A Statistical Basis for Speech Sound Discrimination*

Jennifer L. Anderson, James L. Morgan, and Katherine S. White

Brown University

Key words

- infant speech perception
- non-native contrasts
- phoneme discrimination
- phonological categories
- statistical learning

Abstract

Infants under six months are able to discriminate native and non-native consonant contrasts equally well, but as they learn the phonological systems of their native language, this ability declines. Current explanations of this phenomenon agree that the decline in discrimination ability is linked to the formation of native-language phonemic categories. The goal of this study was to evaluate the role of input statistics in learning these categories: our hypothesis was that relative frequency is a determinant of the relative order in which categories are acquired. English-learning infants of two age groups (6.5 months and 8.5 months) were tested on their ability to discriminate non-native consonant contrasts using the Conditioned Head Turn Procedure. As predicted, older infants were worse in their performance on the more frequent coronal stop contrast than on the less frequent dorsal stop contrast. In contrast, 6.5-month-olds discriminated both contrasts equally well. An adult control group tested with an AX task also discriminated both contrasts equally. These results provide preliminary confirmation of the hypothesis that frequency plays an important role in tuning of phonological systems to properties of the native language. A simple attractor model suffices to account for these and previous results on loss of discrimination of non-native-language contrasts and suggests that the technique of measuring graded loss of multiple contrasts, in combination with observation of input frequencies, can offer a powerful method of assessing infants' phonological representations.

* Acknowledgements: This research was supported by a grant from the National Institute of Health (5 R01 HD32005) to JLM. JLA was supported by the National Science Foundation IGERT Program (9870676). We thank Janet Werker and Catherine Best for contributing their stimuli for use in the studies, Karen Rathbun for her assistance in data collection, and Catherine Best, Heather Bortfeld, Katherine Demuth, Leher Singh, Melanie Soderstrom, and Janet Werker for comments on drafts and development of studies. Requests for reprints should be directed to James Morgan, Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912.

Address for correspondence: Jennifer L. Anderson: Department of Cognitive & Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912; phone: 401-863-1160; fax: 401-863-2255; e-mail: <janderson@microsoft.com>; James Morgan: Department of Cognitive & Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912; phone: 401-863-2462; fax: 401-863-2255; e-mail: <james_morgan@brown.edu>; Katherine S. White: Department of Cognitive & Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912; phone: 401-863-1167; fax: 401-863-2255; e-mail: <katherine_s_white@brown.edu>.
1 Introduction

Infants begin life with the ability to learn any of the world's languages. While adults are often hampered in their ability to identify or discriminate phones that are not used contrastively in the phonological system of their language(s), young infants do not show such language-specific effects. In the early months of life, infants discriminate both native and non-native contrasts with equal ease. As experience with a native language accumulates, evidence of increased attunement to properties of the ambient language emerges. By the end of the first year, infants, like adults, fail to discriminate many non-native contrasts. In this article, we begin to explore whether there may be a simple statistical basis for the particular order in which non-native contrasts are lost.

Adults show difficulty discriminating non-native single-feature contrasts if the speech sounds are not used to distinguish meaning in their native language. A classic example is the difficulty that native Japanese-speaking adults have in discriminating between English /ra/ versus /la/ (Sheldon & Strange, 1982). The same pattern is evident in English-speaking adults' difficulties in discriminating Hindi retroflex and dental stops (Werker, Gilbert, Humphrey, & Tees, 1981), Nthalamkampx/Thompson Salish velar and uvular ejectives (Werker & Tees, 1984), and Czech retroflex and palatal fricatives (Trehub, 1976).

Infants younger than six to eight months, however, can categorically discriminate both contrasts that are phonemic (e.g., for English, /ba/-/pa/; Eimas, Siqueland, Jusczyk, & Vigorito, 1971) and contrasts that are not phonemic in their native language (for a review, see Werker & Tees, 1999). By 10 to 12 months, infants perform like adults in their discrimination of non-native contrasts. This change in discriminative abilities within the first year of life has been reported for English-learning infants for several different consonant contrasts including the Hindi stops, the Salish ejectives, and the Czech fricatives cited above, as well as several different Zulu contrasts — bilabial plosive/implosive, lateral voiced/voiceless fricative, and velar voiceless/ejective distinctions (Best, 1995). Similar loss of discriminatory ability has been reported for Japanese-learning infants' discrimination of English /t/ and /l/ (Kuhl, 1998). The same developmental pattern emerges in the perception of vowel contrasts, although the discriminative decline appears to occur somewhat earlier. Kuhl, Williams, Lacerda, Stevens, and Lindblom (1992) found evidence of distinctive prototype effects for vowels in English- and Swedish-learning infants by six months, and Polka and Werker (1994) found effects of native language experience on perception of vowels between four and six months.

Initially, this developmental pattern was explained as an absolute loss of ability due to lack of exposure (see Werker, 1994 for a detailed historic overview). Such a Maintenance Theory assumes that experience is required to maintain perceptual sensitivities that are already present in the infant (Aslin & Pisoni, 1980). Evidence from Best, McRoberts, and Sithole (1988), however, contradicts such a view. Infants and adults were tested on their ability to discriminate a Zulu apical versus lateral click contrast in which neither sound was like any English sound. Unlike other non-native contrasts previously tested, both older infants and adults showed high levels of
discrimination. Moreover, adults' ability to discriminate certain non-native contrasts can improve with training (see Logan, Lively, & Pisoni, 1991; Werker & Polka, 1993). These findings suggest that the loss of ability is not absolute, but is instead due to perceptual reorganization (Werker, 1994). Two dominant theories used to explain these developmental changes are the Perceptual Magnet Effect (or Native Language Magnet Theory, NLM; Kuhl, 1991, 1995, 2000) and the Perceptual Assimilation Model (PAM; Best, 1993, 1995; Best & McRoberts, this volume; Best, McRoberts, & Sithole, 1988). Though differing in many of their particulars, these theories assume a restructuring of perceptual space or a redistribution of attention rather than an absolute loss of ability.

Kuhl's NLM, originally formulated to account for the changes in vowel discrimination noted earlier but intended to apply to consonant discrimination as well, proposes that experientially derived prototypes function as perceptual magnets for other sounds in speech perception. These prototypes act to warp perceptual space, serving as attractors in speech sound discrimination by pulling other members of the category towards themselves. Thus, nonprototypical members of categories are perceived as more similar to the category prototype than to each other, even though the physical distance between the stimuli may be equivalent. The fashion in which perceptual space is warped is a function of learning about the phonological organization of the native language.

Best's PAM, devised to account primarily for the changes in consonant discrimination noted earlier but intended to account for vowel discrimination as well, hypothesizes that incoming speech sounds are perceptually assimilated to the phonemic categories of the native language whenever possible. Non-native phones are perceived with respect to their similarities and differences to native phones. If a non-native phone is reasonably similar to a native-category phone, then listeners are likely to assimilate it to that native category. In the case of discrimination of non-native contrasts, if both phones are assimilated to a single native category, listeners will show difficulty in discrimination. Therefore, the ability to discriminate non-native contrasts is dependent upon how similar the sounds are to those found in the native language. Best (1993, 1995) offered a taxonomy for classifying how non-native contrasts may relate to native language phonology. First, two non-native phones may be assimilated to a single native phonetic category; such Single Category contrasts should be difficult to discriminate. Second, each of the non-native phones may be assimilated to a separate native phonetic category; Two Category contrasts should be easy to discriminate. Third, one of the native phones may be a good exemplar of a native category, while the other is a poor exemplar (cf. Volaitis & Miller, 1992); Category Goodness contrasts should also be relatively easy to discriminate. Fourth, one or both of the non-native phones may be unlike any native phone, though both are perceived as speech sounds; Uncategorizable contrasts will be easy if one of the phones is similar to a native language phone, but difficult otherwise. Fifth, the two non-native phones may be perceived as nonspeech sounds; the difficulty of discriminating Nonassimilable contrasts depends on the acoustic distance between the phones.

