j-maye@ northwestern.edu

9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

end of the first year infants' perception of speech sounds is markedly affected by the native sound system, and many foreign language contrasts are no longer discriminated: at 10–12 months English infants no longer discriminate Hindi [d] vs. [ɖ] (Werker & Tees, 1984) or Zulu plosive [b] vs. implosive [ɓ] (Best *et al.*, 1990), and Japanese infants no longer discriminate English [r]–[l] (Kuhl *et al.*, 2006; Tsushima, Takizawa, Sasaki, Shiraki, Nishi, Kohno, Menyuk & Best, 1996); and at 8–9 months Spanish infants no longer discriminate Catalan [e]–[ɛ] (Bosch & Sebastián-Gallés, 2003). These changes in infant speech perception have been documented for many types of phonetic contrasts (for review see Jusczyk, 1997), and they reflect infants' development of a language-specific pattern of discrimination.

Phonetic contrasts that are discriminated well by young infants remain well discriminated if they correspond to a phonemic contrast in the native language (i.e. the difference between the two sounds can result in a difference in word meaning, as in the words *rice* vs. *lice*), while those that do not correspond to a native phonemic contrast generally become poorly discriminated (but see Best, McRoberts, LaFleur & Silver-Isenstadt, 1995, for discussion of non-phonemic contrasts that remain well discriminated as well as different assimilation patterns between foreign and native phoneme categories). These two patterns of developmental speech perception were discussed by Aslin and Pisoni (1980) in their description of the trajectories that perceptual development might follow. For initially discriminable contrasts, development may result in *loss* of sensitivity to the contrast (e.g. reduced discrimination of Hindi dental vs. retroflex stops by English-learning infants), or *maintenance* (e.g. continued discrimination of dental vs. retroflex stops by *Hindi*-learning infants). In other words, exposure to a language in which a contrast is phonemic results in maintenance of the distinction, while exposure to a language in which the contrast is nonphonemic leads to decreased sensitivity to the contrast. Outside of the linguistic domain, this same pattern of non-specific sensitivity followed by either maintenance or loss has also been shown for face perception (Pascalis, de Haan & Nelson, 2002). Six-month-old infants are able to discriminate different exemplars of both human and monkey faces, whereas 9-month-olds only show discrimination of human faces. This pattern appears to reflect an initially general, non-species-specific sensitivity to faces at 6 months followed by a loss of sensitivity to the non-experienced monkey faces. When infants are presented with monkey faces during the 3-month period in which they would ordinarily lose sensitivity (i.e. between 6 and 9 months of age), their discriminative ability for monkey faces is maintained (Pascalis, Scott, Kelly, Shannon, Nicholson, Coleman & Nelson, 2005).

A third possible pattern in perceptual development discussed by Aslin and Pisoni (1980) is the *enhancement* of contrasts that are initially discriminated more poorly.[2] Although infants' prodigious ability to discriminate phonetic contrasts has received much attention, certain phonetic contrasts are more difficult for young infants to discriminate. Among these are several fricative contrasts ([s]–[z]: Eilers & Minifie, 1975; Eilers, 1977; [f]–[ʃ]: Eilers, Wilson & Moore, 1977; [f]–[θ]: Eilers *et al.*, 1977; Holmberg, Morgan & Kuhl, 1977; though not all fricative contrasts are difficult for infants: see Trehub, 1976; Eilers *et al.*, 1977; Holmberg *et al.*, 1977), the contrast between prevoiced and short-lag stop consonants, such as the [b]–[p] contrast of French or Spanish (Aslin, Pisoni, Hennessy & Perey, 1981), and between English alveolar and interdental consonants (e.g. [d] vs. [ð]: Polka, Colantonio & Sundara, 2001). For phonetic contrasts that are not initially particularly salient, infants show poor discrimination performance, whereas such contrasts are discriminated well by adults if they are phonemic in the native language (Abramson & Lisker, 1970; Polka *et al.*, 2001; Narayan, 2006). Thus, native language input is sufficient to enhance discrimination by late infancy for some contrasts (Eilers *et al.*, 1977; Kuhl *et al.*, 2006), although enhancement may take longer for certain initially difficult contrasts (Polka *et al.*, 2001; Sundara, Polka & Genesee, 2006).[3]

There may be multiple reasons why some phonetic contrasts are initially more or less difficult to discriminate than others, but some of the variation in difficulty undoubtedly arises from differences in the basic psychoacoustic salience of the contrast in question (see also Burnham's [1986] discussion of robust vs. weak contrasts). Nonlinearities in psychoacoustic space result in the human auditory and perceptual system being more sensitive to some acoustic changes than to others (Stevens, Volkman & Newman, 1937; Jusczyk, Rosner, Cutting, Foard & Smith, 1977; Pisoni, 1977; Cutting & Rosner, 1974; Steinschneider, Volkov, Noh, Garell & Howard, 1999). For example, many languages use the acoustic cue of voice-onset time (VOT) to differentiate between voiced and voiceless stop consonants such as /b/ vs. /p/. These two sounds differ with respect to how much time passes between when the lip closure is released and when the vocal cords begin to vibrate to produce the following vowel sound. Voiceless sounds such as /p/ have a longer lag between closure release and the onset of vocal cord vibration than do voiced sounds, so voiceless sounds are said to have a longer VOT. Young infants exhibit better discrimination at certain regions of a VOT continuum

---

[2] Aslin and Pisoni distinguish between *enhancement*, for contrasts that are initially discriminated poorly, and *induction*, for contrasts that are initially not at all discriminable. In the current study our focus was on *any* improvement in discrimination, and so we have not differentiated between enhancement and induction.

[3] Pisoni *et al.* (1982) found that adult English speakers could learn to identify prevoiced stops rather quickly if the category was given a label ('mba' vs. 'ba' vs. 'pa'), suggesting that the enhancement of difficult contrasts is not limited to infants.

than others (Eimas, Siqueland, Jusczyk & Vigorito, 1971; Aslin *et al.*, 1981), suggesting that humans are born with a heightened sensitivity in particular VOT regions. The fact that certain non-human animals have been found to show the same sensitivity profile in discriminating VOT in speech sounds (Kuhl & Miller, 1975; Sinex & MacDonald, 1989; Steinschneider, Schroeder, Arezzo & Vaughan, 1996) suggests that this innate nonlinearity in VOT perception is nonlinguistic in nature. This conclusion is further confirmed by the fact that human listeners show a similar pattern of heightened sensitivity in the discrimination of nonspeech stimuli that vary along timing continua analogous to VOT (Miller, Weir, Pastore, Kelly & Dooling, 1976; Pisoni, 1977). Additional studies that have found human-like nonlinearities in the discrimination of different kinds of speech stimuli by various non-human animals (e.g. Morse & Snowdon, 1975; Kuhl & Padden, 1983; Dooling, Best & Brown, 1995) suggest that human language may capitalize on nonlinearities in perception by placing many phonetic contrasts at regions of natural psychoacoustic sensitivity (Kuhl, 1978; Burnham, Earnshaw & Quinn, 1987; Stevens, 1972).

