Title: Phonological Knowledge Guides Two-year-olds' and Adults' Interpretation of Salient Pitch Contours in Word Learning

Author 1: Carolyn Quam, Department of Psychology, University of Pennsylvania
Author 2: Daniel Swingley, Department of Psychology, University of Pennsylvania

Address for correspondence:
Carolyn Quam
Institute for Research in Cognitive Science
3401 Walnut St., Suite 400A
University of Pennsylvania
Philadelphia, PA 19104

cmquam@gmail.com
(215) 898-0360
Fax: (215) 573-9247

Phonological Knowledge Guides Two-year-olds' and Adults' Interpretation of Salient Pitch Contours in Word Learning

Carolyn Quam and Daniel Swingley
University of Pennsylvania, 3401 Walnut Street, Suite 400A

Abstract

Phonology provides a system by which a limited number of types of phonetic variation can signal communicative intentions at multiple levels of linguistic analysis. Because phonologies vary from language to language, acquiring the phonology of a language demands learning to attribute phonetic variation appropriately. Here, we studied the case of pitch-contour variation. In English, pitch contour does not differentiate words, but serves other functions, like marking yes/no questions and conveying emotions. We show that, in accordance with their phonology, English-speaking adults and two-year-olds do not interpret salient pitch contours as inherent to novel words. We taught participants a new word with consistent segmental and pitch characteristics, and then tested word recognition for trained and deviant pronunciations using an eyegaze-based procedure. Vowel-quality mispronunciations impaired recognition, but large changes in pitch contour did not. By age two, children already apply their knowledge of English phonology to interpret phonetic consistencies in their experience with words.

Keywords: word learning; word recognition; phonology; prosody

To acquire the phonology of their native language, children must learn to assign appropriate interpretations to various sorts of phonetic variation. This learning process begins early in development. During the first year of life, infants home in on their native language's consonant and vowel categories, becoming better at discriminating some acoustically difficult native contrasts (Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Narayan, 2006) and worse at discriminating pairs of similar sounds that the native language groups into one category (Bosch & Sebastián-Gallés, 2003; Polka & Werker, 1994; Werker & Tees, 1984). By rendering irrelevant segmental distinctions difficult to discriminate, these developmental changes preclude certain linguistic errors. For example, an English-learning child who no longer readily perceives distinctions between dental and alveolar stop consonants is unlikely to mistakenly interpret dental and alveolar realizations of a word-initial /t/ as signaling two separate words.

There is more to phonological interpretation than the categorization of speech sounds, however. A great deal of phonetic variation that is readily perceptible may convey meaning at one or more levels of linguistic structure, in ways that are not universal across languages or retrievable from low-level distributional information in the signal. For example, vowel duration in American English serves functions like helping to signal prosodic boundaries (e.g., Salverda, Dahan, & McQueen, 2003; Turk & Shattuck-Hufnagel, 2000) and lexical stress (Lieberman, 1960), but generally provides only a secondary cue to identification of the vowel itself (e.g., Hillenbrand, Clark, & Houde, 2000). By contrast, many languages, like Japanese and Finnish, have distinct pairs of vowels that differ primarily in duration, so that identifying the exact vowel requires evaluating its duration. Because vowel duration is informative about *something* in all languages, the learner's task is to discover its function in her particular language–not simply whether it can be ignored altogether (Dietrich, Swingley, & Werker, 2007).

Most research on early perceptual development in phonology has been concerned with the changing *discriminability* of native and nonnative speech-sound contrasts, but *interpretation* of the sounds in words likely follows a different developmental course, at least for some phonological features—and may be governed by different learning principles. Two lines of evidence suggest that one- and two-year-olds are still figuring out how to apply their phonological categories in interpreting new words. First, young children do not consistently interpret single phonological-feature changes as indicating lexical distinctions (e.g., Nazzi, 2005; Pater et al., 2004; Stager & Werker, 1997). For example, Stager and Werker (1997) habituated 14-month-olds to the words *bih* and *dih*, paired with two different objects (in the "Switch" procedure). Despite substantial training with the words, infants apparently failed to connect the words to the objects; they did not look longer when the taught word-object pairings were violated than when they were maintained. The same age group succeeded with the dissimilar words *lif* and *neem*, and 17-month-olds succeeded with the similar-sounding words (Werker, Fennell, Corcoran, & Stager, 2002; see also Fennell, Waxman, & Weisleder, 2007; Fennell, 2006;