Although these accounts predict which non-native contrasts should become difficult to discriminate and which should remain easily discriminable, they do not
directly predict either how or when changes in perception of such contrasts should occur. Studies investigating infants' discrimination of non-native contrasts traditionally used measures involving achievement of some criterion of performance. This, coupled with relatively wide separation of the age groups tested, may make it appear as though loss of discrimination is categorical. No evidence, however, exists to support the notion that development in this domain is discontinuous; more likely, previous findings reflect endpoints of a continuous (albeit fairly rapid) process. The more recent adoption of graded measures of performance, such as A’ (e.g., Pegg & Werker, 1997) are in keeping with this view.

With respect to the question of when, the logic of predicting the relative order in which non-native contrasts are lost is straightforward, at least for Single Category contrasts. If non-native contrasts that are discriminable early in development later become indiscriminable due to their relation to emerging native-language phonetic categories, then the order in which non-native contrasts become indiscriminable should correspond to the order in which corresponding native-language categories emerge.

NLM and PAM may be embedded within general models of phonological development that seek to account for development of phonetic categories (Behnke, 1998; Plaut & Kello, 1999). Jusczyk's (1993, 1997) WRAPSA model provides a broad framework in which development of speech perception may be viewed as one component involved in acquiring a receptive lexicon and spoken word recognition skills; models such as PAUSE:mir (Werker & Curtin, in press) and DRIBBLER (Morgan, Singh, Borufeld, Rathbun, & White, 2001) have recently begun to fill in details of this framework. All of these models devote particular attention to the statistical properties of the input; for example, Behnke (1998, p. 43) writes, “The development of phonetic categories] is initially mainly based on the distributional characteristics of incoming signals and is the first step in the direction of language-specific processing of speech.” This concern with input statistics mirrors both recent developments in phonological theory (e.g., Pierrehumbert, 2001, in press) and recent findings in infant speech processing.

In several domains, infants have proven to be highly capable statistical learners. Newborn infants can discriminate sets of words based on distributions of correlated acoustic and phonetic features within the sets (Shi, Werker, & Morgan, 1999). A series of studies has shown that infants are able to use statistical regularities of the input to discover boundaries of word-like units (Aslin, Saffran, & Newport, 1998; Goodisitt, Morgan, & Kuhl, 1993; Saffran, Aslin, & Newport, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Infants are sensitive to the relative frequency of phonotactic sequences (Jusczyk, Luce, & Charles-Luce, 1994) as well to the likelihood of particular sequences occurring within versus between words (Mattys, Jusczyk, Luce, & Morgan, 1999). Recent work by Maye, Werker, and Gerken (2002) has shown that infants' sensitivities to input statistics also extend to learning of phonetic categories. The distributions of speech sounds with which six to eight month old infants were presented were shown to affect the way in which they later classified the sounds: Infants presented with the sounds in a bimodal distribution formed two categories whereas those exposed to a unimodal distribution formed only one.
According to a statistical learning account, the statistics of the input language play a fundamental role in the warping of perceptual space and the formation of native language categories, as indeed Kuhl (2000) has suggested. Maye et al. (2002) have demonstrated that the location of input exemplars in phonetic space influences category development, but the frequencies of exemplars within regions must also play a role. Unless categories are extremely well separated, variation within categories is highly restricted, and infants’ perceptual and representational capacities are non-noisy (a trio of unlikely assumptions), reasonably large numbers of exemplars will be needed for learners to be able to infer whether these are drawn from one or many underlying distributions. Moreover, if phonetic category acquisition were insensitive to exemplar frequencies, then sparse sets of exemplars might suffice to trigger formation of new categories: Even brief exposure to another language might suffice to alter the organization of the learner’s phonological system. Available data suggest to the contrary, however, that exposure to varied (and hence, necessarily, numerous) sets of instances is required for robust category formation (Logan, Lively, & Pisoni, 1991; see also Bomba & Siqueland, 1983; Quinn, Eimas, & Rosenkrantz, 1993). Therefore, we hypothesize that, other things being equal, the order of emergence of native-language phonetic categories will mirror the frequencies with which exemplars of those categories appear in the input. This hypothesis provides a basis for predicting when changes in perception of non-native contrasts should occur: Non-native contrasts relating to categories of sounds with higher frequencies in the native language should be lost earlier.

In particular, for English-learning infants, discriminability should begin to decline earlier for non-native coronal contrasts than for non-native dorsal or labial contrasts. As every contestant on Wheel of Fortune knows, the most frequent English consonantal phonemes are /t/, /d/, /s/, and /n/ — all coronals (French, Carter, & Koenig, 1930; Tobias, 1959). Previous studies investigating infants’ ability to discriminate non-native contrasts have focused on either coronal or dorsal stop consonants. Studies by Werker and colleagues have used coronal and dorsal voiceless stop contrasts whose crucial differentiating factor is place of articulation (e.g., dental vs. retroflex, both of which are coronal, or velar vs. uvular, both of which are dorsal; Werker & Tees, 1984; Werker et al., 1981). The ability to discriminate both coronal and dorsal non-native contrasts disappears between 6 and 10 months, but the relative order in which these contrasts are lost is not known. We hypothesize that, due to the higher frequency of coronals in English, discrimination of non-native coronal contrasts should decline earlier than discrimination of non-native dorsal contrasts.

2 Experiment 1

Infants were tested on their ability to discriminate a coronal Hindi retroflex /dental stop contrast (©/tɛə/) and a dorsal Salish velar /uvular ejective contrast (/kɔə/) in previous studies, infants have been tested on one or the other of these two contrasts. Using the Conditioned Head Turn Procedure, Werker et al. (1981) and Werker and Tees (1984) showed that 6–8-month-old infants were able to discriminate the coronal Hindi contrast at native levels. By 8–10 months, infants showed a slight decline in performance and by 10–12 months, they were no longer able to discriminate the contrast. Testing infants on the dorsal Salish contrast yielded
similar results (Werker & Tees, 1984). Because we were concerned with relative order of loss, we tested the same infants on both contrasts. Although there has been no large-scale study that has systematically tested the same infants on both contrasts, Werker and Tees (1984) reported a small-scale longitudinal study in which six infants were repeatedly tested on both the coronal Hindi and dorsal Salish contrasts. Contrary to our prediction, at eight months, all infants reached criterion on the coronal Hindi contrast, but only half reached criterion on the dorsal Salish contrast. However, whether these initial results reflect differences in discriminative ability or whether they might be ascribable to repeated testing or testing order effects is not clear.

2.1 Method

Subjects. Subjects were 18 full-term infants (10 males, 8 females) recruited from Rhode Island birth records. Subjects were between 7.9 and 9.1 months (mean = 252 days, range = 236 to 273 days) and were from English-speaking homes (exposure to another language was limited to less than 10%). An additional 13 infants were excluded from the study due to crying or fussiness that prevented completion of one or both tests (5), experimenter error (2), and loss of interest in reinforcers (5); one family rescheduled their appointment several times so that their infant was too old when finally tested.

Stimuli. Stimuli were the same as those used in previous experiments by Werker et al. (1981), Werker and Tees (1984), and Best (1995). The coronal contrast was a Hindi voiceless unaspirated retroflex ([t̪a]) versus dental ([t̪a]) distinction. English-speaking adults tend to perceive this as a Single Category contrast (Werker, 1991). Retroflex consonants are produced by curling the top of the tongue back and forming a closure posterior to the alveolar ridge. Dentals are produced by placing the tip (or blade) of the tongue against the back wall of the upper front teeth (Werker & Tees, 1984). The closure for English voiceless coronal stops usually occurs at the alveolar ridge, between the closure points for the Hindi consonants.

The dorsal contrast was an Nthalamkampx (an Interior Salish language spoken in British Columbia) voiceless unaspirated glottalized velar ([k̃]̃) versus uvular ([q̃]̃) stop distinction (Werker & Tees, 1984). English-speaking adults tend to perceive this as either a Single Category or Nonassimilable contrast (Werker, 1991; Best & McRoberts, this volume, characterize both of the contrasts used here as Single Category contrasts). The velar versus uvular place of articulation is the crucial difference. The sounds are produced by obstructing the airflow by raising the back of the tongue either against or behind the velum. In English, only velar stops carry significance. Infants were tested on their ability to discriminate sets of stimuli, three for each speech sound. By using multiple exemplars, subjects were forced to ignore within-category acoustic variability and differentiate the sounds according to phonetic category (Werker & Tees, 1984).