Previous research has shown that phonetic contrasts are mirrored in the distribution of speech sounds produced in a language (Lisker & Abramson, 1964; Magloire & Green, 1999; Sundberg & Lacerda, 1999; Newman, Clouse & Burnham, 2001; Lotto, Sato & Diehl, 2004). For example, in Hindi there are three voicing categories – prevoiced [d], unvoiced [t], and aspirated [t^h] – while in English there are only two. This fact is corroborated by the distribution of stop consonant VOT values produced by native speakers of each language: for Hindi there is a trimodal distribution (the most commonly produced exemplars form three clusters, corresponding to the three Hindi voicing categories), while in English the distribution is bimodal (Lisker & Abramson, 1964; see Figure 1). The difference between the distribution of VOT values in the two languages is also reflected in the discrimination abilities of speakers of the languages. Speakers of a three-category language like Hindi show good discrimination of sounds that cross the 0 msec VOT boundary, while English speakers' discrimination is poor throughout this region (Abramson & Lisker, 1970). Thus, when two regions of the VOT continuum are used contrastively in a language (i.e. when producing a VOT from one region versus the other results in the production of a different word meaning), those two regions will correspond to different modes in the frequency distribution of VOT values produced in the language.

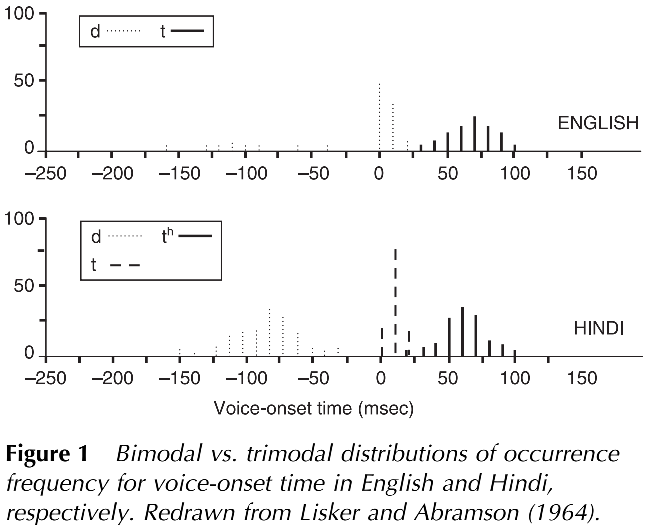

Figure 1 *Bimodal vs. trimodal distributions of occurrence frequency for voice-onset time in English and Hindi, respectively. Redrawn from Lisker and Abramson (1964).*

If infants are able to track this distributional information, then these cues might contribute to developmental changes in speech perception. If this is true, then exposure to a bimodal distribution of sounds should result in enhanced discrimination, while exposure to a unimodal distribution should result in reduced discrimination. Maye *et al.* (2002) provided evidence supporting the latter half of this hypothesis. Namely, they found that two sounds that are discriminable in early infancy were no longer discriminated by infants who had been familiarized to the sounds within a unimodal distribution. This pattern is analogous to Japanese infants' decreased discrimination of [r]~[l] following exposure to Japanese, a language which exemplifies a unimodal distribution of sounds in this phonetic region (Lotto *et al.*, 2004). In the current study, our goal was to test whether the phonetic distribution of the input might also have the converse effect, analogous to English-learners' ultimate improvement at discriminating the English [d]~[ð] contrast. That is, does exposure to a bimodal distribution result in enhanced discrimination of an initially difficult phonetic contrast? In Experiment 1 we asked whether statistical properties of the input speech are sufficient to enhance infants' discrimination of a difficult phonetic contrast. In Experiment 2 we investigated the nature of this facilitation by testing infants' discrimination of a different but featurally related phonetic contrast.

Experiment 1

The goal of this first experiment was to determine whether exposure to a bimodal distribution of speech tokens would enhance discrimination of a contrast previously shown to be difficult for infants to discriminate; namely, the contrast between prevoiced vs. short-lag stop consonants (Aslin *et al.*, 1981). Several studies have investigated infant discrimination of prevoiced vs. short-lag stops, finding mixed results. Studies of young infants (under 6 months) raised in Kikuyu- and Spanish-speaking environments (two languages which utilize the prevoiced/short-lag distinction) have found evidence of discrimination (Lasky, Syrdal-Lasky & Klein, 1975; Streeter, 1976), while infants from English-speaking homes have failed to discriminate this contrast (Eimas *et al.*, 1971; Eimas, 1975; Lasky *et al.*, 1975). Aslin *et al.*

(1981) conducted a detailed investigation of voicing discrimination in infants from English-speaking homes between the ages of 6 and 12 months and found that although infants could make discriminations in both the voicing lead and voicing lag regions, discrimination was considerably better in the lag region where their native language phonemic contrast resides.

We predicted that familiarization to the stimuli presented according to a bimodal frequency distribution would improve infants' discrimination. To ensure that any improvement in discrimination was due to the particular distribution of familiarization rather than to simple exposure to the stimuli, we familiarized some infants to a unimodal frequency distribution of the same stimuli, predicting that discrimination should not be facilitated for these infants. To assess baseline performance at discriminating these stimuli in the absence of prior exposure we familiarized a third group of infants to an unrelated sequence of tones.

Method

Participants

One hundred and fifty-three 8-month-old infants participated in the study. There were 81 males and 72 females whose average age was 8 months and 11 days (ranging from 7 months, 11 days to 9 months, 3 days). Fifty-six of these infants were excluded from analysis due to crying or fussing (23), technical problems (6), exposure to a native language other than English (4), parental interference (5), failure to habituate after 25 trials (4), failure to produce any usable test trials due to looking times < 2 sec on both change trials (1), dishabituation that differed by more than 2 standard deviations from the mean of their familiarization condition (8), failure to dishabituate to the posttest stimulus (4), or falling asleep (1). Twenty-three of the excluded infants were in the Bimodal familiarization condition, 15 were in the Control condition, and 18 were in the Unimodal familiarization condition (conditions are described in detail below). Infants were from English-speaking homes and were recruited based on parental interest in research participation.

Stimuli

We recorded multiple natural tokens of the syllables [da] and [ga] (both prevoiced), and [ta] and [ka] (both unaspirated), as produced by a male speaker of Hindi, at a sampling rate of 44,100 Hz. Four tokens of each of the unaspirated syllables were chosen, from which four dental and four velar continua were made. These syllables were then digitized and edited using SoundEdit 16.2. From the voiceless tokens we removed portions of the unvoiced lag (between release burst and onset of voicing), to create tokens with 0, 7, 14, and 21 msec voicing lag. We then spliced naturally produced prevoicing (from [da] and [ga]) onto the 0 msec lag tokens to create prevoiced stimuli with −100, −75, −50, and −25 msec voicing lead.[4] The result was eight 8-step voicing continua: four continua from [da] to [ta], and four from [ga] to [ka]. Stimuli varied from 299 msec to 423 msec in total duration, with an average length of 338 msec.

