



Learning words' sounds before learning how words sound: 9-Month-olds use distinct objects as cues to categorize speech information

H. Henny Yeung*, Janet F. Werker

Department of Psychology, The University of British Columbia, 2136 West Mall, Vancouver, British Columbia, Canada V6T 1Z4

ARTICLE INFO

Article history:

Received 14 July 2008

Revised 20 April 2009

Accepted 13 August 2009

Keywords:

Infancy

Language acquisition

Phoneme

Phonetic learning

Categorization

Discrimination

Perception

ABSTRACT

One of the central themes in the study of language acquisition is the gap between the linguistic knowledge that learners demonstrate, and the apparent inadequacy of linguistic input to support induction of this knowledge. One of the first linguistic abilities in the course of development to exemplify this problem is in speech perception: specifically, learning the sound system of one's native language. Native-language sound systems are defined by meaningful contrasts among words in a language, yet infants learn these sound patterns before any significant numbers of words are acquired. Previous approaches to this learning problem have suggested that infants can learn phonetic categories from statistical analysis of auditory input, without regard to word referents. Experimental evidence presented here suggests instead that young infants can use visual cues present in word-labeling situations to categorize phonetic information. In Experiment 1, 9-month-old English-learning infants failed to discriminate two non-native phonetic categories, establishing baseline performance in a perceptual discrimination task. In Experiment 2, these infants succeeded at discrimination after watching contrasting visual cues (i.e., videos of two novel objects) paired consistently with the two non-native phonetic categories. In Experiment 3, these infants failed at discrimination after watching the same visual cues, but paired inconsistently with the two phonetic categories. At an age before which memory of word labels is demonstrated in the laboratory, 9-month-old infants use contrastive pairings between objects and sounds to influence their phonetic sensitivity. Phonetic learning may have a more functional basis than previous statistical learning mechanisms assume: infants may use cross-modal associations inherent in social contexts to learn native-language phonetic categories.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Linguists and psychologists have long noted that the perception of speech is dependent on the functional use of particular speech sounds in the native language (Best, McRoberts, & Goodell, 2001; Jakobson & Waugh, 1979; Liberman, Harris, Hoffman, & Griffith, 1957; Trubetsky, 1969). For example, the /d/ sounds in the contexts of “this doll” versus “our doll” are acoustically different, but do not signify any meaningful differences in English. These acous-

tically distinct pronunciations of “doll” nevertheless signify a meaningful contrast in Hindi (i.e., dental [ɖal] and retroflex [ɖʌl] are words for *lentils* and *branch*, respectively). It follows that this phonetic contrast is easily discriminated by adult Hindi speakers, and is difficult for adult English speakers (Stevens & Blumstein, 1975; Werker & Lalonde, 1988). In almost all cases¹ auditory perception is tuned to this native-language pattern from an initial language-general pattern within the first 6–12 months after birth (Werker & Tees, 1984). For example, both English- and Hindi-learning infants discriminate a [ɖa]–[ɖʌ] contrast

* Corresponding author. Tel.: +1 604 822 6408; fax: +1 604 822 6923.
E-mail address: hhyeung@psych.ubc.ca (H.H. Yeung).

¹ See Narayan, Werker and Beddor (in press) for an interesting exception.

at 6–8 months of age, but only the latter continue to do so by 10–12 months (Werker & Tees, 1984).

These findings are striking because native-language sound systems, which are defined by meaningful contrasts among words, are learned before infants otherwise understand or remember any significant numbers of words. The development of this native-language sound system exemplifies the central question in the field of language acquisition: how does a learner acquire linguistic knowledge when linguistic input seems unable to support induction of this knowledge? On the one hand, infants almost certainly have some universal learning biases which help auditory systems draw phonetic boundaries in acoustic space. For example, psychophysical factors may constrain the possible sets of phonetic categories across languages (Kuhl & Miller, 1975; Pisoni, 1979; Werker & Lalonde, 1988). On the other hand, these universal biases cannot explain how infants' sensitivities are attuned to the *specific* phonetic patterns present in their native language. This process has generally been described as a *functional reorganization* of perceptual sensitivity (Werker, 1995) since something about the meaningful or functional status of a phonetic contrast in an infants' native language is claimed to drive discrimination patterns (Best, 1993; MacKain, 1982; Werker & Tees, 1999). It seems unlikely that infants learn language-specific phonetic contrasts by comparing minimally different words from their native language, however, since young infants have receptive lexicons far too small to compute phonetic inventories.

An alternative possibility is that infants learn phonetic boundaries from statistical analyses of speech input, which is generated in turn by the functional use of language by adults (Jusczyk, Bertoni, Bijeljac-Babic, Kennedy, & Mehler, 1990; Kuhl, 1993; Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002; Vallabha, McClelland, Pons, Werker, & Amano, 2007; Werker et al., 2007). Maye and colleagues experimentally tested this hypothesis with infants aged 6–8 months (Maye et al., 2002, 2008). In one study, for example, infants heard sounds drawn from a continuum, along which a single phonetic parameter (i.e., voice-onset time) varied. This continuum crossed the phonetic category boundary between [d] and [t]; similar to two phonetic tokens used contrastively in Spanish, but not in English (Maye et al., 2002). One group of infants heard continuum tokens presented in a frequency-distribution that modeled a language with the phonetic contrast. Tokens near the ends of the continuum were more frequent, generating a bimodal distribution. Infants in this group were better at discriminating the contrast in a subsequent test phase than those who were exposed to another frequency-distribution that modeled a language without the contrast. In this case, tokens from the middle of the continuum were heard more frequently, generating a unimodal distribution.

These experimental situations greatly simplified auditory input, permitting category-induction by a simple statistical analysis (i.e., detecting frequency peaks along a continuum varying on a single acoustic dimension), but more recent work suggests that category boundaries are also learnable from more naturalistic input. Acoustic recordings taken while mothers interacted with their in-

fants during word-teaching sessions were analyzed, and induction of phonetic category boundaries was possible with more advanced statistical techniques (Werker et al., 2007). This analysis assumes, nevertheless, that the learner has the aggregate dataset and knows *a priori* the correct number of phonetic categories. However, subsequent computational modeling of this same dataset yielded similar results, even when the number of phonetic categories was not fixed and when speech tokens were entered iteratively into this model, as a learner might encounter them (Vallabha et al., 2007).