The contrasts were both produced by male native speakers. For the coronal contrast from Hindi, Werker and colleagues selected three exemplars from each category from multiple recordings so that variations in fundamental frequency, duration and intonation were randomized both within and between categories (Werker...
et al., 1981; Werker & Tees, 1984). All of the coronal stimuli displayed similar falling-rising pitch contours, and we further edited them to ensure that duration and RMS amplitude were equal. Werker and Tees note that they had difficulty in eliciting sets of Nthalamkampx stimuli with the same vowel. They identified one instance of [kɔ:] and one instance of [qɔ:] that had similar vowels and then created sets of three stimuli each by splicing these vowels onto different bursts. Consequently, there was more variability among the coronal stimuli than among the dorsal stimuli. We tested whether this affected the intrinsic discriminability of the contrasts in Experiments 2 and 3 below, with six-month-olds and adults, respectively.

In previous reports, the Hindi phones were transcribed with a following vowel [a] and the Salish phones were transcribed with a following vowel [i] (Werker et al., 1981; Werker & Tees, 1984). According to Werker (personal communication; March 7, 2002), these transcriptions reflect the relative values of these vowels in their respective vowel systems. However, these two vowels have very similar formant values; in comparison to English, the two vowels fall in a portion of F1 F2 vowel space labeled as [ɔ] by Peterson and Barney (1952), most closely described as a midcentral vowel. In listening to the vowels we could detect no trace of retroflection and therefore have transcribed them as [ɔ].

To verify that coronal stops occur more frequently than dorsal stops in typical speech to children, we examined transcriptions of mother to child speech obtained from the CHILDES database (MacWhinney, 1999). The four corpora we used (Bernstein's corpora; Brown's Adam, Eve, and Sarah corpora) all involved mother-child dyads. Transcriptions of the mothers' speech began when the children were between one and two years and continued for at least four months (Bernstein, 1982; Brown, 1973). Across these four corpora there were approximately 300,000 words uttered. Each transcribed word type was converted into a phonemic pronunciation using the on-line Wordsmith dictionary; token pronunciations were adjusted for effects of phonological contexts in continuous speech by applying rules of English phonology suggested by Ladefoged (2001). We reasoned that this procedure would suffice to provide rough estimates of phoneme frequencies.

In the four corpora, there were approximately 750,000 consonantal phonemes. Phoneme token frequencies were evaluated with respect to features of both place (labial, coronal, dorsal) and manner of articulation (nasal, liquid/semivowel, fricative, affricate, stop). There are multitudes of ways in which frequency of these phonemes can be computed. Not only could infants be representing sounds in different ways (e.g., features, phones, demisyllables), but they could also be computing frequency with respect to different contexts (e.g., word-initial, syllable-initial, or perhaps overall frequency). However, as is evident in Table 1 (overleaf), across a wide variety of methods of measurement, the result is always the same: in English, voiceless coronal stops are more frequent than voiceless dorsal stops ($\chi^2; p < .001$ for all counts).
TABLE 1
Occurrence of coronal and dorsal sounds in American English child-directed speech

<table>
<thead>
<tr>
<th>Position</th>
<th>Voiceless stops</th>
<th>All oral stops</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coronal /t/</td>
<td>dorsal /k/</td>
</tr>
<tr>
<td>Word Initial</td>
<td>18099</td>
<td>14375</td>
</tr>
<tr>
<td>Syllable Initial</td>
<td>26555</td>
<td>20062</td>
</tr>
<tr>
<td>Any</td>
<td>106832</td>
<td>38325</td>
</tr>
</tbody>
</table>

Procedure. We used a modified version of the Conditioned Head Turn procedure (for a description, see Werker, Polka, & Pegg, 1997). The experimental procedure included two phases, conditioning and testing. Throughout the procedure, the infant was seated on a parent’s lap across a small table from an experimenter in a sound-treated room. A loudspeaker (JBL Sentry 100A) and visual reinforcers were located at a 90-degree angle to the left of the infant and parent. The visual reinforcers included two mechanical toys in smoked Plexiglas boxes, which could be lighted and activated, and four Sony Trinitron televisions arranged in a cube on which an animated movie could be displayed. A Comprehensive Video matrix switcher connected a video cassette player to the televisions, so that the movie could be displayed on all four or any combination of the televisions. To maintain infants’ interest, the reinforcers that were activated varied across trials; once activated, reinforcers remained on for 4500 ms.

Both the experimenter and the parent listened to music over Bose noise-cancellation headphones to prevent them from hearing the stimuli. A second experimenter located outside the room operated the computer and observed the testing session via a Panasonic (BP-WV550) low light video camera and a Panasonic (WV-540) monitor. The experiment was controlled by custom software running on a PC-compatible computer with a TBS Montego II Sound Blaster Pro Emulation soundboard. Stimuli played through an Onkyo Integra stereo preamplifier (P-304) and an Onkyo Integra stereo power amplifier (M-504) and were set to play at a conversational level (66 dB).

For each infant, one stimulus set of the contrast pair being tested was designated as the background set; the other was the target, or change, set. Assignment of stimulus sets to background or target status was counterbalanced across infants. The background stimulus was repeated continuously with 1300 ms interstimulus intervals throughout the procedure except during target trials as noted below.

During the conditioning phase, the infant was given the opportunity to learn the association between a sound change and the activation of the visual reinforcers. In this phase of the experiment, only one exemplar each from the background and target sets were used. When the infant was judged to be in a state of readiness, quietly attending at midline, the experimenter could initiate a trial, during which the target stimulus played three times (again with 1300 ms ISIs). This allowed a response window of approximately five seconds. Afterwards, the background stimulus resumed. Conditioning proceeded by initially activating the reinforcers as soon as the first
target stimulus was presented. After the first three trials, a delay was inserted between presentations of the target stimulus and activation of the reinforcers. The addition of a delay allowed the infants to initiate head turns prior to the activation of the reinforcers. This delay was gradually lengthened to a maximum of 4.5 seconds. Following the standard established by Pegg and Werker (1997), once the infant made three consecutive anticipatory head turns (i.e., before the reinforcers were automatically activated) or reached a maximum of 15 trials, testing began.

In testing, all three exemplars from the background and target sets were used. Presenting infants with three exemplars required them to attend to categorical differences among the speech stimuli in order to discriminate successfully. Between trials, each background stimulus was repeated three times, after which a new background stimulus was chosen (with replacement) from the background stimulus set. As before, when the infant was judged to be in a state of readiness, the experimenter could call for a trial, which was initiated at the first possible interstimulus interval. The type of trial, control or test, was randomly selected by the computer so that change trials occurred on about 55% of the trials and control trials occurred on 45% of the trials. No more than two control or change trials were allowed consecutively. In each trial, one stimulus chosen at random from the appropriate stimulus set was repeated three times with ISIs of 1300 ms. As in conditioning, this allowed a total response window of approximately five seconds. Trials in which a target stimulus was presented and the infant turned were counted as hits; trials in which a background stimulus was presented and the infant did not turn were counted as correct rejections. Correct responses comprised hits and correct rejections.

Seven correct responses out of any eight consecutive trials was adopted as a floating criterion of discrimination (see Aslin & Pisoni, 1980; Jusczyk, Shea, & Aslin, 1984; Kuhl, 1985). Testing concluded when the infant reached this criterion or completed 26 trials. If on the 26th trial, the infant was within two trials of reaching criterion, an additional five trials were presented. Infants who failed to reach the criterion and had fewer than three hits, or fewer than five head turns total (hits plus false alarms) were judged to have lost interest in the reinforcers and to no longer be on task. Data from these infants were discarded.

During testing, if an infant had three consecutive incorrect responses, a retraining stage was entered. In retraining, the infant was presented with only the conditioning stimuli, and every trial presented was a change trial. If no head turn occurred on the first three retraining trials, the reinforcers were automatically activated on the fourth trial. After two consecutive correct responses or five retraining trials, testing resumed. Each infant was limited to three retraining sets, which were not included in the analyses.