Design

Infants were randomly assigned to one of three conditions which differed with respect to the auditory stimuli they were familiarized to before the discrimination test. Thirty-two infants were familiarized to the stimuli from the phonetic continuum presented according to a bimodal frequency distribution (the Bimodal condition), such that stimuli near the endpoints were presented the most frequently, while stimuli from the center of the continuum were presented infrequently (see Figure 2). Thirty-four infants were familiarized to the same stimuli but presented according to a unimodal frequency distribution (the Unimodal condition), such that stimuli from the center of the continuum were presented the most frequently (see Figure 2). In order to determine baseline discrimination of the stimuli in the absence of any prior exposure, 31 infants were tested in a third condition in which they were familiarized to irrelevant auditory stimuli (a repeating sequence of tones) prior to the discrimination test (the Control condition). Tones rather than some other distribution of speech sounds were used in this control condition to ensure that infants were neither primed to attend to the post-habituation speech contrast

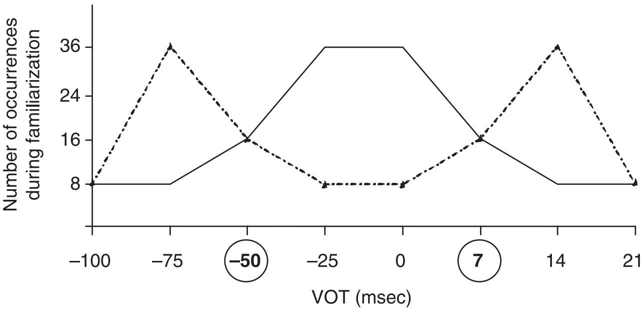

Figure 2 *Presentation frequency for continuum stimuli during familiarization phase. The frequency distribution for the Bimodal (Experiment 1) and Generalization (Experiment 2) conditions is shown by the broken line; the frequency distribution for the Unimodal condition (Experiment 1) is shown by the solid line. VOT values of −50 and 7 msec were used during the test phase.*

---

[4] Our continua were intended to represent the voicing values produced in natural speech in a language with a prevoiced vs. short-lag stop contrast. Thus, as in natural language, there is a larger degree of variation in the prevoiced range (in our experiment, 25 ms steps) than in the lag range (7 ms steps).

nor given distracting speech information that may have interfered with their discrimination. Moreover, a no-familiarization control was not employed because it would not balance the duration and amount of auditory stimulation preceding the habituation and test phases. Following familiarization, all three groups of infants were tested on their discrimination of stimuli from different sides of one of the two continua (da/ta or ga/ka).

Procedure

The experiment was conducted in two phases: familiarization and test. For the duration of the experiment, infants were seated on a parent's lap in front of a 15″ video monitor (ViewSonic ViewPanel VG150) in an IAC single-wall sound-attenuated chamber. Auditory stimuli were presented to the infant through a speaker (Boston Acoustics CR6) located below the video monitor. Parents listened to masking music through headphones (Peltor Headset). The experimenter testing the infant sat outside of the booth and viewed the infant's behavior on an 8″ black and white TV monitor by means of a video camera located inside the booth (Sony HyerHAD Hi Resolution), but could not hear any of the auditory stimuli. The experiment was controlled by a Macintosh G4 computer, using a custom program written for infant looking time experiments. The program files for the various familiarization and test conditions were stored under code names such that the experimenter was blind to each infant's assigned condition.

Infants in the Bimodal and Unimodal conditions were each familiarized to all 32 tokens of the experimental stimuli (i.e. eight tokens from each of the four continua) at one place of articulation (half of the infants in each condition were familiarized to the [da]~[ta] stimuli, the other half to [ga]~[ka]). Stimuli were presented in random order according to a bimodal or unimodal frequency distribution (see Figure 2), with an inter-stimulus interval of 1 second. The full set of familiarization stimuli comprised 64 syllables (given that certain tokens were presented multiple times, according to the appropriate frequency distribution). Infants in the Bimodal and Unimodal conditions were familiarized to two blocks of familiarization stimuli, for a total familiarization time of 170 seconds. Infants in the Control condition were familiarized to irrelevant auditory stimuli during familiarization, consisting of a repeating sequence of tones, each 330 msec long with no pauses between tones, for a total familiarization time of 170 seconds. Infants in all three conditions watched a silent cartoon video clip on the video monitor while listening to familiarization stimuli. The presentation of familiarization stimuli was not contingent on infants' looking behavior.

Infants were then tested on their discrimination of the same contrast that they heard during familiarization (i.e. [da]–[ta] or [ga]–[ka]; for infants in the Control condition, half were tested on the dental contrast, half on the velar contrast). Discrimination was tested through a habituation procedure, in which the dependent measure was the recovery of looking time from the final two habituation trials to the two test trials.[5] On each habituation trial, a colorful bullseye appeared on the video monitor, and the infant heard the four +7 msec lag tokens (token 6 from the continuum), from the place of articulation heard during familiarization, presented in random order.[6] The speech stimuli continued to play until the infant looked away from the screen for 2 seconds (up to a maximum of 60 seconds trial length), at which point the trial ended. Habituation was assessed via a moving window that compared the infant's looking time for the first three trials to that of each subsequent set of three trials. The habituation criterion was satisfied when a window was reached in which looking time was at or below 50% of the initial window. Infants who did not meet the habituation criterion within 25 trials were excluded from analysis. Following habituation, two change trials were presented. The change trials were identical to the habituation trials, except that the auditory stimulus was from the other side of the continuum, with a VOT of −50 msec (token 3 from the continuum). Experimenters were blind with respect to when the infants had reached the habituation criterion and when they were presented with the change stimuli. Importantly, the continuum tokens presented during the habituation phase (token 6) and test phase (token 3) occurred with equal frequency during the familiarization phase in both the Unimodal and Bimodal conditions. Discrimination is indicated by an increase in looking time between the average of the two final habituation trials and the average of the two change trials. Following the two change trials, each infant was

---

[5] Maye *et al.* (2002) tested infant discrimination using a preference procedure during which each infant listened to the two test conditions (single phoneme vs. two alternating phonemes) in a paradigm in which trial length was controlled by the infant. Thus, during test each infant received slightly different exposures to the two stimulus types. In the present habituation procedure, each infant listened to a series of trials with one stimulus type, followed by two trials with a different stimulus type, and the duration of all trials was infant-controlled. Thus, there were more stimulus exposures during test in the habituation than in the preference procedure, and therefore more variability across infants. However, the overall duration of exposure did not differ among the three conditions of Experiment 1. Moreover, an advantage of the habituation procedure is that all infants must meet a criterion of response decrement prior to the change trials, whereas in the preference procedure the listening times are dependent on spontaneous differences in attention to one class of stimuli over another. It is possible for an infant to be able to discriminate two stimuli but yet have no preference for one over the other. We chose to use a habituation design in the present study because it affords a more direct test of discrimination, despite the fact that this sacrifices an element of control over the shape of the frequency distribution presented during the test phase.

[6] Aslin *et al.* (1981) found that discrimination was poorer when the background stimulus was short lag (e.g. [ta]–[da]) than when the background stimulus was prevoiced (e.g. [da]–[ta]). Knowing that order of presentation would affect the difficulty of discrimination, and given our desire to test the discrimination of a difficult contrast, we used the more difficult order of presentation for all infants in all conditions, rather than counterbalancing the order within each condition.

presented with a posttest trial on which they heard a trisyllabic nonsense word ('bupoki', produced by a synthetic female voice), which was presented repeatedly until the participant looked away from the screen for more than 2 seconds. This trial was included to ensure that any failure to dishabituate to the change stimuli was not due to general unresponsiveness. Infants whose looking times showed no numerical increase either between habituation and change trials or between change trials and the posttest trial were excluded from analysis (this was true for two infants in the Bimodal condition and one infant each in the Unimodal and Control conditions).