A purely statistical approach may still have its limitations in the domain of developmental speech perception. Firstly, it is not known whether all the computational analyses mentioned above are available to infants. Neither is it known whether these analyses would be successful given input from less circumscribed situations (i.e., not limited to explicit teaching contexts), or situations where the input is not balanced for absolute frequency (i.e., where one phonetic category is much more frequent than its acoustic neighbors, as is the case with many phonetic contrasts; e.g., Dillon, Idsardi, & Phillips, 2008). A more robust method of categorization might necessarily rely on *functional* relations between real-world referents and phonetic forms, similar in spirit to earlier conceptions of phonology as a causal factor in driving perceptual patterns (Best, 1993; MacKain, 1982; Trubetskoy, 1969; Werker, 1995; Werker & Tees, 1999). Secondly, it is not known whether the learning mechanism of choice may change through development. For example, while Maye and colleagues report learning from statistical distributions at 6–8 months of age, learning of this type may be subject to stricter attentional demands by 10 months of age (Yoshida, Pons, & Werker, submitted for publication). Indeed, other research at this older age suggests that exposure to a non-native language affects infants' phonetic sensitivity to a greater degree when it comes from contingent social interaction with an experimenter, compared to when input comes from non-contingent audio or video recordings of the same experimenter's play session with another infant (Kuhl, Tsao, & Liu, 2003). These data together suggest that by 10 months of age, infants' phonetic learning mechanisms are heavily influenced the *functional* role of speech in social contexts, which may in turn be mediated by attentional factors.

As mentioned earlier, functional accounts are weakened by the fact that infants do not have vocabularies of any substantial size by the age at which phonetic sensitivity becomes language-specific. However, testing infants' comprehension of words may underestimate the amount of information present in linguistic input from which infants can learn. For example, research in cognitive development supports the idea that infants from 6 to 12 months of age link speech sounds and conceptual knowledge before they readily learn labels for objects (Waxman, 2002; Xu, 2007). Waxman and colleagues report that 6-, 9-, and 12-month-old infants can group different objects into a single visual category when those objects are paired with a consistent phonetic form (e.g., "daxy" for every object) and not when paired with a tone, emotional expressions, or inconsistent word-forms (e.g., "daxy" for one object, and "blicket" for

another) (Balaban & Waxman, 1997; Fulkerson & Waxman, 2007; Waxman & Braun, 2005; Waxman & Markow, 1995). Other researchers report that pairing phonetic forms with objects can influence young infants' abilities to track and individuate these objects (Xu, 2002), as well as modify infants' expectations about their internal properties (Graham & Kilbreath, 2007) and kind-membership (Dewar & Xu, 2007). This work suggests that young infants use the co-occurrence of a unique phonetic form with different objects to help structure and enrich their conceptual categories.

While these studies show that *verbal* cues can shape conceptual categories, the converse has been left relatively unexplored: whether infants use *visual* cues to structure their phonetic categories. A recent study from Teinonen, Aslin, Alku, and Csibra (2008) suggests that infants can use audiovisual matching between visual displays of lip and tongue movement on the one hand, and speech sounds on the other, to learn phonetic categories. This occurs even when statistical information is uninformative. While this provides important evidence that infants can use perceptual cues in vision to categorize in the auditory domain, the crucial link to conceptual development remains untested: whether infants can use more visually abstract cues (i.e., possible referents) to categorize phonetic information.

There is supporting literature from animal and adult learning studies, suggesting that functional cues to category identity may help in learning phonetic contrasts. In one study, for example, birds given feedback about phonetic category membership were able to learn both consonant (Kluender, Diehl, & Kileen, 1987) and vowel (Kluender, Lotto, Holt, & Bloedel, 1998) categories after extensive training. In another study (human) adults heard ambiguous synthesized phonetic tokens that had an acoustic cue to one phonetic category (e.g., a /ba/-like formant transition), but another acoustic cue to a contrasting phonetic category (e.g., a /da/-like burst release). Adults were given explicit feedback about these ambiguous tokens in a training phase that biased them to categorize using either one acoustic cue or the other, and they generalized this classification pattern to novel speech tokens in a subsequent test phase (Francis, Baldwin, & Nusbaum, 2000). Importantly, adults can learn not only from explicit training, but can also learn from mere exposure to category cues. Adults who were merely exposed to two distinct pictures (i.e., a rat vs. a pot) paired contrastively with two phonetic categories discriminated these phonetic categories in a subsequent test phase better than adults who had seen only one picture paired with both phonetic categories (Hayes-Harb, 2007).

These kinds of studies are evocative of learning mechanisms described in the perceptual learning literature (Hall, 1991; Miller & Dollard, 1941). The phenomenon of *acquired distinctiveness* describes enhanced differentiation of two target stimuli resulting from a previous pairing of distinct events or responses with the targets, and *acquired equivalence* describes its converse, where the same event or response is paired with two target stimuli and impairs subjects' subsequent discrimination and differentiation (Hall, 1991; Miller & Dollard, 1941). Experimental evidence for these phenomena was first reported in rats (Lawrence, 1949), but has been replicated in adult humans (Hall,

1991), and young children (Norcross & Spiker, 1957; Reese, 1972).

The current experiments present evidence for a phonetic learning mechanism similar in principle to, but importantly different from, acquired distinctiveness. In the perceptual learning literature, acquired distinctiveness often implies a learned, associative link between differentiating events and target stimuli: for example, associating two distinct responses to two targets (Hall, 1991). As suggested by the research from cognitive development reviewed above, however, the mere co-occurrence of target stimuli and distinct visual objects may provide sufficient cues for categorization without necessarily implying a long-lasting, learned association. This kind of learning mechanism would help explain how infants learn language-specific phonetic contrasts without first learning contrastive words. In Hindi contexts, for example, the pairing of one visual cue (e.g., lentils) and its label (e.g., [ɖal]) against a contrasting, distinct visual cue (e.g., a branch) and a similar label (e.g., [ɖal]) may decrease the perceived similarity of those two phonetic forms. In an English context, the pairing of a visual cue (e.g., a doll) and both labels (e.g., [ɖal] in a phrase like "this doll is pretty" and [ɖal] in a phrase like "your doll is pretty") may increase perceived similarity between these phonetic forms. An experimental version of the Hindi context presented to English-learning infants is reported in this series of experiments, and asks whether the pairing of speech with distinct visual cues can be informative for learning phonetic categories, all without requiring that infants learn minimal pairs of words.