Thiessen, 2007; Yoshida, Fennell, Swingley, & Werker, 2009). Phonetic similarity also appears to play a stronger role among young children, relative to adults, in determining whether they treat phonological changes as indicating separate words—even when it is clear that children can *perceive* the phonological changes. Swingley and Aslin (2007) and White and Morgan (2008) found that 1.5-year-olds, upon viewing a familiar object (like a car) and a novel object, did not assume that a novel phonological neighbor of the familiar object's label (such as *gar*) referred to the novel object, though they did make this inference with more phonologically distinct nonwords, and did show some sensitivity to the mispronunciations. Swingley and Aslin's (2007) participants also showed much worse performance in learning novel words that were phonological neighbors of familiar words than in learning nonneighbors.

The second line of evidence that young children are still learning how to apply their phonological categories to word learning comes from findings that they appear to be more open-minded than older children about what they will treat as a word. Children under 18 months sometimes interpret noisemaker sounds, melodies, and gestures as words, while older toddlers do not. Namy (2001) successfully taught 17-month-olds gestures, sounds, and pictograms as object-category labels by embedding the symbols in familiar labeling routines. Namy and Waxman (1998) similarly found that 18-month-olds were willing to interpret both gestures and novel words as category labels, but found that 26-month-olds were reluctant to learn gestures as category labels, and required more practice with gestures before they would do so. Finally, Woodward and Hoyne (1999) found that 13-month-olds could learn the pairing of a new toy with either a novel word or a noisemaker sound, while 20-month-olds did not. These findings suggest fundamental changes around 18–20 months of age in children's expectations about how their language uses sound for reference (see also Roberts, 1995 and Fulkerson & Haaf, 2003).

As discussed, correct interpretation of phonological variation in word learning appears to follow a more protracted developmental course than the learning of language-specific phonetic categories. The present study further investigates children's interpretations of potentially relevant acoustic variability, focusing on interpretation of highly salient pitch contours. Pitch is a particularly interesting dimension of variation that English learners must interpret at appropriate levels of structure. In English, pitch varies systematically at the phrasal level (e.g., to mark yes/no questions, convey intonational meaning, and demarcate phrases), but it cannot contrast words. Since pitch is not contrastive in English, we might expect a particular word, like "good," to vary greatly in its pitch realization across tokens, because the pitch realization is not constrained by an underlying lexical tone. In English infant-directed speech, however, frequent words like "good" and "no" exhibit some consistency in their pitch patterns across tokens, probably because they tend to occur with particular pragmatic meanings and in stereotyped lexical contexts (Quam, Yuan, & Swingley, 2008). English-learning children must learn to interpret this pitch consistency at the phrasal level rather than the word level, even though it is potentially ambiguous between the two. Here, we

address whether English-learning toddlers correctly avoid attribution of pitch regularities to the word level when learning a novel word.