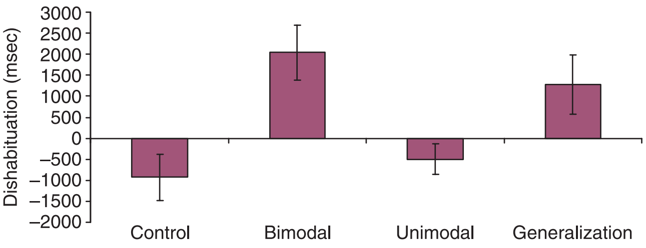

Figure 3 *Mean dishabituation (average looking time for two change trials minus average for final two habituation trials, msec) for infants in the Control, Bimodal, Unimodal (Experiment 1), and Generalization (Experiment 2) conditions, collapsed across velar and dental places of articulation. Error bars indicate 1 standard error.*

Results

To assess habituation, we compared the mean for each infant's first three habituation trials to the mean of their last three habituation trials in a 3 (condition) × 2 (first 3 vs. last 3) ANOVA. As expected, there was a main effect of habituation (F[1, 94] = 194.14, $p < .001$), with looking time decreasing significantly between the beginning and end of the habituation phase. There was a significant main effect of familiarization condition (F[2, 94] = 3.653, $p < .05$), with longer looking times for infants in the Control condition than infants in either the Bimodal or Unimodal conditions, both for the first three habituation trials (Control: 19.9 s; Bimodal: 12.9 s; Unimodal: 16.1 s) and the last three habituation trials (Control: 7.8 s; Bimodal: 5.6 s; Unimodal: 5.8 s). The interaction between familiarization condition and the degree of habituation was significant (F[2, 94] = 4.71, $p < .05$), reflecting the fact that looking times for infants in the Control condition showed a greater decrease across the habituation phase. However, the three familiarization conditions did not differ with respect to each infant's total exposure to the speech stimuli across the entire habituation phase ($F < 1$, *ns*; Control: 117.5 s; Bimodal: 110.3 s; Unimodal: 107.1 s).

To assess discrimination, we analyzed mean looking times for the last two habituation trials and for the two change trials. To eliminate outliers, infants whose dishabituation differed by more than 2 standard deviations from the mean of their familiarization condition were excluded from the analysis. This was true for four infants in the Bimodal condition and two infants each in the Unimodal and Control conditions. Average looking times for each condition are shown in Table 1. For each infant, we calculated individual difference scores by subtracting their mean looking time on change trials from their mean on the final two habituation trials. Average difference scores for each condition are shown in Figure 3. We then conducted a 3 (familiarization condition: Bimodal, Unimodal, Control) × 2 (place of articulation: velar vs. dental) ANOVA on these difference scores. This analysis revealed a main effect of familiarization condition (F[2, 91] = 8.599, $p < .001$). There was not a significant effect of place of articulation (F[1, 91] = 1.403, $p > .23$) and no interaction between place of articulation and familiarization condition ($F < 1$, *ns*), indicating that the VOT contrast was discriminated equally well at both dental and velar places of articulation. In subsequent analyses data were pooled across the two places of articulation within each familiarization condition.

To isolate the differences underlying the main effect of familiarization condition we conducted paired samples t-tests. Infants in the Bimodal condition dishabituated to the change stimulus (t[31] = 3.088, $p < .005$), while looking times for infants in the Unimodal and Control conditions continued to decrease on change trials (mean differences: Unimodal = −501 ms; Control = −925 ms), suggesting that only infants in the Bimodal condition discriminated the contrast.[7] In addition, independent

Table 1 *Average looking time (msec) for each condition (Experiments 1 and 2) on the final two trials of habituation and the two change trials. Standard errors are given in parentheses*

	Final habituation trials	Change trials
Bimodal	4807 (362)	6844 (628)
Unimodal	5362 (420)	4861 (360)
Control	6466 (672)	5540 (478)
Generalization	5421 (453)	6697 (740)

---

[7] One difference between the Control group and infants in the other familiarization conditions is that the Control group was familiarized to (potentially less interesting) nonspeech stimuli at a more rapid rate of presentation (three tones per second, compared to one syllable per 1.338 seconds for the other conditions). Thus a concern is that infants in the Control group may have failed to dishabituate due to greater overall inattentiveness. However, a comparison of looking times during familiarization indicates that infants in the control condition were at least as attentive as infants in the speech conditions, with a mean looking time of 13.5 sec per familiarization trial (compared to 11.2 sec and 10.4 sec for infants in the Unimodal and Bimodal conditions, respectively). Moreover, all conditions were run to the same criterion of habituation prior to presentation of the two change trials (looking time decreased to at least 50% of initial trials, measured over a moving window of three trials). Finally, there was no difference among the familiarization conditions in the likelihood of recovering to the posttest stimulus that assessed generalized fatigue with the testing situation (two infants in the Bimodal condition and one infant each from the Unimodal and Control conditions failed to dishabituate to the posttest).

sample *t*-tests found that the change in looking time for infants in the Bimodal condition significantly differed from that of infants in both the Control (*t*[61] = 3.428, *p* = .001) and Unimodal conditions (*t*[64] = 3.436, *p* = .001). There was no significant difference between Unimodal and Control groups (*t* < 1, *ns*).

Discussion

The results from Experiment 1 support our hypothesis that exposure to a bimodal distribution results in enhanced discrimination of a difficult contrast. The Control group's failure to dishabituate confirms that this contrast is difficult for infants to discriminate in the absence of prior familiarization, while the Bimodal group's dishabituation shows that discrimination of this contrast is enhanced following exposure to a bimodal distribution. Moreover, the fact that infants in the Unimodal condition did not discriminate the contrast rules out the possibility that facilitation in the Bimodal condition was due to simple exposure to the speech stimuli. Rather, facilitation was due to the shape of the distribution to which infants were familiarized. Familiarization to a phonetic continuum whose frequency distribution is indicative of a phonetic contrast (i.e. a bimodal distribution) results in enhanced discrimination of stimuli lying within different modes along the distribution; familiarization to the same continuum but a frequency distribution indicative of the *absence* of a phonetic contrast (i.e. a unimodal distribution) does not.

Experiment 2

The results of Experiment 1, combined with the findings of Maye *et al.* (2002), support the notion that the phonetic distribution exhibited in the speech that infants hear affects their patterns of speech perception, resulting in both the loss of sensitivity to non-native contrasts as well as the facilitation of difficult contrasts. In Experiment 2 we investigate the level at which infants encode these contrasts during the course of distributional learning. In particular, we ask whether familiarization at one place of articulation affects discrimination of the analogous contrast at another place of articulation. In particular, does familiarization to a bimodal distribution of sounds at one place of articulation (e.g. velar) also enhance discrimination of a similar contrast at a different place of articulation (e.g. dental)?

During the course of normal language acquisition, infants presumably have ample exposure to all phonetic categories of the language. Thus, it would never be the case that an infant would need to generalize knowledge about one phonetic contrast to posit the possible existence of an analogous but unheard contrast. However, the use of an artificial language paradigm enables us to ask the question whether such generalization would occur. If infants do generalize phonetic learning to analogous contrasts, it would indicate that the manner in which infants encode phonetic information during acquisition involves some abstraction from the raw acoustic input. Specifically, it would suggest that infants' sensitivity to a particular acoustic dimension (in this case, VOT) has been altered. In other words, the infants have begun to acquire something like a (native language-specific) phonetic or acoustic feature.

To test the hypothesis that infants learn phonetic contrasts at the featural level, we familiarized a new group of infants to a bimodal distribution at one place of articulation and then tested their discrimination at the other place of articulation. If learning occurs at the level of the segment, these infants should not discriminate the untrained contrast, because (like the Control group in Experiment 1) they have had no prior exposure to the tested contrast. However, if learning occurs at the level of the acoustic or phonetic feature, these infants should show facilitated discrimination of the untrained contrast, due to having been familiarized to a bimodal distribution of the same featural contrast.