2. Experiment 1: baseline

English-learning infants' perceptual sensitivity for the aforementioned Hindi phonetic contrast (i.e., a dental alveolar stop [ɖa] vs. a retroflex alveolar stop [ɖɑ]) is well studied; discrimination performance declines from 6–8 to 10–12 months of age (Werker & Tees, 1984). Infants at 9 months of age are in the midst of perceptual reorganization, and this age group was tested in the hope that perception of this particular contrast might show some sensitivity to our experimental manipulations in Experiments 2 and 3.

A standard habituation paradigm is not feasible for testing discrimination in this series of experiments, because in subsequent experiments (i.e., Experiments 2 and 3) both categories of Hindi sounds are presented during a familiarization phase presented prior to the test phase assessing phonetic sensitivity. Instead, an alternating/non-alternating paradigm was used to test phonetic discrimination (Best & Jones, 1998; Mattock, Molnar, Polka, & Burnham, 2008; Maye et al., 2002; Teinonen et al., 2008; Yoshida et al., submitted for publication). This experiment establishes English-learning 9-month-olds' baseline sensitivity for the Hindi dental-retroflex contrast using this particular discrimination task.

2.1. Methods

Participants. Twenty 9-month-old infants (11 female; mean age = 9;5, range = 8;20–9;20) were recruited from a

database of families who had previously expressed interest in research studies. Care-givers were informed about the study prior to their participation, and their infants were given a *t*-shirt and a certificate as a token of thanks. As measured by parental report, infants were exposed to at least 80% English, and less than 1% of any South Asian language, which commonly have dental-retroflex distinctions. Data from an additional 2 infants were not included due to experimenter error (1 male; 1 female).

2.1.1. Stimuli

Naturally produced infant-directed stimuli were elicited from a native Hindi-speaking female (age = 35 years). Stimuli were recorded in a sound-attenuated booth on a Radio Shack unidirectional dynamic microphone (model 33–3009) connected to a preamp set at maximum gain. Tokens consisting of CV-syllables (i.e., consonant–vowel syllables), contrasting on a Hindi phonetic contrast (the voiced dental alveolar stop [ɖa] versus the voiced retroflex stop [ɖʌ]), were elicited in sentential frames and then excised from these frames. Six dental and six retroflex tokens that had similar rising pitch contours were selected. Another native speaker of Hindi independently classified the category of each excised token with 100% accuracy. Two tokens of each kind were used in Experiment 1 and in the test phases of Experiments 2 and 3, while the remaining four were used in the familiarization phases of Experiments 2 and 3. The average length of the test trial tokens used here was 503 ms (dental = 506 ms; retroflex = 500 ms).

2.1.2. Procedure

Infants were tested in a quiet, softly lit room while sitting on their care-giver's lap. Care-givers were instructed not to speak, and to keep their infants calm while themselves listening to masking music over headphones; they were seated 36 in. away from a black curtain. A 42 in. plasma screen was positioned in the middle of the curtain, and a slit for a video camera (Sony Digital Handycam, model DRC-TRV25) was positioned 22 in. above the floor, and 6 in. under the bottom edge of the screen. An experimenter in another room could see the infant's face through the camera's video display, and controlled stimulus presentation with computer software (Habit X: Leslie Cohen at the University of Texas, Austin). Sounds were presented free-field over speakers hidden behind the curtain at approximately 60–62 dB.

The study began with a single 12-s warm-up trial showing a moving toy on the screen paired with a nonsense word. Then, a non-object stimulus (i.e., a static unbounded black-and-white checkerboard pattern) was displayed on each trial as infant looking-time was recorded. Each checkerboard trial lasted 10 s, and the first checkerboard was silently presented to give infants a chance to look at the novel visual stimulus (i.e., the 'silent checkerboard' trial). Following the first silent checkerboard trial, four pairs of two types of test trials containing auditory stimuli were presented. In one trial type, two unique tokens from the same phonetic category were presented (i.e., non-alternating trials), while in the other, two unique tokens from the contrasting category were presented (i.e., alternating tri-

als). Looking-time to non-alternating and alternating trial types was compared, and infants were assumed to discriminate the phonetic contrast if they look longer at one type of trial over the other (Best & Jones, 1998).

As described above, four novel auditory tokens (two dental; two retroflex) were used in testing. The non-alternating test trials contained either the two dental tokens, or the two retroflex tokens. Alternating test trials contained two tokens, one from each of the two phonetic categories. All four pairings of the two dental and two retroflex tokens were presented twice in eight test trials. In a single trial a token was played every 1500 ms, and alternated between the two tokens for a total of 10 s. When a trial finished, the same colorful pattern used to attract infants' attention in the familiarization phase was displayed if infants were no longer looking at the screen; otherwise, the next trial began immediately. For half the infants the 1st, 3rd, 5th, and 7th trials were of the alternating type, and the 2nd, 4th, 6th, and 8th trials were of the non-alternating type. For the other half of infants, the order of trial types was reversed.

Videos of the test trials were digitized from DAT recordings and then converted to QuickTime movies. Looking-time to test trials was coded frame-by-frame at a rate of 29.97 frames per second by a trained coder. Looking time to the first pair of test trials (i.e., the 1st and 2nd test trials) was analyzed as Pair 1. For example, for non-alternating trials in Pair 1, half the infants contributed means from their 1st test trial, and half the infants contributed means from their 2nd test trial, depending on which of the counter-balanced orders was presented. A similar analysis was done for the 3rd and 4th test trial (Pair 2), the 5th and 6th test trial (Pair 3), and the 7th and 8th test trial (Pair 4).