The curious case of pitch variation

Pitch is relevant at the lexical level in some but not all languages. In tone languages, words with very different meanings can differ only in their tone. For example, in Thai, *khaa* means *a grass* when pronounced with a mid tone, *to kill* when pronounced with a low tone, and *leg* when rising (Gandour, 1978). All of the world's languages—tone and nontone alike—convey meaning through phrasal intonation (e.g., the English phrase "oh, great" can mean very different things depending on its intonation). What makes tone languages special is that they use pitch contrastively, to distinguish words. There is mixed evidence about whether tone categories are clarified in infant-directed speech (IDS) or distorted by the exaggerated pitch patterns typical of IDS. Papousek and Hwang (1991) found that Mandarin speakers reduced or even neglected tone information in order to produce simple intonation contours to their two-month-old infants. In contrast, Liu, Tsao, and Kuhl (2007) found that tones in Mandarin IDS to 10- to 12-month-olds were not distorted by the sweeping pitch patterns of IDS, and were in fact *exaggerated* in a manner comparable to the exaggeration of vowel categories found in IDS (Burnham, Kitamura, & Vollmer-Conna, 2002). This difference could arise because parents' speech needs to convey different information to children of different ages: intonational meaning to younger infants and tone and segmental information to older infants (Kitamura & Burnham, 2003; Stern, Spieker, Barnett, & MacKain, 1983.) Thai speakers appear to exaggerate pitch contours in IDS to children from birth to 12 months, without causing much distortion of tones (Kitamura, Thanavishuth, Burnham, and Luksaneeyanawin, 2002). Even in speech to two-month-olds, however, Mandarin speakers appear to expand their pitch range and raise their pitch mean *less* than speakers of nontone languages, though they still produce the same intonational meanings (M. Papousek, H. Papousek, & Symmes, 1991).

Recent research has asked whether the acquisition of tone contrasts parallels that of consonant and vowel categories. The perceptual reorganization by which infants become worse at discriminating nonnative sound contrasts, but maintain good discrimination of native contrasts, occurs as early as six months for vowels: English-learning six-month-olds fail to discriminate some German vowel contrasts (Polka & Werker, 1994), and Spanish learners fail to discriminate the Catalan /ɛ/-/e/ contrast by eight months (Bosch & Sebastian-Galles, 2003). The reorganization is evident slightly later for consonants: while six-month-old English learners easily discriminate Hindi and Salish consonant contrasts, twelve-month-olds fail to do so (Werker & Tees, 1984).

Perceptual reorganization for tone seems to follow a similar trajectory; recent studies suggest that infants learning tone languages develop adult-like tone perception within the first year. Mattock and Burnham (2006) found that English learners failed to discriminate Thai tones by nine months, but Chinese learners—who were acquiring a tone language—did not undergo the same

worsening of discrimination with age. Harrison (2000) tested English-learning and Yoruba-learning six- to eight-month-old infants' perception of Yoruba tones. The Yoruba-learning infants were more sensitive than the English learners to changes in fundamental frequency (f0), but only in the region surrounding a tone boundary (190 versus 210 Hz). This response aligned with that of adult native speakers of Yoruba, providing evidence that the infants were already responding in an adult-like way to the tone contrasts.

Adults' perception of tones also suggests that listeners are shaped by their native-language structure. Mandarin speakers perceive Mandarin tones quasi-categorically, apparently assimilating the tones to linguistic categories, while French speakers perceive them continuously (suggesting French speakers perceive the tones psychophysically vs. linguistically; Halle, Chang, & Best, 2004). Finally, there is evidence that tones, like other speech sounds, form classifiable clusters. An unsupervised learning algorithm can learn the four tone categories of Mandarin from pitch movement in syllables extracted from fluent speech (Gauthier, Shi, & Xu, 2007).

Evidence from children's productions suggests that the reliability of the realization of tones affects their age of acquisition. Hua and Dodd (2000) found early acquisition of tones in Putonghua (Modern Standard Chinese, a variety of Mandarin). For children between the ages of eighteen months and 4.5 years, tone errors were rare relative to consonant and vowel errors. The distribution of production errors across the age groups suggested that Putonghua-learning children acquire tones first, then vowels and syllable-final consonants, then syllable-initial consonants. In another language, Sesotho, words' surface forms often diverge from their underlying tones because of pervasive tone sandhi. Demuth (1995) found a slower, more item-specific acquisition of tone in Sesotho than had been found for lexical tone languages. This suggests that the reliability of the mapping between underlying tone and surface form has a large impact on the speed of acquisition of a tone (see also Ota, 2003).