Methods

In this experiment, 44 8-month-olds were familiarized in a manner identical to the Bimodal condition in Experiment 1. That is, during familiarization they heard the prevoiced/short-lag contrast at one place of articulation (17 to the velar stimuli, 18 to the dental stimuli), presented according to the bimodal frequency distribution shown in Figure 2. They were then tested on their discrimination of the prevoiced/short-lag contrast at the other place of articulation. For example, half of the infants were familiarized to a bimodal distribution of the dental stimuli, and then tested on their discrimination of the velar contrast.

The participants were 23 males and 21 females whose average age was 8 months and 8 days (ranging from 7 months, 22 days to 9 months, 2 days). Nine of these infants were tested but excluded from analysis due to crying or fussing (3), technical problems (1), exposure to a native language other than English (3), ear infection at the time of testing (1), or dishabituation that differed by more than 2 standard deviations from the mean (1). Infants were from English-speaking homes, and were recruited based on parental interest in research participation. In all other respects, the methods were identical to Experiment 1.

Results

As in Experiment 1, we analyzed mean looking time for the last two habituation trials and for the two change trials. To eliminate outliers, one infant whose dishabituation differed by more than 2 standard deviations from the mean was excluded from analysis. Average looking time data are shown in Table 1, and average difference score is shown in Figure 3.

Infants in Experiment 2 (the Generalization group) showed dishabituation to the change stimuli that was significant in a one-tailed though not a two-tailed *t*-test ($t[34] = 1.798$, two-tailed $p = .081$, one-tailed $p < .05$). Although these infants' recovery of looking time was more modest than for infants in the Bimodal condition of Experiment 1, the fact that looking time changed in the predicted direction suggests that they discriminated the voicing contrast at the untrained place of articulation. An ANOVA assessing dishabituation as a function of place of articulation found no effect of place of articulation ($F[1, 34] < 1$, *ns*), indicating that discrimination was equally robust at both dental and velar places of articulation.

We also compared these infants with the infants from Experiment 1 in a 4 × 2 ANOVA (4 familiarization conditions × 2 places of articulation) conducted over individual difference scores (average change trial duration – average of final 2 habituation trials). This analysis found a significant effect of familiarization condition, ($F[3, 124] = 5.445$, $p < .005$), but no effect of place of articulation ($F[1, 124] = 1.29$, $p > .25$) or interaction ($F < 1$, *ns*). To determine which of the three Experiment 1 conditions that infants in the Generalization condition were most similar to, we compared the Generalization group with each of the other familiarization conditions in independent samples *t*-tests.

We first compared the Generalization and the Control conditions – two conditions in which infants were tested on discrimination of speech sounds they had not been familiarized to. Infants in the Generalization condition showed a significantly greater increase in looking time on change trials than infants in the Control condition ($t[64] = 2.401$, $p < .05$). This difference indicates that familiarization to a bimodal distribution of speech sounds at a different place of articulation facilitated infants' discrimination of a featurally analogous contrast, above the baseline discrimination of unfamiliarized infants.

We then compared the Generalization and Bimodal conditions – two conditions in which infants were familiarized to a bimodal distribution of the VOT continuum. There was no significant difference in dishabituation between these two conditions ($t < 1$, *ns*), providing further confirmation that bimodal familiarization increased discrimination of the voicing contrast regardless of whether or not the familiarization stimuli had the same place of articulation as the test stimuli.

Finally, we compared the Generalization and Unimodal conditions. Although these two conditions differed in both place of articulation and training distribution, the Unimodal condition provides something akin to a baseline measure for infants who have been familiarized to speech stimuli. That is, infants in the Unimodal condition received familiarization to speech stimuli but in a distribution that was not predicted to facilitate discrimination. Thus, as in the Bimodal vs. Unimodal comparison for Experiment 1, a comparison of the Generalization and Unimodal conditions provides a measure of how much discrimination benefit is provided by familiarization to a bimodal distribution, above and beyond simple exposure to the speech stimuli. The difference between these two conditions was significant ($t[50]^8 = 2.235$, $p < .05$), providing further evidence that familiarization to a bimodal distribution results in improved discrimination of a featurally analogous contrast.

Discussion

The results from Experiment 2 indicate that exposure to a bimodal distribution at one place of articulation facilitates discrimination at a second, untrained place of articulation. In other words, the infants in Experiment 2 appear to have extracted the featural properties of the input speech after less than 3 minutes of exposure, by attending to the particular acoustic dimension that is relevant for discriminating these contrasts. It is important to note that this generalization of voicing across place of articulation is not the result of infants failing to discriminate the place of articulation differences used in this experiment. There is ample evidence that infants as young as 2–3 months of age can discriminate stop consonants varying in place of articulation (Eimas, 1974; Walley, Pisoni & Aslin, 1984). Our finding of feature generalization is in line with research showing that 6- and 9-month-old infants treat sets of sounds forming a natural class as more similar than sets with no unifying phonetic features (Hillenbrand, 1983, 1985; Jusczyk, Goodman & Baumann, 1999). At least by the age of 6 months, infants appear to encode speech sounds at an abstract level, on the basis of the featural relationships between the sound categories of a language. This featural information also appears to aid 9-month-old infants in the acquisition of phonotactic patterns (Saffran & Thiessen, 2003).

The results from Experiment 2 are particularly interesting in light of a previous finding that adult participants who learn to discriminate a contrast via exposure to a bimodal distribution do *not* generalize to an untrained place of articulation (Maye & Gerken, 2001). However, the methods used in the adult study were not completely analogous to those used in this experiment; in the adult task, participants were asked to make metalinguistic judgments about the stimuli in a minimal pair task, whereas with the infants we simply assessed discrimination. It is possible that generalization occurs at the level of discrimination but is not robust enough to influence metalinguistic judgments. Thus, one future direction for this research will entail a replication of the adult study, employing a methodology more closely resembling that of the current study, in order to

---

[8] Levene's test for equality of variance indicated unequal variance in the Unimodal vs. Generalization conditions, thus the degrees of freedom were adjusted from 67 to 50. This adjustment did not affect the significance level obtained.

determine whether the discrepant findings are due to methodological differences. A second possibility is that infants, who are in the process of acquiring a first language, encode speech sounds at a different level of abstraction than adults. Infants appear to extract the featural properties of the input speech, while adult learning may be restricted to the segmental level.

General discussion

The overall results from these experiments have shown that exposure to a bimodal distribution of speech sounds results in enhanced discrimination of a difficult speech contrast. We have shown that, for a phonetic contrast that is not discriminated by infants in the absence of familiarization, infants familiarized to a unimodal distribution remain unable to discriminate the contrast, whereas familiarization to a bimodal distribution results in enhanced discrimination. Combined with the results from the Maye *et al.* (2002) study, these data suggest that infants' sensitivity to the distributional properties of speech may account for both the loss and enhancement patterns seen in infants' development of native language speech perception. These data also parallel those obtained in the domain of face perception, where the typical specialization for species-specific faces can be affected by altering the nature of the faces to which infants are exposed during the relevant developmental time-window (Pascalis *et al.*, 2002, 2005).