2.2. Results and discussion

A 2×4 (Type \times Pair) repeated-measures ANOVA yielded no significant interaction ($F(3, 57) = 0.61$; $p > 0.05$), nor was there a main effect of Type ($F(1, 19) = 0.30$; $p > 0.05$), indicating that there was no difference in looking-time between non-alternating (7.14 s) and alternating (7.01 s) test trials. However, there was a main effect of Pair ($F(3, 57) = 10.44$; $p < 0.001$), indicating that infants habituated to these trials over the duration of the test phase. A significant linear ($F(1, 19) = 25.96$, $p < 0.001$) component was observed in this looking decline. Looking time is charted in Fig. 1.

In summary, infants were habituating to the stimuli presented in Experiment 1, and did not treat non-alternating and alternating trial types differently. This suggests that at 9-months, English-learning infants do not succeed at discriminating a Hindi dental and retroflex phonetic contrast in an alternating/non-alternating discrimination task.

3. Experiment 2: familiarization with consistent visual cues

Nine-month-old English-learning infants in Experiment 1 failed to discriminate the contrast between Hindi dental and retroflex stop consonants using a non-alternating/

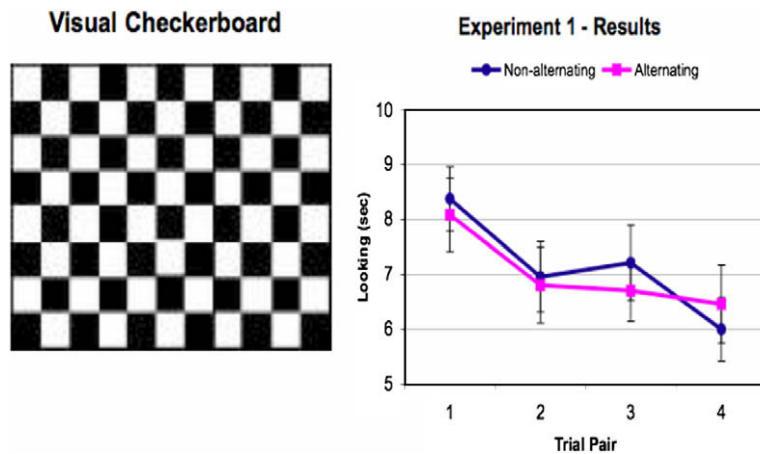


Fig. 1. In a procedure designed to measure phonetic sensitivity, infants saw a static visual checkerboard (left) and heard trials where speech tokens were played from either one phonetic category (i.e., non-alternating trials) or two phonetic categories (i.e., alternating trials). Infants did not look consistently longer at either non-alternating trials (dark blue line) or alternating trials (light pink line). Error bars indicate standard errors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

alternating procedure to test phonetic sensitivity. While in the midst of perceptual reorganization, 9-month-old infants may still be sensitive enough to benefit from the brief training periods commonly used in behavioral studies on infants. In Experiment 2, English-learning infants this age were familiarized with visual cues that promoted phonetic categorization: each one of two distinct objects was consistently paired with a unique Hindi phonetic category.

In a familiarization phase, infants were presented with eight unique syllables from two Hindi categories. Speech tokens were presented along with videos of two distinct objects, where the pairing between a particular object and a particular phonetic category was *consistent*: the object was a reliable cue for phonetic category membership. After exposure to the object–sound familiarization, the test phase used in Experiment 1 was administered to assess infants' ability to discriminate the phonetic contrast.

3.1. Methods

3.1.1. Participants

Twenty 9-month-old infants (10 female; mean age = 9;4, range = 8;13–9;19) were recruited from the same database used in Experiment 1. Care-givers were similarly informed, and the same language criteria were used to exclude infants from the analysis. Data from an additional 7 infants were not included due to failure to meet the language criteria (1 female), fussiness (3 males; 1 female), and experimenter error (2 males).

3.1.2. Stimuli

Four tokens from each Hindi phonetic category were used in the familiarization phase, and these tokens were recorded in the same session from the same speaker as those described in Experiment 1. The two additional tokens from each category used in Experiment 1 were again used in the test phase here. Average length of the familiar-

ization tokens was 527 ms (dental = 523 ms; retroflex = 531 ms).

3.1.3. Procedure

The testing apparatus and instructions to the care-givers were identical to Experiment 1, with the addition of a familiarization phase before the test phase. The study began with the warm-up trial as in Experiment 1, and then the familiarization phase began as one of two different objects appeared on the left side of the screen. It paused for 1000 ms, rotated as it moved to the right side of the screen (250 ms), paused for another 1000 ms, subsequently rotated as it moved back (250 ms), and paused for another 1000 ms before disappearing. If infants were still looking at the video display, this sequence was repeated, but if not, then a colorful pattern appeared on the screen until a look was made back to the screen. If infants did not look for more than 2 s during this sequence, it was repeated.

During each of these sequences, two of the four tokens from a single Hindi category (either dental or retroflex) were presented along with the onset of object-movement. The audio and visual stimuli were synchronized to increase the likelihood that infants would encode the sound–object pairing (Gogate & Bahrick, 2001). In four movement-sequences displaying a single object, eight auditory tokens (four unique) were presented to infants. This was followed by four movement-sequences with a new object presented in a similar fashion along with the CV-exemplars from the other Hindi category (see Fig. 2). Object–sound pairings alternated in this way until infants accumulated a total of 2.5 min of looking. The length of the familiarization period lasted ~4–5 min, including delays introduced by the computer and the time it took for infants to re-fixate on the screen after looking away. The pairings between dental and retroflex sounds and the two novel objects were counter-balanced across infants, as well as the order of dental and retroflex blocks. As soon as infants had accumulated enough looking-time, the test phase began. The test phase