Beyond acquisition of tones, we can ask how perception and interpretation of pitch cues to other levels of structure develop. In English, pitch demarcates phrase boundaries (Gussenhoven, 2004), marks yes/no questions (with a terminal rise), and cues lexical stress, e.g., helping distinguish the noun *PERmit* from the verb *perMIT* (Fry, 1958; for reviews, see Ladd, 1996, and Gussenhoven, 2004, Chapter 2). Because of contrastive stress pairs like these, there is a sense in which pitch can help contrast words in English. But in these cases other correlated cues to stress, including vowel quality, vowel duration, and amplitude, contribute strongly to the contrast. Cutler and Clifton (1984) found that adults were slower to identify words when the acoustic cues to stress were naturally produced to stress the wrong syllable. This mispronunciation effect occurred even when the unstressed vowel was unreduced—meaning the vowel-quality cue was essentially neutralized—but the effect was greater when the unstressed vowel was reduced. It is not yet known whether listeners can exploit an *isolated* pitch cue to stress in word recognition.

Pitch also conveys highly complex intonational meanings in adult-directed speech, through particular, stereotyped contours. The ToBI transcription system (Pierrehumbert, 1980; Beckman, Hirschberg, & Shattock-Hufnagel) was developed to characterize different intonation contours in English as a series of High and Low tones, and has led to the identification of certain, fairly reliably realized intonational meanings. For example, the 'fall-rise' or 'rise-fall-rise' pattern conveys uncertainty or incredulity in some sentential contexts (Ward & Hirschberg, 1985; Hirschberg & Ward, 1992), while the 'continuation rise' contour can convey that the speaker is about to continue talking (Bolinger, 1989).

For very young infants who have not begun learning words, the meaning of caregivers' speech is carried entirely by prosodic characteristics, particularly intonation. The distinctive pitch characteristics of infant-directed speech (IDS) complement the infant's developing auditory system; the higher f0 mean and wider f0 range make the speech more interesting and easier for the infant to tune in to (Fernald, 1992). Infants prefer listening to IDS over adult-directed speech (ADS; Fernald, 1985), a preference driven primarily by IDS's pitch characteristics (Fernald & Kuhl, 1987; Katz, Cohn, & Moore, 1996). Some pragmatic functions of speech are expressed more clearly in IDS than in ADS; listeners are more successful at identifying the pragmatic functions of content-filtered IDS utterances than comparable ADS utterances (Fernald, 1989). Considering the clarity of intonational meaning in IDS, it is not surprising that infants can categorize utterances from different emotional classes before they know many words (Moore, Spence, & Katz, 1997).

Despite the relevance of pitch at nonlexical levels of structure, and the clear importance of pitch in parental communication to infants, the English-learning child must learn to disregard intonational pitch as a lexically contrastive feature when establishing new lexical entries and in recognizing words. A recent study by Singh, White, and Morgan (2008) provides some evidence for development in infants' categorization of word forms varying in pitch. Singh et al. familiarized infants to words in isolation and tested their recognition of those words in sentences, using a procedure that evaluates infants' preference for familiarized versus novel materials. When the pitch realization matched between familiarization and test, both 7.5-month-olds and 9-month-olds preferred to listen to the sentences containing familiarized words. When the familiarized words were realized with different pitch, however, only the 9-month-olds preferred to listen to the familiarized words, suggesting that the younger infants failed to recognize them. In the second half of the first year, therefore, infants appear to become better able to recognize words despite changes in pitch. Still, this leaves open the phonological status of linguistic pitch in two ways. First, the pitch manipulation tested by Singh et al. (2008) involved an absolute change in the words' pitch levels, produced by raising or lowering all pitch samples by six semitones. Nine-month-old infants might still be thrown off by changes in intonation *contour* (e.g., Trehub & Hannon, 2006). Second, developmental changes in infants' matching of different realizations of a word form may bear more on a general property of

infant memory (e.g., a decrease over development in the number of perfectly matching features required for a new stimulus to be matched to a prior one) than on children's *interpretation* of how speech conveys meaning.