Infants were tested on both contrasts in two sessions on the same day, counterbalanced for the order of presentation of the contrasts (coronal or dorsal).

---

1 We had intended this to be 25 trials, as is the norm, but our software mistakenly began numbering trials from 0 rather than 1.
2.2 Results & Discussion

Conditioning proceeded in a similar fashion for both of the contrasts tested. The rate at which trials were presented during conditioning did not differ across the contrasts. For the coronal contrast, there was an average of 12.07 background stimuli (approximately 18 seconds) between trials; for the dorsal contrast, there was an average of 11.48 background stimuli (approximately 17 seconds) between trials, $t(17) = 0.76, p > .4$. Also, the average number of conditioning trials per infant did not differ across contrasts (coronal: 12.77, dorsal: 11.33, $t(17) = 1.57, p > .1$). However, whereas 14 of 18 infants reached the conditioning criterion (3 consecutive anticipatory head turns in fewer than 15 trials) on the dorsal contrast, only 11 did so on the coronal contrast. Ten infants reached the conditioning criterion on both contrasts; three infants did so on neither. We consider later whether performance in testing was associated with performance during conditioning.

Two types of dependent measures, both regularly used in previous work, were used to assess infants' performance in the testing phase. First, subjects were scored categorically on their ability to discriminate each contrast using a criterial measure. Second, A' scores were calculated to provide graded measures of performance. A' is the more suitable measure, assuming that loss of discriminability of non-native contrasts is a continuous process; we include measures of criterial performance for compatibility with previous reports.

On the coronal contrast, seven of 18 subjects reached the pre-established floating criterion of seven correct responses on any eight consecutive trials (within 15 trials on average). On the dorsal contrast, 11 of 18 subjects reached criterion (within 17.3 trials on average). Three infants reached criterion on both contrasts; three others reached criterion on neither. Eight infants reached criterion on the dorsal contrast, but not the coronal, whereas four infants reached criterion on the coronal contrast, but not the dorsal. This asymmetry is consistent with the hypothesis that sensitivity to the coronal contrast is lost first, but the difference is not significant by McNemar's Q, $p > .35$.

The 7-of-8 floating criterion has been used in a number of studies employing the conditioned head turn procedure (Kuhl, 1985; Polka, Jusczyk, & Rvachew, 1995; Werker, Polka, & Pegg, 1997). The justification for this criterion is that the binomial likelihood of seven correct responses in eight trials is less than 0.05 ($p = .035$), so that claims that individual infants can discriminate particular contrasts may be justified. However, application of the 7-of-8 criterion is problematic, for at least two reasons. First, subjects may perform at a level that exceeds this criterion, but nonetheless fail to meet it. Consider a subject who performs correctly on six consecutive trials, then misses two trials, and then gets six more correct. This subject has not reached the 7-of-8 criterion, even though the likelihood of 12 correct in 14 trials is five times less than that of seven of eight. This may be straightforwardly remedied by considering that any performance whose likelihood is equal to or less than the likelihood of 7-of-8 meets criterion: What matters is not any particular sequence of responses, but rather whether infants' performance exceeds the standard.

This raises the second problem: What is the likelihood of 7-of-8? The true
likelihood of reaching this criterion is not .035, unless subjects receive only eight trials. When applied as a floating, rather than blocked, criterion, likelihood increases according to the total number of trials. In nine trials, the likelihood of reaching 7-of-8 by chance is .049; in 10 trials, the likelihood is .061. To estimate the likelihood of 7-of-8 in larger numbers of trials, we conducted a series of Monte Carlo simulations, each including 10,000,000 random sequences of $n$ trials; results of these simulations are shown in Table 2. Also shown in Table 2 are the number correct needed in $n$ trials to achieve likelihoods less than or equal to those of 7-of-8.\(^2\)

**TABLE 2**

Floating 7-of-8 criterion likelihoods

<table>
<thead>
<tr>
<th>Trials</th>
<th>Criterion Likelihood</th>
<th>Likelihood-equivalent number correct</th>
<th>Trials</th>
<th>Criterion Likelihood</th>
<th>Likelihood-equivalent number correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>.035</td>
<td>7</td>
<td>18</td>
<td>.156</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>.049</td>
<td>8</td>
<td>19</td>
<td>.167</td>
<td>13</td>
</tr>
<tr>
<td>10</td>
<td>.061</td>
<td>8</td>
<td>20</td>
<td>.178</td>
<td>13</td>
</tr>
<tr>
<td>11</td>
<td>.074</td>
<td>9</td>
<td>21</td>
<td>.188</td>
<td>14</td>
</tr>
<tr>
<td>12</td>
<td>.086</td>
<td>9</td>
<td>22</td>
<td>.199</td>
<td>14</td>
</tr>
<tr>
<td>13</td>
<td>.097</td>
<td>10</td>
<td>23</td>
<td>.210</td>
<td>14</td>
</tr>
<tr>
<td>14</td>
<td>.109</td>
<td>10</td>
<td>24</td>
<td>.220</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>.121</td>
<td>11</td>
<td>25</td>
<td>.231</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>.133</td>
<td>11</td>
<td>26</td>
<td>.241</td>
<td>16</td>
</tr>
<tr>
<td>17</td>
<td>.144</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Applying the likelihood-equivalent criteria, performance on the coronal contrast remains the same: Seven of 18 subjects reach criterion. Given that the likelihood of reaching the floating criterion by chance within 26 trials is 0.24, we cannot rule out the possibility that seven of 18 infants might do so by chance (binomial $p = .12$). However, under this standard, 14 of 18 subjects reached criterion on the dorsal contrast (within 16.6 trials on average; the likelihood of this occurring by chance is $p < .0001$). Five subjects reached criterion on both contrasts; two reached criterion

\(^2\) These likelihoods perhaps render the criterion unsuitable as a basis for claims about individual subjects. However, the criterion may be used as a basis for binomial analyses of group performance, with a substantial advantage over the default assumption of $p = q = .50$ that must be used elsewhere (e.g., in assessing the numbers of subjects who listened longer to one stimulus set over another in the Headturn Preference Procedure). For example, under the default assumption, the likelihood of 8 of 10 infants reaching criterion is .055, whereas with the probability of reaching 7-of-8 within 25 trials, the likelihood drops to .0002, and only 6 of 10 must reach criterion ($p = .013$) to exceed the .05 standard.
on neither. Under this criterion, the asymmetry between the two contrasts is more pronounced: Nine subjects reached criterion on the dorsal contrast, but not the coronal, whereas only two showed the opposite pattern. However, as before, this asymmetry falls short of significance, McNemar’s Q, \( p < .07 \).

Although we believe that the likelihood-equivalent approach to calculating a criterion of performance is logically more consistent, arguably, neither form of criterion is entirely relevant to our present enterprise. Our fundamental concern is the relative discriminability of the coronal and dorsal contrasts tested; we do not wish to advance claims that infants can discriminate this contrast or cannot discriminate that one. Use of continuous measures that can be analyzed with parametric statistics are best suited for deciding questions of relative discriminability. Thus, as a more sensitive measure of discrimination, subjects’ A’ scores were calculated for each of the contrasts. A’ is a measure of sensitivity similar to d’ but designed for use on a smaller number of trials (Grier, 1971), which takes into account the proportion of hits and false alarms to give an estimate of a subject’s ability to detect the sound change given the overall frequency of head turns. Scores range up to 1. By definition, .5 corresponds to a chance level of performance (hits = false alarms), whereas A’ scores that are significantly below chance suggest that subjects are using a strategy different from that required for the task. Polka (1991) defined adult “native-like” performance at .95, corresponding to a 10% error rate; however, we would expect the performance of infants to be somewhat lower.