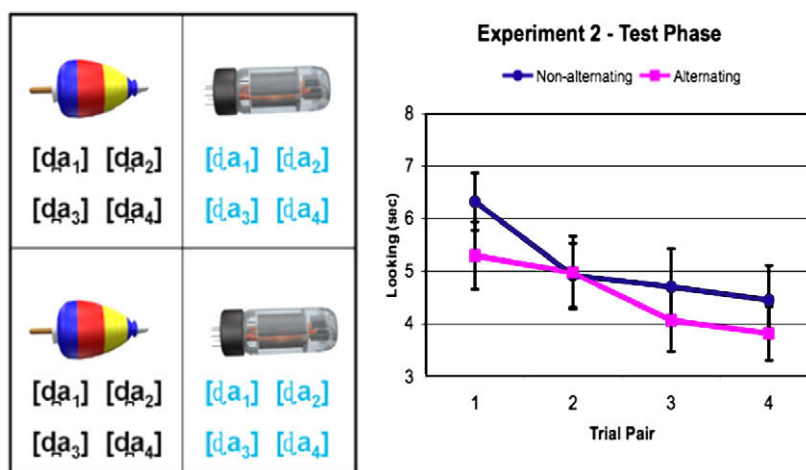


Fig. 2. Infants were familiarized with sound–object pairings before testing phonetic sensitivity. Four unique tokens (subscripts 1–4) from either dental [ɖa] or retroflex [ɖa] categories were presented multiple times with one object in each familiarization trial. Familiarization trials had a consistent pairing between objects and phonetic categories. In the test phase, infants looked consistently longer during non-alternating trials (dark blue line) than during alternating trials (light pink line). Error bars indicate standard errors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

was identical to Experiment 1, as was the analysis of the results.

3.2. Results and discussion

Infants' looking times to the eight test trials were entered into a repeated-measures ANOVA with factors of Type (alternating or non-alternating) \times Pair (1, 2, 3, or 4). The 2×4 (Type \times Pair) ANOVA yielded no significant interaction ($F(3, 57) = 0.61$; $p > 0.05$), but there was a main effect of Type ($F(1, 19) = 4.63$; $p < 0.05$), indicating that infants looked on average longer at the non-alternating (5.10 s) compared to the alternating (4.54 s) trial type. There was also a main effect of Pair ($F(3, 57) = 4.94$; $p < 0.01$), and a significant linear contrast on this main effect ($F(1, 19) = 9.77$, $p < 0.01$), indicating that infants looked progressively less at each pair of test trials. Looking time is charted in Fig. 2.

These results suggest that infants were habituating to all the trials in the test phase, but looked longer overall at only one of the trial types. Because the two trial types differed only in whether phonetic tokens comprised a within- or between-category contrast, infants' preference to look at one over the other suggests discrimination of the categories. Moreover, a looking-preference was observed for the non-alternating trial type. This replicates previous work showing a novelty preference for non-alternating trials following a familiarization phase² (Maye et al.,

2002; Yoshida et al., submitted for publication), and suggests that these infants were able to discriminate the non-native phonetic contrast after only a brief period of training.

It remains unclear, however, whether infants depended on the consistent pairing between objects and speech sounds to form phonetic categories. Firstly, infants may have simply relied on statistical cues also present in the familiarization phase. Because infants heard naturally produced exemplars from two phonetic categories, they were essentially exposed to two distinct clusters of sounds in phonetic space, a more leptokurtic version (i.e., with “sharper” peaks) of the bimodal frequency-distributions than used in previous studies modeling statistical learning from auditory input (Maye et al., 2002, 2008; Yoshida et al., submitted for publication). If infants did not pay attention to the visual stimuli at all, then computing the statistical properties of the auditory input alone may still have provided information needed to learn the phonetic categories. Secondly, previous research on perceptual learning in several domains suggests that mere pre-exposure to target stimuli during a training period improves subsequent discrimination (Gibson, 1969; Gibson & Walk, 1956). To guard against these possibilities, a second study was run.

4. Experiment 3: familiarization with inconsistent visual cues

In Experiment 3, infants were familiarized with the same audio input as in Experiment 2, but visual cues were inconsistent and thus uninformative: both distinct objects were paired with both Hindi categories. If infants were not attending to the pairing between speech sounds and visual stimuli, but depending on the statistical distribution of sounds in the familiarization phase to learn the contrast, or benefiting from simple pre-exposure to the auditory tokens, then presenting infants with the same auditory input (but eliminating the correlation between the types of

² A particular direction of preference (i.e., either alternating $>$ non-alternating or vice versa) is not necessary to interpret discrimination of the phonetic contrast. While both looking patterns have been used to suggest discrimination, the present pattern is found in studies where phonetic discrimination follows a lengthy (i.e., 2 minutes or more) familiarization phase (Maye et al., 2002; Teinonen et al., 2008; Yoshida et al., submitted for publication), whereas the opposite pattern has been found when the familiarization phase is brief (i.e., 30 s or less) or absent (Best & Jones, 1998; Mattock et al., 2008). In all cases, equal looking to both trial types is interpreted as a failure to discriminate.

sounds and objects) should not affect infants' ability to discriminate. However, if the object–sound pairings were influencing perceptual sensitivity, then presenting infants with *inconsistent* pairings in the familiarization phase should affect infants' successful discrimination.

4.1. Methods

4.1.1. Participants

Twenty 9-month-old infants (10 female; mean age = 9;4, range = 8;14–9;22) were recruited from the same database. Care-givers were similarly informed as in Experiments 1 and 2, and the same language criteria were used to exclude infants from the analysis. Data from an additional 5 infants were not included due to fussiness (1 male; 2 females), experimenter error (1 female), and being out of range of the video camera (1 female).

4.1.2. Stimuli and procedure

The stimuli, testing apparatus, and instructions to the care-givers were identical to Experiment 2. Familiarization trials were also of the same structure as those used in Experiment 2: eight auditory tokens (four unique) were presented in four movement-sequences displaying a single object. This was followed by four movement-sequences with the other Hindi category. Over the course of this familiarization phase, however, both objects were shown with each block of phonetic tokens. Thus, infants saw the same two distinct-looking objects, but object–sound pairings were *inconsistent*: objects were paired with exemplars from both phonetic categories. In Experiment 2, for example, infants saw the “top” object and the “tube” object in a regular order (e.g., top–tube–top–tube–top–tube), while infants in Experiment 3 saw an irregular order (e.g., top–tube–tube–top–tube–top–top–tube). Importantly, infants received the same overall distribution of sounds. As in Experiment 2, the pairings between dental and retroflex sounds and the two novel objects were also

counter-balanced across infants, as well as the order of dental and retroflex blocks. The test phase was identical to Experiment 1 and 2, as was the analysis of the results.