These results highlight infants' sensitivity to fine-grained phonetic variation, since in order to track a phonetic distribution an infant must be able to encode differences between phonetically similar tokens (i.e. tokens that might fall within a single adult phonetic category) in order to note their relative frequency of occurrence. The same sensitivity has been found in distributional phonetic learning experiments with adult subjects (Maye & Gerken, 2000, 2001; Hayes, 2003), and is evident in word recognition in both adults (Andruski, Blumstein & Burton, 1994; McMurray, Tanenhaus & Aslin, 2002) and infants (McMurray & Aslin, 2005). These findings may seem to fly in the face of received wisdom regarding the well-studied phenomenon of categorical perception, in which both adult and infant listeners are found to be more sensitive to some regions of phonetic space (that line up with category boundaries, for adults) than to other regions (i.e. within-category variation; Liberman, Harris, Hoffman & Griffith, 1957; see Repp, 1984, for review). However, in the categorical perception literature subjects' discrimination of within-category variation is typically found to be above chance, and higher than would be predicted by their categorization responses (e.g. Liberman *et al.*, 1957; Miyawaki, Strange, Verbrugge, Liberman, Jenkins & Fujimura, 1975; Best, Morrongiello & Robson, 1981; Repp, 1984), indicating that listeners do retain some sensitivity to within-category differences. Furthermore, studies in which subjects are asked to rate the 'goodness' of a stimulus as a token of some category (e.g. 'how good of a "t" is it?') consistently report gradient responses, indicating that not all members of a category are perceived as equally good exemplars (Massaro & Cohen, 1983; Miller & Volaitis, 1989; Kuhl, 1991), again highlighting the fact that listeners do indeed perceive fine within-category phonetic detail, to a degree that may be underestimated by their performance on discrimination tasks alone. The human perceptual system is clearly more sensitive to some regions of acoustic-phonetic space (that tend to line up with phonetic category boundaries) than other regions (that tend to fall within phonetic categories), some of which are innate (as evidenced by the findings of categorical perception in young infants; e.g. Eimas *et al.*, 1971), and others that are induced by native language phonetic categories (e.g. Abramson & Lisker, 1970). However, the present study adds to a growing body of research demonstrating that despite the categorical nature of overt discrimination abilities,[9] both infants and adults continue to encode fine-grained, within-category phonetic detail.

Despite the fact that listeners have been found to encode greater perceptual detail for speech sounds than is evident in overt discrimination tasks, we cannot rule out the possibility that the infants in this study did not resolve the phonetic continua tested here into the full eight continuum steps. It is thus important to consider how this would have affected infants' perception of the familiarization distributions in the present experiments. Consider the most extreme case in which infants perceived these phonetic continua as only two categories with no within-category variation. If the category boundary divided these continua at their midpoints, then because both the unimodal and bimodal distributions were symmetrical, both groups would be exposed to an equal number of exemplars in the two categories and no difference in performance would be expected. This was clearly not the case as we did find differences in performance between the Unimodal and Bimodal groups. However, it is possible that the category boundary for infants is offset from the middle of the continua. In this case, infants in the Unimodal and Bimodal conditions would be exposed to a different number of exemplars from the two categories. For example, if the category boundary was at step 6, Bimodal infants would hear 76 tokens to the left of this boundary and 44 tokens to the right, whereas Unimodal infants would hear 104 tokens to the left and 16 to the right. During test, infants were habituated to stimulus 6 and recovery scores to stimulus 3 were assessed. Because prior to test infants in the Unimodal group had heard exemplars of the category corresponding to stimulus 3 more than six times as often

---

[9] McMurray and colleagues (2002) found that the extent to which subjects evidence categorical perception is influenced by task-related factors. When presented with a more naturalistic task than is typical in categorical perception studies, responses are less categorical.

as exemplars from the other category, while infants in the Bimodal group heard only twice as many such exemplars of the stimulus 3 category, stimulus 3 would be less novel to infants in the Unimodal group, and recovery after habituation to stimulus 6 might be less. Unfortunately, there are no data available to resolve this issue of where the category boundary lies for infants. However, given the difficulty that infants have with discriminating the phonetic contrast in the Control (tone) condition, it seems unlikely that the boundary is clearly specified, and in fact there may be little or no ability to make the relevant phonetic discrimination prior to exposure to the bimodal distribution. But even if infants are not sensitive to the full eight-step continuum and perceive it as an asymmetrical distribution, we have demonstrated that a simple distributional cue (frequency of occurrence of category X vs. category Y) biases the subsequent discriminability of a phonetic contrast that was difficult to detect prior to exposure. It would be interesting to explore this issue in future research by shifting the distributions in one direction or the other on the continuum in a manner that changes the relative asymmetry between the Unimodal and Bimodal conditions (e.g. making them equally asymmetrical) while retaining the basic unimodal–bimodal distinction.

Although we have discussed the development of infants' speech perception as a process of maintenance, enhancement, or loss, it is unlikely that perceptual development is a simplistic, binary parameter-setting mechanism in which infants are born with some discrete set of potential phonetic contrasts from which they must simply determine the subset of contrasts that are active in the language being learned. Rather, perceptual development is probably best characterized as a process of shaping an infant's discrimination profile such that it matches the phonetic categories of the native language according to any number of continuous acoustic dimensions. We believe this to be the case because language learners ultimately encode the unique phonetic characteristics of their language and the particular acoustic cues signaling each contrast. For example, while both English and Spanish distinguish voiced vs. voiceless consonants, in English VOT is a primary cue to this contrast, while in Spanish listeners place more weight on the duration of the closure interval before the stop is released (Zampini & Green, 2001; Martinez Celdrán, 1993). This difference in cue weighting presumably reflects the fact that English speakers do not produce a reliable difference in closure duration between voiced and voiceless stops (Crystal & House, 1988), while Spanish speakers do (Green, Zampini & Magloire, 1997). Recent research has demonstrated in the laboratory that adult subjects can appropriately adjust their cue weighting in response to whether or not particular cues reliably signal a contrast (Goudbeek, Smits, Swingley & Cutler, 2005; Holt & Lotto, 2006).

Furthermore, for any two languages or dialects in which a given acoustic cue is relevant (e.g. languages that use VOT to distinguish voiced and voiceless stops such as /b/ vs. /p/), the actual acoustic realization of this cue may differ between the two linguistic systems (Cho & Ladefoged, 1999). For example, the VOT boundary between /b/ and /p/ is around 0 msec lag for European French, 7 msec lag for Canadian French, 25 msec lag for Canadian English (Caramazza & Yeni-Komshian, 1974), and up to 40 msec lag for Danish (depending on vowel context; Christensen, 1984). These details are not lost on the speakers of a language; they are part of what enables listeners to identify a speaker's foreign or dialectal accent (Evans & Iverson, 2004; Clopper & Pisoni, 2004), or an individual's personal ideolect or pronunciation habits (Sancier & Fowler, 1997; Allen & Miller, 2004). Thus, for infants learning a language, it is not enough to simply determine whether their language has a voicing contrast, for example, or whether VOT is a relevant acoustic parameter; they must become sensitive to the particular region of VOT that differentiates voiced from voiceless stops in the language or dialect they are learning.

Previous researchers have proposed models in which the process of phonetic category acquisition is one of attunement (e.g. Kuhl, 1993; Lacerda, 1995), rather than the simple addition or subtraction of fixed category boundaries. In particular, attunement is predicted by the notion of phonetic prototypes, wherein phonetic categories are warpings of perceptual space on the basis of the distribution of the specific phonetic properties of the linguistic input (Repp, 1977; Samuel, 1977, 1982; Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992; Miller, 1994; Miller & Eimas, 1996). Kuhl and colleagues (1992) found that at 6 months American and Swedish infants were already sensitive to the phonetic particularities of their native language, as the two groups of infants responded differently to a vowel token that is prototypical for American English [i], but is an atypical token of Swedish [i]. In future research this process of attunement can be explored by presenting infants with phonetic distributions in which the modes are positioned at slightly different locations, which should result in somewhat different patterns of sensitivity.