A’ scores were computed for each subject and compared across contrasts to chance levels of performance. Results are shown in Figure 1. Dorsal A’ scores were significantly higher than chance (\( M = .73, SD = .19, t(17) = 5.26, p < .0001 \)), whereas coronal A’ scores were not (\( M = .48, SD = .40, t(17) = -0.26, p > .75 \)). A 2 x 2 Contrast by Order mixed ANOVA on these data showed a significant main effect of Contrast, \( F(1, 16) = 4.95, p < .05 \). The order in which the contrasts were tested made no significant difference, \( F(1, 16) = 2.88, p > .10 \), nor was there a significant Contrast by Order interaction, \( F(1, 16) < 1, NS \). Thirteen of 18 infants had higher A’ scores for the dorsal contrast than for the coronal contrast, binomial \( p < .05 \).

We may also ask whether the two contrasts engendered different biases in responding. In sessions testing the coronal contrast, infants turned on 32.1% of the trials (hits plus false alarms); in sessions testing the dorsal contrast, infants turned on 39.4%. Twelve of 18 infants showed higher rates of turning to the dorsal contrast (binomial \( p = .12 \)). The difference in turning rate across contrasts was not significant, \( t(17) = 1.27, p > .2 \). Infants did, however, turn more during the first session (42.4%) than during the second session (29.2%), \( t(17) = 2.59, p < .02 \).

As noted earlier, we followed Pegg and Werker (1997) in allowing infants to proceed to testing after 15 conditioning trials, even if they had not reached the criterion of three consecutive anticipatory head turns. Our motivation for doing so was two-fold. First, limiting the length of the sessions increased the likelihood that infants could successfully complete both tests, thereby helping to minimize subject mortality. Second, restricting the variance in amount of exposure to both the contrasts and the reinforcers was in keeping with our general strategy of counterbalancing. Our intent was to treat infants in the same fashion during both sessions; that this was accomplished
is demonstrated by the data reported at the beginning of this section. Nevertheless, 14 of 18 infants reached the conditioning criterion on the dorsal contrast, whereas only 11 of 18 infants did so on the coronal contrast. Here we consider whether differences in test performance across the contrasts might have been associated with differences in conditioning.

On the dorsal contrast, there was an association between achievement of the conditioning criterion and achievement of the (likelihood-equivalent) test criterion, Fisher exact $p = .02$. Of the 14 infants reaching the conditioning criterion, 13 also reached the test criterion. However, this pattern did not hold for the coronal contrast. Of the 11 infants reaching the conditioning criterion on this contrast, only five also reached the test criterion. On this contrast, there was no association between the achievement of these two criteria, Fisher exact $p = .42$. For both contrasts, subjects failing to reach the conditioning criterion also tended to fail to reach the test criterion, by a ratio of three to one. We also considered whether number of conditioning trials might be correlated with A’ scores for the coronal contrast. Although the statistic had the expected sign, it fell well short of significance, $r = -0.14$, $p > .25$ (1-tail).

Ten infants reached the conditioning criterion on both contrasts. Half of these were tested first on the dorsal contrast, and half were tested first on the coronal contrast. All 10 of these infants reached the test criterion on the dorsal contrast (binomial $p < .0001$), whereas five reached the test criterion on the coronal contrast (binomial $p < .07$). Within this subsample, A’ scores were significantly higher on the dorsal contrast ($M = .79$, $SD = .13$) than on the coronal contrast ($M = .48$, $SD = .37$), $t(9) = 2.44$, $p < .05$. Eight of 10 had higher A’ scores on the dorsal contrast than on the coronal contrast, binomial $p < .06$. The essential question is whether continuing
conditioning until all infants had reached criterion would have substantively changed our results. The data suggest that, if anything, the differences between the two contrasts might then have been more pronounced. There is no indication that the differences we observed in test performance were attributable to differences in conditioning.

Werker and Tees (1984) and many subsequent studies include a control condition that we omitted. When an infant failed to reach test criterion, before deciding that that infant could not discriminate the contrast, Werker and Tees required that the infant pass criterion on an English-language contrast, /ba/ versus /da/. Werker and Tees (p. 54) note, “This was done to ensure that the failure of the infant was due to an inability to readily perceive the sound difference, and was not due to nonspecific factors such as boredom, dirty diapers, etc.”

One motivation for eliminating this control was feasibility: The likelihood of getting infants through three sessions in a single day is remote. More important, however, we believe that, given our design, this additional session is logically unnecessary: Counterbalancing the order of testing served to strain out extraneous factors. The essential use of this additional session is to provide a basis for interpreting null results by assessing continued interest in reinforcers and ensuring that the infant has remained on task. As noted earlier, we do not advance claims that infants fail to discriminate either or both of the contrasts tested, but rather make the weaker assertion that performance differs across the two, in a fashion that is predictable from the statistical properties of input speech. Our argument does not rest on null results, so the interpretive difficulty with which Werker and Tees were concerned does not arise.

Moreover, the efficacy of the control session is unclear, given its lack of simultaneity with the behavior of interest. Between sessions, performance may vary for any of a multitude of reasons: For example, the diaper gets changed or becomes soiled, and the infant’s performance changes for better or worse accordingly. Thus, an infant who was inattentive during a session testing a non-native contrast may become attentive on a following control session testing a native contrast, or vice-versa. No necessary logical relation holds between performance in the two sessions. Interest in reinforcers may be gauged by observing infant’s overall turning rate within a session; data from infants failing to turn at some criterion rate may be excluded, as we did here. Turning rate provides a clue to whether an infant remains on task; beyond this, counterbalancing provides the most appropriate solution. Recall that half the infants were tested first on the coronal contrast, and half were tested first on the dorsal contrast, so that any extraneous, nonspecific influences on performance were randomized across the two contrasts.

The results of this experiment indicate that 8.5-month-old English-learning infants are better able to discriminate a non-native dorsal contrast, corresponding to a lower frequency English phonological category, than they are to discriminate a non-native coronal contrast, corresponding to a higher frequency category. There was a tendency for more infants to reach criterion on the dorsal contrast than on the coronal contrast (either the traditional 7-of-8 criterion, or its probabilistic equivalent); more important, infants were significantly more sensitive to the dorsal contrast than the coronal contrast, as measured by A’ scores. These results are consistent with our
statistical account: The more frequent category appears to be drawing in the non-native contrasts sooner.

However, it is possible that the 8.5-month-olds were better at discriminating the dorsal contrast for reasons other than frequency. Perhaps the reason for the discrepancy in performance across the contrasts was simply because the dorsal contrast was inherently easier to discriminate or because the dorsal exemplars were less variable than were the coronal exemplars. These possibilities were evaluated in the next two experiments.

3 Experiment 2

Werker and Tees' (1984) results provide an initial argument against the hypothesis that the coronal and dorsal contrasts we tested differed in their inherent discriminability. Younger infants (6- to 8-month-olds) were tested on either the coronal or the dorsal contrasts; the proportion of subjects reaching criterion on the two contrasts was equivalent. This was, however, a between-subjects test. To evaluate explicitly whether the dorsal contrast was easier to discriminate, it is necessary to perform a within-subjects test on infants who are young enough to discriminate both contrasts.

3.1 Method

Subjects. Twenty-four English-learning infants between six and seven months old were tested in this procedure (11 males, 13 females; $M = 199$ days, range = 180–212 days). An additional 29 infants were tested and subsequently excluded from analysis due to crying or illness (11), failure to turn during testing (7), A' scores significantly below chance on both contrasts (3), failure to complete the second Session (3), parental or sibling interference (3), experimenter error (1), and foreign language exposure in excess of 10% (1).

Stimuli. The stimuli used were the same Salish (dorsal) and Hindi (coronal) contrasts described in Experiment 1.

Procedure. Infants were tested using the CHT procedure outlined in Experiment 1. To accommodate the younger infants, two changes were made to the procedure. First, in pilot tests, many of the younger infants made headturns on change trials immediately following the 4.5 second window, thus not activating the reinforcer. Because of this, we increased the reinforcer window to 4.75 seconds. This slight increase aided in capturing the slower responses of the younger infants.