As confirmation that this new sequence of pairings did not affect how interested infants were in the familiarization stimuli, a test was conducted on the number of trials needed to reach the looking-time criterion in the infant-controlled familiarization procedure. Since each infant accumulated an equal amount of looking in the familiarization phase (see Methods section), a difference in the number of familiarization trials it took to reach this criterion would indicate whether these infants were more or less interested in the stimuli. The number of trials it took to reach criterion each study did not differ significantly (mean = 56.3 trials in Experiment 2 *versus* mean = 56.8 trials in Experiment 3; $t(38) = 0.49$; $p > 0.05$).

4.2. Results and discussion

Results from the test phase suggested that these infants looked equally at the two types of trials (Fig. 3). A 2×4 (Type \times Pair) repeated-measures ANOVA yielded no significant interaction ($F(3, 57) = 0.59$; $p > 0.05$), nor was there a main effect of Type ($F(1, 19) = 0.43$; $p > 0.05$), indicating that there was no difference in looking-time between non-alternating (5.13 s) and alternating (4.93 s) test trials. However, there was a main effect of Pair ($F(3, 57) = 11.56$; $p < 0.001$), indicating that infants habituated to these trials over the duration of the test phase. Significant linear ($F(1, 19) = 18.39$, $p < 0.001$) and quadratic ($F(1, 19) = 5.84$, $p < 0.05$) components were observed in this looking decline. In summary, infants were habituating to the stimuli in the test phase, but did not treat non-alternating and alternating trial types differently. Importantly, the distribution and number of auditory tokens to which infants were familiarized was identical to Experiment 2, yet there was no evidence that this group of infants was able to discriminate the Hindi phonetic contrast.

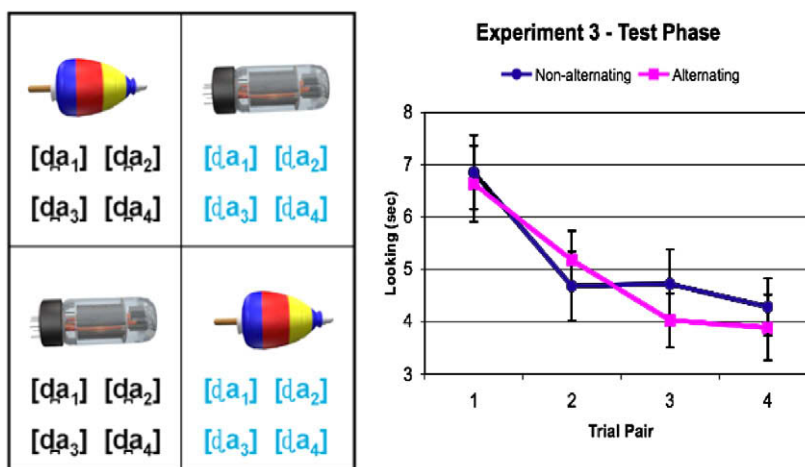


Fig. 3. Infants were familiarized with sound–object pairings before testing phonetic sensitivity. Four unique tokens (subscripts 1–4) from either dental [d̪a] or retroflex [ɖa] categories were presented multiple times with one object in each familiarization trial. Familiarization trials had an inconsistent pairing between objects and phonetic categories. In the test phase, infants did not look consistently longer during non-alternating trials (dark blue line) than during alternating trials (light pink line). Error bars indicate standard errors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. General discussion

The central learning problem in language acquisition is exemplified in the field of developmental speech perception. Phonetic sensitivity is re-organized along acoustic dimensions useful for contrasting words in the native language, but this happens *before* infants learn a lexicon with a sufficiently rich phonemic inventory to compute sound contrasts from comparing minimally different words. Researchers have previously suggested that the *statistical* information inherent in the auditory input to young infants provides information important for establishing phonetic categories (Jusczyk et al., 1990; Kuhl, 1993; Maye et al., 2002; Vallabha et al., 2007; Werker et al., 2007). Yet purely statistical analyses that focus only on auditory input may encounter problems when larger corpora and different phonetic contrasts are analyzed.

These current studies suggest a radical alternative to the dominant, yet problematic statistical approach to induction in this domain. In Experiment 1, 9-month-old English-learning infants did not differentiate a non-native Hindi dental-retroflex consonant contrast in a discrimination task. In Experiment 2, these infants were able to discriminate the non-native phonetic contrast after a brief learning period, where distinct-looking objects were paired with tokens from each phonetic category. In Experiment 3, infants did not show discrimination when these pairings were no longer informative about category identity. An essentially *functional* mechanism of phonetic reorganization is described, where the pairing of speech with an explicitly categorical cue (e.g., with distinct objects) guides perceptual learning.

The results from Experiment 2, in particular, are reminiscent of *acquired distinctiveness*, a term from the perceptual learning literature that describes differentiation of two target stimuli resulting from the pairing of distinct events with those targets (Hall, 1991; Miller & Dollard, 1941). These data do not provide explicit evidence for *acquired equivalence*, the converse phenomenon, and future work will need to test infants on a phonetic contrast in which baseline phonetic sensitivity differentiates between phonetic categories, and experience in a linguistic environment collapses this distinction. Regardless, these data provide an explicit example of a functional learning mechanism in infant speech perception similar to these mechanisms discussed in the perceptual learning literature, extending existing work on adult- and animal-learning of phonetic categories (Francis et al., 2000; Hayes-Harb, 2007; Kluender et al., 1987, 1998).