The distinction between interpretation and simple acoustic matching is also an issue for studies showing similar improvement in children's ability to recognize words despite changes in talker's voice or affect. At 10.5 months—but not at 7.5 months—infants successfully generalize familiarized words from male to female voices,[1] or when the affect changes (from happy to neutral or vice-versa) across familiarization and test (Singh, Morgan, & White, 2004; see also Houston & Jusczyk, 2003). Studies of how infants match different tokens of a word form are informative about foundational mental capacities that underlie language acquisition, but they do not necessarily indicate how phonetic variation is interpreted referentially. Even adults are better at recognizing a word when it is spoken by the original voice (Palmieri et al., 1993; Goldinger, 1996). Rather than tuning out irrelevant information completely, we apparently become more adept, over development, at focusing on essential properties of words, like phonemes and stress patterns. One way to view this process follows Jusczyk (1993) in proposing that exposure to the native language leads the system to weight relevant features more heavily and irrelevant features less heavily.

Learning how pitch is used in English requires separating pitch from the lexical level and learning intonational categories cued by pitch. Young children's speech does contain a range of intonational contours that often sound familiar enough to be interpreted referentially by adults, but few studies have shown that young children analytically separate the intonational characteristics of words from their segmental characteristics (see Vihman, 1996, for a review). Galligan (1987) reports a case study of two children, who amid their second year each used single words with more than one intonational contour in ways that could be interpreted as being appropriate for the communicative context. This sort of evidence suggests that children attempt to interpret and produce sentence intonation, and may succeed in separating the pitch properties of an utterance from the utterance's lexical context. However, it does not necessarily follow that children command a linguistic system that rules out pitch contours as relevant for distinguishing words. Establishing this stronger claim requires an empirical test, like the current one, in which the child's experience with a word provides evidence for a (grammar-inconsistent) interpretation in which the word has intrinsic pitch, and the child must attribute that consistent pitch pattern to the intonational level rather than the lexical level. The apparent difficulty of this correct attribution depends upon whether one assumes that toddlers interpret speech in a holistic fashion, encoding words as a mass of relatively unanalyzed sensory properties, or in an analytical fashion, potentially attributing various phonetic properties of a word token to separate linguistic levels of interpretation.

---

[1] Male versus female voices differ more in their fundamental frequency than two female or two male speakers (Houston & Jusczyk, 2000).

The issue of interpreting pitch at the appropriate levels of structure has hardly been addressed in the developmental speech perception literature, in which discussion of holistic or analytic representations has focused on segmental phonology (consonants and vowels) rather than intonation. In that context, the analytic viewpoint holds that young children's lexical representations can be described using the conventional inventory of consonants and vowels (e.g., Swingley, 2003), whereas the holistic viewpoint argues either that children's knowledge of the sounds of words is less clearly specified (many features are missing) or that children's lexical representations are not made up of a sequence of categories at all (e.g., Metsala & Walley, 1998; Storkel, 2002; for discussions, see Swingley, 2007; Vihman & Croft, 2007; Werker & Curtin, 2005).

More generally, the notion that children interpret speech analytically is at variance with simple exemplar models in which the lexicon provides the sole level of organization relevant to word recognition (see Goldinger, 1998, for what he describes as an "extreme" model of this sort, and for discussion of more richly structured alternatives). If the recognition of words depends entirely on the overall phonetic or acoustic match between the current token and the mass of previously experienced tokens of that word, prior experience with a particular word's realizations should trump phonological generalizations derived from analysis of the other known words of the language. Listeners do retain voice- or otherwise token-specific information about experienced words (e.g., Goldinger, 1996; Nygaard & Pisoni, 1998), which rules out models in which formal linguistic content *alone* guides behavior. But the existence of such effects does not imply that phonological analysis is unnecessary (Pierrehumbert, 2006). Studies supporting exemplar models rarely calibrate the effects of nonphonological information, like talker's-voice characteristics, against a phonological baseline. In Experiment 1, we test the hypothesis that adults will weigh much more heavily those phonetic changes that are *relevant* for distinguishing words in English, than changes that, though perceptually salient, are not lexically contrastive. If we find that adults are sensitive to changes in pitch contour, this will support the holistic, or exemplar, perspective; we will then be in a position to assess the relative importance of lexically relevant and irrelevant phonetic variation within that perspective. If we find that adults show large effects of lexically relevant changes, but not changes in pitch contour, this will support analytic views of speech interpretation—or exemplar views in which the phonetic dimension of pitch is weighted extremely weakly.