Second, once we began to test the infants, it became evident that performance on the second test was greatly reduced. There were a much higher number of infants discarded for crying or failure to turn to the reinforcer compared to the 8.5-month-olds. For these reasons, after testing 12 infants in one-day sessions, we tested the remaining infants on both contrasts over two different days. The second testing session always occurred within three days of the first. As in Experiment 1, testing was counterbalanced for both the order of presentation of the contrasts (coronal or dorsal) and for the syllable that was used as the background stimulus (i.e., retroflex or dental, velar or uvular).
3.2 Results & Discussion

Fourteen of 24 infants reached the conditioning criterion (3 consecutive anticipatory head turns in fewer than 15 trials) on the dorsal contrast, as did 12 of 24 infants on the coronal contrast. Nine infants reached the conditioning criterion on both contrasts; seven infants did so on neither. As was the case for the older infants, there was no significant difference in the number of conditioning trials for the two contrasts (Coronal: 12.08, Dorsal: 11.29, t(23) = .806, p > .4).

Nineteen of 24 infants reached the 7-of-8 criterion on the dorsal contrast (within 15.37 trials on average), whereas 17 of 24 infants reached the 7-of-8 criterion on the coronal contrast (within 13.82 trials on average). Fourteen infants reached this criterion on both contrasts. Five infants reached this criterion on the dorsal contrast but not on the coronal contrast whereas three infants exhibited the reverse pattern. There was no difference in the proportion of subjects reaching the 7-of-8 criterion on the two contrasts by McNemar's Q, p > .70.

Twenty-two of 24 infants reached the likelihood-equivalent criterion on the dorsal contrast (within 14.77 trials on average), as did 18 of 24 infants on the coronal contrast (within 12.78 trials on average). Seventeen infants reached this criterion on both contrasts. Five infants reached this criterion on the dorsal contrast but not on the coronal contrast, whereas only one infant exhibited the reverse pattern. By McNemar's Q, this asymmetry was not significant, p > .20.

A' scores were calculated for the infants for both contrasts and were analyzed in a 2 × 2 × 2 Contrast (coronal vs. dorsal) by Group (same day vs. different days) by Order mixed ANOVA. There were no main effects or interactions; except for the main effect of Order, F(1, 20) = 1.03, NS, in all instances F(1, 20) < 1. As Figure 2 indicates, both coronal and dorsal A' scores were significantly higher than chance levels of performance, coronal, t(23) = 6.603, p < .0001; dorsal, t(23) = 10.992, p < .0001.

Results from this experiment fail to provide evidence in support of the hypothesis that there was an inherent difference in the difficulty of the coronal and dorsal contrast tested: 6.5-month-old infants discriminated both contrasts equally well. This was evident in the proportion of subjects reaching criterion, as well as in the subjects' A' scores on the two contrasts. These results are consistent with previous between-subjects tests of these contrasts (Werker et al., 1981; Werker & Tees, 1984).

Younger and older infants' A' scores on the two contrasts across experiments were compared in a 2 × 2 Contrast by Age Group mixed ANOVA, which revealed a significant Contrast by Age Group interaction, F(1, 40) = 4.53, p < .05. A' scores for the coronal contrast were significantly different, t(40) = 3.02, p < .01, whereas A' scores for the dorsal contrast were not, t(40) = .85, p > .40. By 10–12 months, as shown by Werker and Tees (1984), English-learning infants' abilities to discriminate both coronal and dorsal non-native contrasts have declined. This pattern of results is consistent with our prediction that the coronal category is acquired earlier, due to its higher frequency of occurrence in input; consequently, decline in discriminability appears earlier for the non-native coronal contrast than for the non-native dorsal contrast.
4 Experiment 3

To further evaluate the possibility that there might be an inherent difference in the discriminability of the coronal and dorsal contrasts tested here, or that these contrasts might vary in their relation to corresponding English categories, we tested adult English speakers on their perception of these contrasts. Adult perception of each of the contrasts used here has been examined previously, though in independent studies (Polka, 1991; Werker & Logan, 1985). Both of these studies used AX tasks.

Werker and Logan tested English speakers on the Hindi dental versus retroflex voiceless stop contrast used in Experiments 1 and 2 here, including four exemplars of each (the 3 exemplars of each used here were drawn from this larger set). Groups of subjects were tested using ISIs of 250, 500, and 1500 ms in five 96-trial blocks; we discuss only results from the last group, because this ISI was most similar to that used with the infants in Experiments 1 and 2 (1300 ms). Werker and Logan reported likelihood of hits and false alarms; from the data given (see their Fig. 3), we calculate that average A’ was about 0.70. Across blocks, subjects showed considerable improvement, from an approximate A’ of 0.60 in the first block to an approximate A’ of 0.80 in the fifth block (see their Fig. 4). Werker and Logan did not report on subjects’ characterization of the stimuli, though elsewhere Werker (1991) has noted that English-speaking adults tend to perceive this as a Single Category contrast.

Polka examined effects of phonemic and phonetic familiarity on identification and discrimination of non-native dorsal contrasts, using both AX and ID tasks; we
note only results from the former. In her first experiment, English and Farsi speakers were tested on the Salish velar versus uvular glottalized voiceless stop contrast used in Experiments 1 and 2 here. Speakers of both languages performed significantly above chance, but significantly below native level (A' = 0.95); for Farsi speakers, there was no benefit of familiarity with the phonemic dimension underlying the contrast (Farsi contrasts velar [g] vs. uvular [g] voiced stops). English speakers showed significant improvement in the AX task between the first and second 90-trial halves. In her second experiment, English speakers were tested on the Salish contrast as well as the Farsi dorsal contrast; subjects performed slightly better on the Farsi contrast, suggesting that there might be some influence of the phonetic familiarity of [g]. Subjects offered varying descriptions of the Salish stimuli; some English speakers characterized these as “funny k's” (consistent with perception of this as a Single Category contrast), whereas others perceived them as unusual combinations of speech sounds (consistent with perception of this as a non-assimilable contrast).

In this experiment, we tested English-speaking adults on an AX task in a within-subjects design on the coronal and dorsal contrasts used in Experiment 1 and 2. Polka instructed some subjects that one of the Salish sounds was similar to /k/, whereas the other was not; other subjects received illustration of velar versus uvular places of articulation. To better foster comparability with infants' performance, like Werker and Logan, we instructed subjects simply that they would hear two syllables in each trial and should judge whether these were the same or different.

4.1 Method

Subjects. Participants included 16 native English-speaking adults, 14 right-handed and two left-handed; 14 female and two male. Partial data from one additional participant were discarded because the software crashed partway through the experiment.

Stimuli. The stimuli used were the same Hindi and Salish contrasts described in Experiment 1. Within each contrast, each of the six stimuli was paired with itself and every other stimulus to form a block of 36 pairs. Within each pair, the syllables were separated by a 1500 ms silence.

Procedure. Participants were tested individually with an AX task, using a PC-compatible computer and Koss headphones. Participants were instructed that they would hear pairs of syllables and should press one key on the keyboard if they thought the syllables were the same or another key if they thought the syllables were different. Participants first received a series of practice trials using endpoint stimuli from an English [da]-[ta] VOT continuum on which feedback was provided. No feedback was provided on trials with experimental stimuli.

Four blocks of 36 syllable pairs each were presented. The first half included one block of coronal stimuli and one block of dorsal stimuli, as did the second half. Subject to this constraint, all possible orderings of blocks were employed (coronal-dorsal-coronal-dorsal, coronal-dorsal-dorsal-coronal, dorsal-coronal-dorsal-coronal, and dorsal-coronal-coronal-dorsal). Within each block, syllable pairs were ordered randomly for each participant. Subjects had an opportunity to pause after every 18 trials.
Participants' responses were scored as hits if they responded “different” to pairs of stimuli drawn from different non-native categories and were scored as correct rejections if they responded “same” to pairs of stimuli drawn from a single non-native category. A' scores were calculated as described earlier. In addition, latencies from the onset of the “X” item to the key press were recorded. Three seconds after the participant's response, the next trial began.