Why infants did not use information from statistical distributions (nevertheless present in the familiarization phase) to learn the phonetic contrast in Experiment 3, is an open question. One possibility is that visual cues were sometimes in conflict with statistical cues: in other words, the statistical distribution of acoustic information suggested two categories, but visual cues did not consistently support this categorization. It is possible that infants rely on statistical information only when visual cues are not conflicting, or altogether unavailable. In previous adult and animal studies of phonetic category learning the effects of *both* functional and statistical mechanisms have

been observed. For example, birds in phonetic training studies simultaneously show two pecking patterns. One pattern was a result of the *functional* nature of the training procedure: birds pecked more for test tokens with extreme values on an acoustic dimension, since high values on that dimension were reinforced. Another (different) pattern resulted from statistical information in the learning phase: birds also pecked more for test tokens that were closer to the training tokens' centroid in acoustic space (Kluender et al., 1998; Lotto, 2000). Furthermore, Hayes-Harb (2007) found that adults showed improved learning when statistical information from frequency-distributions correlated with category labeling cues, over situations when these cues were in conflict. Future work will be needed to show how both functional and statistical cues interact in infant perception.

Yet another possibility is that functional (i.e., visual) cues are more powerful than statistical cues at this stage of development. The claim that this pattern of results is related to developmental change is consistent with previous literature. Statistical learning was originally tested at 6–8 months, an age at which sensitivity to non-native contrasts is still present (Maye et al., 2002, 2008), but using statistical cues seems to become more difficult at older ages (i.e., 10 months) after which native contrasts have been established (Yoshida et al., submitted for publication). Furthermore, Kuhl et al. (2003) suggest that, by at least 10 months, language exposure affects infants' phonetic sensitivity in social, interactive contexts much more than when exposure occurs in non-social, non-contingent audio or video recordings. At 9 months, the age tested here, we may be documenting a transition in the relative strength of each learning mechanism.

This point raises the important question of what phonetic learning might look like in earlier periods of development. For example, the influence of native-language exposure is seen for vowels at a younger age than for consonants: at 6 as opposed to 10–12 months (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994). It remains an empirical question whether the mechanism for phonetic learning described here is available at this age, and how statistical mechanisms might be weighted in comparison. Other work suggests, at the very least, that labeling can affect the categorization of objects at 6 months (Fulkerson & Waxman, 2007). It seems plausible that these labelings may have a similar effect on phonetic categories. Alternatively, it may be the case that only statistical mechanisms are available at young ages, or weighted more heavily.

Clarifying what learning mechanism(s) are available throughout development requires an examination of the kinds of cues available in the regular course of care-givers' interactions with infants. Observational studies have confirmed that concordant, contingent, and synchronous word–object pairings are common in mother–child interactions, particularly in ostensive labeling situations; moreover, the amount of contingent word–object pairings in these interactions is correlated with vocabulary size at older ages (Gogate, Bahrick, & Watson, 2000; Gogate, Bolzani, & Betancourt, 2006). These ostensive contexts may also be beneficial for learning native-language sound systems as

well, and we suggest that the act of labeling an object may provide the infant with explicit cues needed for phonetic categorization. The naturalistic pairing of words and objects in infant-directed interactions, coupled with a demonstration that infants can use such cues, helps explain how the classical learning problem is simplified: if explicitly categorical information is available to infants, then an inductive strategy for deriving phonetic categories seems possible, one which overcomes some of the potentially unreliable characteristics of statistical learning.

Somewhat paradoxically, these results offer no support for the idea that infants are *remembering* whole word-forms accurately. Indeed, there is quite a bit of evidence that infants this age have trouble recalling phonetic detail in certain types of familiar words. While 10- and 11-month-old infants discriminate mispronunciations of many familiar words from their correctly pronounced counterparts (e.g., *dinner* vs. *ninner* or *cup* vs. *tup*), they ignore more subtle mispronunciations in unstressed, or word-final positions (e.g., French: *bonjour* vs. *ponjour*; Dutch: *paart* vs. *paarp*; English: *dinner* vs. *didder*) (Hallé & Boysson-Bardies, 1996; Jusczyk & Aslin, 1995; Swingley, 2005; Vihman, Nakai, DePaolis, & Hallé, 2004). Moreover, studies testing memory for newly learned word–object pairings have shown that young infants easily confuse word-forms differing by a single consonant (i.e., *bin* versus *pin*) and their referents (Stager & Werker, 1997; Thiessen, 2007). Only in cases where the task demands are made easier (Fennell & Werker, 2003; Swingley & Aslin, 2002; Thiessen, 2007; Yoshida, Fennell, Swingley, & Werker, 2009) or in later stages of infancy – around 17 months of age – can infants remember object labels which are minimally contrastive (Werker, Fennell, Corcoran, & Stager, 2002). These data serve as a potent reminder that learning words, as such, does not drive tuning of perceptual sensitivity.

If not word-learning, then a kind of learning mechanism similar to the phenomenon of acquired distinctiveness may drive language-specific perceptual tuning. However, it is important to note that this research is agnostic as to whether the kind of learning reported here is a domain-general process, or if links between speech sounds and potential referents are privileged. For example, research in cognitive development shows that hearing phonetic forms, and not tones, paired with objects may help infants define categories of objects' shapes and properties (Balaban & Waxman, 1997; Waxman, 2002). By 9–10 month of age, contrasting verbal cues, and not contrasting tones or emotional expressions, allow infants to individuate, track, and distinguish between kinds of objects (Dewar & Xu, 2007; Xu, 2007). Further research is needed to show whether cues for phonetic categorization are more effective when they are ostensible referents (e.g., distinct objects like those used here), or whether any salient cue will do.

Since speech sounds can influence young infants' categorization, tracking, and representation of objects before they have a sizable lexicon, it seems reasonable to expect that young infants can also use cues from their conceptual worlds to shape speech perception. Moreover, the influence of speech-on-concepts and concepts-on-speech may persist without requiring that infants learn the referential

functions of words. This idea revives, at least in spirit earlier notion of functional reorganization (Best, 1993; Mackain, 1982; Trubetskoy, 1969; Werker, 1995). In other words, young infants may be learning phonetic sensitivity from adults' use of particular speech sounds as functional elements of language. Although 9-month-old infants are not old enough to remember phonetic detail in words, or fully understand their referential properties, infants may still benefit from hearing words used in an interactive context, where the pairing of speech sounds with visual referents can act as a cue for perceptual learning.