The results of Experiment 2 showed that infants generalized from a newly learned contrast to an analogous contrast exhibiting the same featural distinction. This finding may have implications regarding the types of representations formed by infants as they listen to speech sounds. If infants compute phonetic distributions at the level of the phonetic feature, it would suggest that they do not encode the sounds as holistic exemplars or bundles of features, but rather in terms of the individual features that signal contrasts between sets of sounds. That is, rather than learning that [b] and [p] are contrastive, for example, they learn to attend to a particular region of the VOT continuum. This process may also account for Kikuyu infants' ability to discriminate the voiced–voiceless contrast that is attested to in their Bantu dialect at the alveolar but not bilabial place of articulation (Streeter, 1976).[10] As a result, the infant's discrimination profile grows to match the appropriate pattern of the native language, which sets the stage for subsequent developments in language acquisition. These results from infants contrast with previous studies of adults which suggest that adults do not learn phonetic categories at the level of the feature; for adults, familiarization to a phonetic distribution affects discrimination of the familiar contrast, but not a featurally analogous contrast (Maye, 2000; Maye & Gerken, 2000). Future research is needed to determine whether this difference is indeed indicative that phonetic learning occurs in a fundamentally different way at different ages.

In summary, the present experiments have demonstrated that exposure to a phonetic distribution indicative of a contrast (i.e. a bimodal distribution) facilitates the discrimination of a difficult phonetic contrast. This process may account for the improved discrimination of difficult contrasts (e.g. improved discrimination of [d]~[ð] by English speakers) between infancy and adulthood for those contrasts that are attested to in the native language input. In conjunction with previous research (Maye *et al.*, 2002), these results suggest that sensitivity to the distributional properties of speech sounds plays an important role in shaping the infant's discrimination profile to match the native language categories. In addition, the present results suggest that infants encode this phonetic information at a relatively abstract level, in terms of featural dimensions of contrast. These findings have broad implications for our understanding of phonetic development during the first year of life.

References

Abramson, A.S., & Lisker, L. (1970). Discriminability along the voicing continuum: cross-language tests. In *Proceedings of the Sixth International Congress of Phonetic Sciences*. Prague: Academia.

Allen, J.S., & Miller, J.L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, **115**, 3171–3183.

Andruski, J.E., Blumstein, S.E., & Burton, M.W. (1994). The effect of subphonetic differences on lexical access. *Cognition*, **52** (1), 63–187.

Aslin, R.N., & Pisoni, D.B. (1980). Some developmental processes in speech perception. In G.H. Yeni-Komshian, J.H. Kavanagh, & C.A. Ferguson (Eds.), *Child phonology, 2: Perception* (pp. 67–96). New York: Academic Press.

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

Best, C.T., McRoberts, G.W., & Goodell, E. (1990). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, **109** (2), 775–794.

Best, C.T., McRoberts, G.W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative speech contrasts. *Infant Behavior and Development*, **18**, 339–350.

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: Human Perception and Performance*, **14** (3), 345–360.

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

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, **46**, 217–243.

Burnham, D.K. (1986). Developmental loss of speech perception: exposure to and experience with a first language. *Applied Psycholinguistics*, **7**, 207–239.

Burnham, D.K., Earnshaw, L.J., & Quinn, M.C. (1987). The development of categorical identification of speech. In B.E. McKenzie & R.H. Day (Eds.), *Perception in infancy: Problems and issues* (pp. 237–275). Hillsdale, NJ: Lawrence Erlbaum.

Caramazza, A., & Yeni-Komshian, G.H. (1974). Voice onset time in two French dialects. *Journal of Phonetics*, **2**, 239–245.

Cho, T., & Ladefoged, P. (1999). Variations and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, **27**, 207–229.

Christensen, J.B. (1984). The perception of voice-onset-time: a cross-language study of American English and Danish. *Annual Report of the Institute of Phonetics University of Copenhagen*, **18**, 163–184.

Clopper, C.G., & Pisoni, D.B. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, **32**, 111–140.

Crystal, T.H., & House, A.S. (1988). The duration of American English stop consonants: an overview. *Journal of Phonetics*, **16**, 285–294.

Cutting, J.E., & Rosner, B.S. (1974). Categories and boundaries in speech and music. *Perception and Psychophysics*, **16**, 564–570.

Dooling, R.J., Best C.T., & Brown, S.D. (1995). Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). *Journal of the Acoustical Society of America*, **97**, 1839–1846.

Eilers, R.E. (1977). Context-sensitive perception of naturally produced stop and fricative consonants by infants. *Journal of the Acoustical Society of America*, **61**, 1321–1336.

Eilers, R.E., & Minifie, F.D. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research*, **18**, 158–167.

Eilers, R.E., Wilson, W.R., & Moore, J.M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research*, **20**, 766–780.

Eimas, P.D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception and Psychophysics*, **16**, 513–521.

---

[10] The process of learning to attend to the appropriate dimensions of contrast in a language could be accomplished by either creating/moving perceptual boundaries, or by warping the perceptual space such that particular regions of some acoustic cue (e.g. VOT) become more or less salient.

Eimas, P.D. (1975). Speech perception in early infancy. In L.B. Cohen & P. Salapatek (Eds.), *Infant perception, 2: From sensation to cognition* (pp. 193–231). New York: Academic Press.

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

Evans, B.B., & Iverson, P. (2004). Vowel normalization for accent: an investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America*, **115**, 352–361.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds 'l' and 'r'. *Neuropsychologia*, **9**, 317–323.

Goudbeek, M., Smits, R., Swingley, D., & Cutler, A. (2005). Acquiring auditory and phonetic categories. In H. Cohen & C. Lefebvre (Eds.), *Categorization in cognitive science* (pp. 497–513). Amsterdam: Elsevier.

Green, K.P., Zampini, M.L., & Magloire, J. (1997). An examination of word-initial stop closure interval in English, Spanish, and Spanish-English bilinguals. *Journal of the Acoustical Society of America*, **102**, 3136.

Hayes, R. (2003). How are second language phoneme contrasts learned? Doctoral Dissertation, University of Arizona.

Hillenbrand, J. (1983). Perception organization of speech sounds by infants. *Journal of Speech and Hearing Research*, **26** (2), 268–282.

Hillenbrand, J. (1985). Perception of feature similarities by infants. *Journal of Speech and Hearing Research*, **28** (2), 317–318.

Holmberg, T.L., Morgan, K.A., & Kuhl, P.K. (1977). Speech perception in early infancy: discrimination of fricative consonants. *Journal of the Acoustical Society of America*, **62**, Supplement 1, S99(A).

Holt, L.L., & Lotto, A.J. (2006). Cue weighting in auditory categorization: implications for first and second language acquisition. *Journal of the Acoustical Society of America*, **119**, 3059–3071.

Jusczyk, P.W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P.W., Goodman, M.B., & Baumann, A. (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language*, **40**, 62–82.

Jusczyk, P., Rosner, B., Cutting, J., Foard, C., & Smith, L. (1977). Categorical perception of nonspeech sounds by two-month-old infants. *Perception and Psychophysics*, **21**, 50–54.

Kuhl, P.K. (1978). Predispositions for the perception of speech-sound categories: a species-specific phenomenon? In F.D. Minifie & L.L. Lloyd (Eds.), *Communicative and cognitive abilities: Early behavioral assessment* (pp. 229–256). Baltimore, MD: University Park Press.

Kuhl, P.K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, **50** (2), 93–107.