### 4.2 Results & Discussion

Mean A' scores for the coronal and dorsal contrasts are shown in Figure 3. Performance on the coronal contrast ($M = 0.72, SD = 0.08$) was similar to that reported by Werker and Logan ($M \approx 0.70$) and was significantly better than chance, $t(15) = 10.53, p < .0001$ but worse than native-level performance, $t(15) = -10.90, p < .0001$. Performance on the dorsal contrast ($M = 0.74, SD = 0.16$) was very similar to that reported by Polka ($M = 0.77, SD = 0.13$); as in Polka's experiments, performance was significantly better than chance, $t(15) = 6.06, p < .0001$, but also significantly worse than native-level performance, $t(15) = -5.19, p < .0001$. A $2 \times 2$ Contrast by Half repeated-measures ANOVA conducted on the A' scores revealed no main effect of Contrast (coronal vs. dorsal), $F(1, 15) = 0.25, p > .50$, Half (first vs. second) $F(1, 15) = 2.92, p > .10$, and no Contrast by Half interaction, $F(1, 15) = 0.21, p > .50$.

**Figure 3**

Adults' discrimination of coronal and dorsal nonnative contrasts in an AX task

![Graph showing mean A' scores for coronal and dorsal contrasts](image)

Although we found no difference in overall accuracy for the two contrasts, this does not rule out the possibility of more subtle differences in performance. Such differences might have arisen, for example, if the coronal stimuli were more similar to
the corresponding English alveolar [t] than the dorsal stimuli were to the corresponding English velar [k]. For example, participants might have taken longer to respond correctly that dental (coronal) stimuli were different from retroflex stimuli than they did to respond correctly that velar (dorsal) stimuli were different from uvular stimuli, due to increased competition in the former case. To test this, we analyzed participants' latencies. Mean latencies and standard deviations were calculated for each participant, and latencies more than two standard deviations from the participant's mean were discarded. Overall, latencies on dorsal stimulus pairs \( M = 1349.09 \text{ms}, \ SD = 375.64 \) were slightly longer than latencies on coronal stimulus pairs \( M = 1294.66, \ SD = 367.29 \), but this difference was not significant, \( t(15) = 0.92, \ p > .35 \). Separate analyses conducted on latencies for each of the four response types also failed to show any differences; these are shown in Table 3. Thus, there is no evidence that the coronal stimuli were any more or less different from English [t] than the dorsal stimuli were from English [k].

**TABLE 3**

Latencies by response type, Experiment 3

<table>
<thead>
<tr>
<th>Response type</th>
<th>Coronal ( M )</th>
<th>Coronal ( SD )</th>
<th>Dorsal ( M )</th>
<th>Dorsal ( SD )</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit</td>
<td>1392.42</td>
<td>482.83</td>
<td>1491.94</td>
<td>422.72</td>
<td>1.09</td>
<td>&gt; .25</td>
</tr>
<tr>
<td>Miss</td>
<td>1409.65</td>
<td>436.87</td>
<td>1479.12</td>
<td>509.87</td>
<td>1.04</td>
<td>&gt; .30</td>
</tr>
<tr>
<td>False Alarm</td>
<td>1579.92</td>
<td>610.03</td>
<td>1646.65</td>
<td>673.27</td>
<td>0.58</td>
<td>&gt; .50</td>
</tr>
<tr>
<td>Correct Rejection</td>
<td>1532.84</td>
<td>548.38</td>
<td>1510.27</td>
<td>518.03</td>
<td>0.21</td>
<td>&gt; .50</td>
</tr>
</tbody>
</table>

*Note:* Latencies are in ms. One subject had no false alarm responses on dorsal stimuli, thus on this comparison, \( df = 14 \). On all other comparisons, \( df = 15 \).

The results of this experiment partially replicate those of Werker and Logan (1985) and Polka (1991): Overall performance was very similar across their studies and ours. However, whereas Werker and Logan and Polka both found improvement across blocks on the stimuli they tested, we did not. On the view that improvement over time provides information about the treatment of a particular contrast, this could be interpreted as an important difference. For example, a significant proportion of Polka's subjects may have perceived the Salish contrast as non-speech, whereas our subjects perceived these stimuli as speech sounds. Perhaps, then, improvement occurs when subjects perceive the sounds as non-speech, but not as assimilating to a single native-language category. However, contrary to the suggestion by Best (1993), Werker and Logan's data suggest that improvement is not diagnostic of the distinction between Single Category and Nonassimilable contrasts. In their study, there was improvement for a contrast subjects described as falling within a single native-language category. Furthermore, there were methodological differences that can easily account for discrepant findings across the studies. Differences in instructions between our study and Polka's were noted earlier. In addition, Werker and Logan's subjects received 480 trials on the Hindi contrast; Polka's subjects received 180 trials...
per contrast (Salish and Farsi), and subjects tested on both contrasts were tested on each contrast on different days. In contrast, our participants received only 72 trials per contrast and were tested on both the Hindi (coronal) and Salish (dorsal) contrasts in a single session. In light of these methodological differences, variance in improvement across studies does not appear to be meaningful. There is no indication on the basis of our results that the two contrasts were perceived as differing qualitatively in their assimilability to native-language categories.

Like 6.5-month-olds, English-speaking adults discriminate the dorsal and coronal contrasts equally well (or equally poorly). Taken together, results from Experiments 2 and 3 indicate that there is no significant inherent difference in the discriminability of the two non-native contrasts we tested. Nor is there any evidence that these two non-native contrasts were perceived as falling in different categories in Best’s (1993, 1995) taxonomy. The explanation for the difference in performance on these two contrasts in Experiment 1 by 8.5-month-olds thus must be sought elsewhere.

5 General Discussion

The goal of this study was to examine whether frequency of occurrence in input plays a role in the acquisition of phonological categories. We investigated this by examining infants’ relative abilities to discriminate two non-native contrasts whose corresponding English categories differ in frequency. Dominant theories accounting for loss of discrimination of non-native contrasts in infants all tie this loss of ability to learning of native categories.

If frequency does play a role in the developing phonological system, we would expect infants to learn more frequent categories earlier than less frequent categories. This would result in earlier declines in discrimination performance on non-native contrasts relating to the more frequent categories. An examination of transcriptions of speech to children confirmed that coronal stops are more frequent than dorsal stops. In Experiment 1, 8.5-month-old infants discriminated a non-native coronal stop contrast significantly worse than they discriminated a non-native dorsal stop contrast. In Experiment 2, 6.5-month-old infants performed equally well on the two contrasts, as did adults in Experiment 3, indicating that 8.5-month-old infants’ poor performance on the coronal contrast was not due to inherent differences in discriminability of the contrasts tested. Taken together, these results are consistent with the hypothesis that frequency plays a role in determining the order in which native language phonological categories are acquired and, hence, the order in which non-native contrasts are lost.

An alternative explanation of our results might appeal to the “special” status of coronals in phonology. Across languages, coronal consonants are more likely to appear in phonological inventories than are consonants at other places of articulation (Ferguson, 1963; also see contributions to Paradis & Prunet, 1991). Furthermore, within languages, inventories of consonants tend to be richer for coronals than for alternative places of articulation (Maddieson, 1984; also see contributions to Paradis & Prunet, 1991). Thus, it is likely that the high relative frequencies of coronals observed in

Language and Speech
English reflect a widespread, if not universal, pattern. These observations have led phonologists to suggest that coronals are unmarked (see contributions to Paradis & Prunet, 1991). Hence, the contrasts that we tested may be distinguished by their markedness as well as by differences in the frequencies of corresponding English categories.

A markedness explanation of our results is less adequate than a frequency-based account for several reasons. If these two accounts are to be distinguishable, there must be some frequency-independent means by which learners can judge which phonological categories are unmarked. Learners have no access to phonological inventories of other languages, so the cross-linguistic commonness of coronals is not relevant. Learners might be innately disposed to classify coronals as unmarked. Alternatively, they might observe that there is a larger inventory of coronals in their language. Or learners might observe that coronals behave differently than do other phonological categories with respect to certain phonological processes. Some alternations apply uniquely to coronals; for example, in English word-final coronal nasals may assimilate in place of articulation to a following labial or dorsal obstruct (but not vice versa). In other instances, coronals may be more likely to result from certain phonological processes, such as neutralization. We consider each of these accounts of the origin of markedness in turn.