Acknowledgements

We thank Dr. Laurel Fais, Marisa Cruickshank, Jasmine Cady, Emily Chevrier, Sarah Heller, Ramesh Thiruvengadaswamy, Jason Lau, Zarrin Hosseinzadeh, and Nisha Malhotra for their technical and infant recruitment assistance on this project, and we are very grateful to the parents and infants who made this research possible. Thanks also to anonymous reviewers and to Colin Phillips for their comments on an earlier version of this article. Portions of this work formed part of H.H.Y.'s MA thesis, supervised by J.F.W. This work was funded by a National Science Foundation GRF (to H.H.Y.), and a Canadian Natural Sciences and Engineering Research Council Grant (to J.F.W.).

References

- Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, 64(1), 3–26.
- Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies & S. de Schonen (Eds.). *Developmental neurocognition: Speech and face processing in the first year of life*. NATO ASI series D: Behavioural and social sciences (vol. 69, pp. 289–304). Norwell, MA: Kluwer Academic Publishers.
- Best, C. T., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development*, 21, 295.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). 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.
- Dewar, K., & Xu, F. (2007). Do 9-month-old infants expect distinct words to refer to kinds? *Developmental Psychology*, 43(5), 1227–1238.
- Dillon, B., Idsardi, W., & Phillips, C. (2008). Learning vowel categories: Challenges from Turkish and Inuktitut. In *Poster session presented at the annual meeting of the cognitive science society*. Washington, DC.
- Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, 46(2), 245–264.
- Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception and Psychophysics*, 62(8), 1668–1680.
- Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition*, 105(1), 218–228.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. East Norwalk, CT: Appleton-Century-Crofts.
- Gibson, E. J., & Walk, R. D. (1956). The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *Journal of Comparative and Physiological Psychology*, 49(3), 239–242.
- Gogate, L. J., & Bahrick, L. E. (2001). Intersensory redundancy and 7-month-old infants' memory for arbitrary syllable–object relations. *Infancy*, 2(2), 219–231.
- Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, 71(4), 878–894.

- Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy*, 9(3), 259–288.
- Graham, S. A., & Kilbreath, C. S. (2007). It's a sign of the kind: Gestures and words guide infants' inductive inferences. *Developmental Psychology*, 43(5), 1111–1123.
- Hall, G. F. (1991). *Perceptual and associative learning*. Oxford, UK: Clarendon Press.
- Hallé, P. A., & Boysson-Bardies, B. d. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, 19, 463–481.
- Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, 23(1), 1–31.
- Jakobson, R., & Waugh, L. R. (1979). *The sound shape of language*. Bloomington, IN: Indiana University Press.
- Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1–23.
- Jusczyk, P. W., Bertoncini, J., Bijeljac-Babic, R., Kennedy, L. J., & Mehler, J. (1990). The role of attention in speech perception by young infants. *Cognitive Development*, 5, 265–286.
- Kluender, K. R., Diehl, R. L., & Kilean, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, 237, 1195–1197.
- Kluender, K. R., Lotto, A. J., Holt, L. L., & Bloedel, S. L. (1998). Role of experience for language-specific functional mappings of vowel sounds. *Journal of the Acoustical Society of America*, 104(6), 3568–3582.
- Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–274). Norwell, MA: Kluwer Academic Publishers.
- 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., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, 100(15), 9096–9101.
- 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.
- Lawrence, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, 39(6), 770–784.
- Lieberman, 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, 268–358.
- Lotto, A. J. (2000). Language acquisition as complex category formation. *Phonetica*, 57, 189–196.
- MacKain, K. S. (1982). Assessing the role of experience on infants' speech discrimination. *Journal of Child Language*, 9(3), 527–542.
- Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, 106(3), 1367–1381.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101–B111.
- Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11(1), 122–134.
- Miller, N. E., & Dollard, J. (1941). *Social learning and imitation*. New Haven, CT: Yale University Press.
- Narayan, C., Werker, J. F., & Beddor, P. (in press). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*.
- Norcross, K. J., & Spiker, C. C. (1957). The effects of type of stimulus pretraining on discrimination performance in preschool children. *Child Development*, 28, 79–84.
- Pisoni, D. B. (1979). Perception of speech sounds as biologically significant signals. *Brain Behavior and Evolution*, 16(5–6), 330–350.
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20(2), 421–435.
- Reese, H. W. (1972). Acquired distinctiveness and equivalence of cues in young children. *Journal of Experimental Child Psychology*, 13(1), 171–182.
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388(6640), 381–382.
- Stevens, K. N., & Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics*, 3(4), 215–233.
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13(5), 480–484.
- Swingle, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, 8(5), 432–443.
- Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*.
- Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, 56, 16–34.
- Trubetskoy, N. S. (1969). *Principles of phonology*. Berkeley, CA: University of California Press.
- Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences of the United States of America*, 104(33), 13273–13278.
- Vihman, M. M., Nakai, S., DePaolis, R. A., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50(3), 336–353.
- Waxman, S. R. (2002). Links between object categorization and naming: Origins and emergence in human infants. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 213–241). New York, NY: Oxford University Press.
- Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants. *Cognition*, 95(3), B59–b68.
- Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month old infants. *Cognitive Psychology*, 29, 257–302.
- Werker, J. F. (1995). Exploring developmental changes in cross-language speech perception. In L. R. Gleitman, & M. Liberman (Eds.), *Language: An invitation to cognitive science*, vol. 1 (2nd ed.). *An invitation to cognitive science* (pp. 87–106). Cambridge, MA: MIT Press.
- Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, 3(1), 1–30.
- Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, 24(5), 672–683.
- Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, 103(1), 147–162.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49–63.
- Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, 50, 509–535.
- Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, 85(3), 223–250.
- Xu, F. (2007). Sortal concepts, object individuation, and language. *Trends in Cognitive Sciences*, 11(9), 400–406.
- Yoshida, K., Fennell, C., Swingle, D., & Werker, J. F. (2009). Fourteen-month-olds learn similar-sounding words. *Developmental Science*, 12(3), 412–418.
- Yoshida, K., Pons, F., & Werker, J. F. (submitted for publication). Attention aids infants' phonetic learning.