Kuhl, P.K. (1993). Early linguistic experience and phonetic perception: implications for theories of developmental speech perception. *Journal of Phonetics*, **21**, 125–139.

Kuhl, P.K., & Miller, J.D. (1975). Speech perception by the chinchilla: voiced–voiceless distinction in alveolar plosive consonants. *Science*, **190**, 69–72.

Kuhl, P.K., & Padden, D.M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America*, **73**, 1003–1010.

Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, **9** (2), F13–F21.

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

Lacerda, F. (1995). The perceptual magnet effect: an emergent consequence of exemplar-based phonetic memory. In K. Elenius & P. Branderyd (Eds.), *Proceedings of the XIIIth International Congress of the Phonetic Sciences, vol. 2* (pp. 140–147). Stockholm: Royal Institute of Technology.

Lasky, R.E., Syrdal-Lasky, A.K., & Klein, R.E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, **20**, 215–225.

Liberman, A.M., Harris, K.S., Hoffman, H.S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, **54**, 358–368.

Lisker, L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, **20**, 384–482.

Lotto, A.J., Sato, M., & Diehl, R.L. (2004). Mapping the task for the second language learner: the case of Japanese acquisition of /r/ and /l/. In J. Slifka, S. Manuel, & M. Matthies (Eds.), *From sound to sense: 50+ years of discoveries in speech communication*, Research Laboratory of Electronics at MIT, Cambridge, MA. Electronic conference proceedings.

McMurray, B., & Aslin, R.N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, **95** (2), B15–B26.

McMurray, B., Tanenhaus, M.K., & Aslin, R.N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, **86**, B33–B42.

Magloire, J., & Green, K.P. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica*, **56**, 158–185.

Martinez Celdrán, E. (1993). La percepción categorial de /b-p/ en espanol basada en las diferencias de duración. *Estudios de Fonética Experimental*, **5**, 224–239.

Massaro, D.W., & Cohen, M.M. (1983). Phonological context in speech perception. *Perception and Psychophysics*, **34** (4), 338–348.

Maye, J. (2000). The acquisition of speech sound categories on the basis of distributional information. Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In S.C. Howell, S.A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Boston University Conference on Language Development* (pp. 522–533). Somerville, MA: Cascadilla Press.

Maye, J., & Gerken, L. (2001). Learning phonemes: how far can the input take us? In A.H-J. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 480–490). Somerville, MA: Cascadilla Press.

Maye, J., Werker, J.F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, **82** (3), B101–B111.

Miller, J.L. (1994). On the internal structure of phonetic categories: a progress report. *Cognition*, **50**, 271–285.

Miller, J.L., & Eimas, P.D. (1996). Internal structure of voicing categories in early infancy. *Perception and Psychophysics*, **58**, 1157–1167.

Miller, J.L., & Volaitis, L.E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception and Psychophysics*, **46** (4), 505–512.

Miller, J.L., Weir, C.C., Pastore, R., Kelly, W.J., & Dooling, R.J. (1976). Discrimination and labeling of noise-buzz sequences with varying lead times: an example of categorical perception. *Journal of the Acoustical Society of America*, **60**, 410–417.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J., & Fujimura, A. (1975). An effect of linguistic experience: the discrimnination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, **18**, 331–340.

Morse, P.A., & Snowdon, C.T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception and Psychophysics*, **17**, 9–16.

Narayan, C. (2006). Follow your nose: non-native nasal consonant discrimination in infancy. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 411–422). Cambridge, MA: Cascadilla Press.

Newman, R.S., Clouse, S.A., & Burnham, J. (2001). The perceptual consequences of acoustic variability in fricative production within and across talkers. *Journal of the Acoustical Society of America*, **109** (3), 1181–1196.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, **64**, B9–B17.

Pascalis, O., de Haan, M., & Nelson, C.A. (2002). Is face processing species-specific during the first year of life? *Science*, **296**, 1321–1323.

Pascalis, O., Scott, L.S., Kelly, D.J., Shannon, R.W., Nicholson, E., Coleman, M., & Nelson, C.A. (2005). *Proceedings of the National Academy of Sciences*, **102**, 5297–5300.

Pisoni, D.B. (1977). Identification and discrimination of the relative onset of two component tones: implications for voicing perception in stop consonants. *Journal of the Acoustical Society of America*, **61**, 1352–1361.

Pisoni, D.B., Aslin, R.N., Perey, A.J., & Hennessy, B.L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, **8** (2), 297–314.

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/~/D/ discrimination: evidence for a new developmental pattern. *Journal of the Acoustical Society of America*, **109**, 2190–2201.

Polka, L., & Werker, J.F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 421–435.

Repp, B.H. (1977). Dichotic competition of speech sounds: the role of acoustic stimulus structure. *Journal of the Acoustical Society of America*, **3**, 37–50.

Repp, B.H. (1984). Categorical perception: issues, methods, findings. In N.J. Lass (Ed.), *Speech and language: Advances in basic research and practice, 10* (pp. 1249–1257). New York: Academic Press.

Saffran, J.R., & Thiessen, E.D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, **39**, 484–494.

Samuel, A.G. (1977). The effect of discrimination training on speech perception: noncategorical perception. *Perception and Psychophysics*, **22**, 321–330.

Samuel, A.G. (1982). Phonetic prototypes. *Perception and Psychophysics*, **31**, 307–314.

Sancier, M.L., & Fowler, C.A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, **25**, 421–436.

Sinex, D.G., & McDonald, L.P. (1989). Synchronized discharge rate representation of voice onset time in the chinchilla auditory nerve. *Journal of the Acoustical Society of America*, **85**, 1995–2004.

Steinschneider, M., Schroeder, C., Arezzo, J., & Vaughan, H. (1996). Physiological correlates of the voice onset time boundary in primary auditory cortex of the awake monkey – temporal response patterns. *Brain and Language*, **48**, 326–340.

Steinschneider, M., Volkov, I.O., Noh, M.D., Garell, P.C., & Howard, M.A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of Neurophysiology*, **82**, 2346–2357.

Stevens, K.N. (1972). The Quantal nature of speech: evidence from articulatory-acoustic data. In E.E. David & P.B. Denes (Eds.), *Human communication: A unified view* (pp. 51–66). New York: McGraw-Hill.

Stevens, S.S., Volkman, J., & Newman, E.B. (1937). A scale for the measurement of the psychological magnitude of pitch. *Journal of the Acoustical Society of America*, **8**, 185–190.

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

Sundara, M., Polka, L., & Genesee, F. (2006). Language experience facilitates discrimination of /d-∂/ in monolingual and bilingual acquisition of English. *Cognition*, **100** (2), 186–199.

Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica*, **56**, 186–199.

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

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. (1996). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6–12 months: language-specific developmental changes in speech perception abilities. *The Emergence of Human Cognition and Language*, **3**, 57–61.

Walley, A.C., Pisoni, D.B., & Aslin, R.N. (1984). Infant discrimination of two- and five-formant voiced stop consonants differing in place of articulation. *Journal of the Acoustical Society of America*, **75**, 581–589.

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

Werker, J.F., & Tees, R.C. (1984). Developmental changes across childhood in the perception of nonnative speech sounds. *Canadian Journal of Psychology*, **37**, 278–286.

Zampini, M.L., & Green, K.P. (2001). The voicing contrast in English and Spanish: the relationship between perception and production. In J.L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 23–48). Malden, MA: Blackwell.

Received: 29 November 2005
Accepted: 16 January 2007