On the view that learners have some innate sense of markedness, the unmarked status of coronals may be consistent with the notion that such categories are learned earlier than others. However, on this view, it is not at all clear why there should be a relation between “less marked” and “less discriminable.” To the contrary, if learners are predisposed to learn multiple coronal categories (which must be distinguished by finer gradations than are the sparser sets of categories at other places of articulation), they ought to have enhanced discriminative abilities for these less marked sounds. Moreover, if there were an innate relation between “less marked” and “less discriminable” it ought to be developmentally constant: An innate markedness hypothesis cannot account for the fact that 6.5-month-olds discriminate both coronal and noncoronal contrasts, whereas 8.5-month-olds discriminate only the latter.

Changes in discriminative abilities across age might be accommodated by the view that learners develop a sense of markedness from observations of phonological processes or phonological inventories in the native language. However, on this view, the (eventual) unmarked status of coronals is irrelevant to the order in which phonological categories are acquired. This is because in order to observe that one category behaves differently than others with respect to phonological alternations, these categories must have already been acquired. Similarly, noting that inventories of consonantal categories are not uniformly distributed with regard to place of articulation requires that these categories already be learned. Regardless of whether markedness is learned by observing phonological processes or inventories, learning

---

3 Evidence of preference for unmarked categories or phonotactic sequence should appear early in development; indeed, Smolensky, Davidson and Jusczyk (in press) found that 4.5-month-old infants prefer unmarked syllable sequences with assimilated nasals (e.g., /umpo/) to marked sequences with unassimilated coronal nasals (e.g., /unpo/).
which categories are marked requires induction to be performed over already existing categories. Thus, on this view, the order of acquisition of native language categories — and hence the order of loss of non-native contrasts — must be independent of markedness. The innate markedness and the learned markedness accounts can offer at best post hoc accounts of the present results.

In contrast, the frequency-based account that we advocate offers a straightforward explanation of these results. One way to conceptualize this phenomenon is in terms of an attractor model. The DRIBBLER (Dimensionally Reduced Item-Based Lexical Recognition) model we have been developing (Morgan et al., 2001) is similar to the attractor models proposed by Hinton and Shallice (1991), Tabor, Juliano, and Tanenhaus (1997), and Tabor and Tanenhaus (1999) used for syntactic processing. DRIBBLER is an attempt to flesh out Jusczyk's (1993, 1997) WRAPSA model. Like WRAPSA, DRIBBLER posits a front-end encoder that takes as its input the output of domain-general sensory analyzers. This encoder seeks to reduce the dimensionality of representation (roughly from phonetic space to phonemic space) by extracting covariation from word-like stretches of speech. Phonological categories emerge as nexuses of position-specific variation (see also Pierrehumbert, 2001). Order of emergence depends jointly on frequency and distribution of exemplars (Maye, Werker, & Gerken, 2002).

In this model, items are projected to locations in representational space determined by the encoder. Perceptually similar items will be projected near one another in space; clusters of exemplars emergently constitute categories. The location of each item is stored for future use. Categorization is competitive: The movement of the current input item under the combined gravitational attraction of all previously stored exemplars is simulated, and the endpoint of the trajectory reached within criterial time determines the perceptual categorization of the item. In general, items will be attracted to clusters according to (1) the perceptual similarity of the new item to previous exemplars, (2) the presence of neighboring, competing clusters, and (3) the density of the clusters — clusters containing more exemplars will more strongly attract new items.

In terms of the present data, when a listener hears a foreign speech sound, the sound will gravitate toward a perceptually similar native category. For more frequent sounds, clusters will have coalesced earlier and be denser than for less frequent sounds. Thus, if an infant hears a non-native coronal sound, that exemplar will be likely to gravitate quickly to a dense native cluster. Conversely, if an infant hears a non-native dorsal or labial sound, that sound would be less likely to gravitate to the sparser native cluster within criterial time, but rather may be perceived as an exemplar of a new category.

In older speakers, classification of coronals should occur more rapidly than classification of dorsals or labials. One prediction following from this is that phoneme identification should be less affected by higher-level (e.g., lexical) information for coronals than for sounds at the other places of articulation, particularly when these sounds occur in word-initial position. Indeed, several studies of top-down influences on phoneme identification in adults have found effects for dorsals and labials, but not for coronals (Burton, Baum, & Blumstein, 1989; Newman, Sawusch, & Luce, 1997; Pitt & Samuel, 1993).
If one were to adhere to the idea that coronals are learned earlier because they are unmarked, it would be expected that coronals would always be learned first. The frequency-based account, however, does not necessarily predict this. In those cases in which coronals are not the most frequent sounds, the two hypotheses make divergent predictions. Determination of precisely what these cases are may provide a means of ascertaining the nature of infants’ phonological representation of input.

Currently, we are unsure of the fashion in which infants might compute statistics pertaining to speech sounds. Intuitively, there appear to be several types of representations that infants could be considering. These representations can vary both in the size of the unit and in the context with respect to which the frequency of the unit is tabulated. For example, infants could be considering the overall frequency of phonetic features. Alternatively, at the other end of the spectrum, infants could be sensitive to CV demisyllables with respect to their word-relative position.

For the comparison that we examined, we considered the frequency of phones. For these counts, context is irrelevant; in English, the outcome is always the same: the voiceless coronal stop /t/ is always more frequent than the voiceless dorsal stop /k/. However, if we consider the frequency of demisyllables, we find that coronal-stop-onset demisyllables are not always more frequent than dorsal-stop-onset demisyllables. For instance, coronal stop onsets followed by /t/ are more frequent than dorsal stop onsets followed by /t/ (a pattern clearer for voiced than voiceless stops). However, when the same onsets are followed by /e/, the pattern reverses (again more clearly for voiced than for voiceless stops): Coronal stop onsets followed by /e/ are less frequent than are dorsal stop onsets followed by /e/. (Recall that the coronal and dorsal stimuli used here had similar midcentral vowels; the closest English vowel category is /a/. Because coronals occur with much higher frequency before /a/ than do velars, the present results are neutral with respect to the size of representational unit that infants may have used.)

Thus, opposing predictions may be made about the relative rate of loss depending on whether frequency is computed by phone/feature or by demisyllable. By testing infants on their abilities to discriminate non-native contrasts that relate to native-language categories differing in relative frequency depending on the unit of representation assumed, we hope to learn more about the types of representations infants are using. The technique of measuring graded loss of multiple contrasts, in combination with observation of input frequencies, can offer a powerful method of assessing infants' phonological representations.

The nature of these representations, in turn, bears on whether the statistical approach we advocate is sufficient to account for the order in which native language categories are acquired, or whether it must be supplemented with additional mechanisms. Recall that loss of discrimination of non-native consonant contrasts, on which we have focused, begins at about eight months and is largely complete by the end of the first year, whereas effects of native language experience on perception of vowels begin sometime between four and six months (Polka & Werker, 1994). In our tabulation of phones in child-directed speech, however, we found a ratio of consonants to vowels of about 1.5:1. If infants are computing statistics of speech sounds in terms of phones without regard to context (i.e., with respect to phonemic categories),
frequency alone cannot account for why perceptual changes occur first for vowels. In this case, some supplemental mechanism would be required. One might posit that segments with high amplitude and long duration (such as vowels) are especially salient to learners and may thus receive early preferential processing. Behnke's (1998) MAPCAT model assumes such a filter, which is instrumental there in determining that vowel categories are acquired first. However, if infants' tabulations are restricted to counting phones as they occur in particular syllabic positions (i.e., with respect to allophonic, rather than phonemic, categories), then the ratio of consonants to vowels reverses: in the child-directed speech we examined, we found a ratio of syllable-initial consonants to syllable-medial vowels of about 1:1.5. Thus, if the representation on which frequency counts are based is suitably constrained, the statistical account can suffice to explain why changes in discrimination are seen earlier for vowels than for consonants.

The experiments in this article provide evidence suggesting that the order in which infants lose the ability to discriminate non-native contrasts may be accounted for, at least in part, by a simple frequency-based theory. The ramifications of this conclusion are twofold. By examining the frequencies of occurrence of sounds in a language, we can make predictions about the order in which native phonological categories are acquired. Moreover, examining the relative order of loss of discrimination of non-native contrasts gives us a novel way to explore the granularity of phonetic information that infants are representing and exploiting in learning their native languages.

References