Overview of the two experiments

We taught both adults and 2.5-year-olds a new word, always pronounced with a consistent, salient pitch contour, and then tested their interpretations of a nonphonemic change in the word's pitch contour versus a phonemic change in the word's vowel. We first tested adults, in Experiment 1, in order to establish the mature interpretation of these changes. In Experiment 2, we tested 2.5-year-olds in the same task. We selected an age at which children should treat

the vowel change as relevant, since we wanted to compare interpretations of the pitch-contour change to this phonological baseline. Seventeen- to twenty-month-olds sometimes struggle to differentiate similar-sounding words in teaching contexts (Swingley & Aslin, 2007), so we wanted to ensure that processing constraints (e.g., failure to remember which version of the word was taught and which was the change) would not prevent children from interpreting a mispronunciation as a new word. Our selection of 2.5-year-olds for Experiment 2 was also motivated by evidence of developmental change in children's interpretation of pitch cues to emotion, over the ages of 3 to 4 years (Quam & Swingley, 2009). This suggests an especially protracted learning course for interpretation of pitch structure in English.

Experiment 1

Three questions led us to test adults as well as 2.5-year-olds. First, although adult native-English speakers are naturally expected to have acquired the phonology of English, in which pitch contours cannot be interpreted lexically, adults might still recognize words best when the test instances are most similar to the training instances. This result would be consistent with the episodic-lexicon model and with evidence that adults retain subsegmental and indexical information about words (e.g., Goldinger, 1996; Nygaard & Pisoni, 1998). A comparison between adults' and children's sensitivity to changes in pitch contour could also shed light on whether children's interpretation of nonphonemic dimensions becomes adult-like through the fine-tuning of attention weights to different acoustic dimensions (Jusczyk, 1993). Second, despite their knowledge of native phonology, adults could choose to interpret the highly salient pitch change as relevant, treating a word with altered, "mispronounced" pitch as a worse version of the newly learned word than the word with the original pitch contour. Third, adults could interpret the vowel change either as an entirely new word, referring to a different object, or as a mispronunciation of the taught word. We were interested in whether adults, who have reached the endpoint of phonological development, would be uniform in their responses, or whether we would still see individual variation in interpretations of the two changes.

Method

Participants

Twenty-four adults, nine male, and all native speakers of English were included in the analysis. (One of these participants was also a native speaker of Spanish; his responses were typical.) All participants but one were undergraduates (the exception was a postdoctoral researcher), assumed to be between 17 and 23 years old. Ten more participated but were excluded: six for experimenter error / equipment failure, two for failure to follow instructions to fixate the pictures, and two for their language backgrounds (one was a nonnative speaker of English, the other was a native bilingual of English and Chinese).

Apparatus and Procedure

We used a language-guided looking procedure to investigate how adults would interpret a phonological (vowel-quality) versus nonphonological (pitch-contour) change in a newly learned word. Since adults participated in essentially the same experiment as the toddlers in Experiment 2, the stimuli were designed for children. To make this experience less odd, adult participants were told before the study that they would be helping to calibrate an experiment designed for two-year-olds.

Participants sat in front of a large display screen, on which they viewed pictures. Concealed speakers played recorded sentences that referred to the pictures, and a hidden video camera in the center of the display captured participants' eye movements, which were later coded by hand.

The experiment lasted twenty minutes and consisted of four phases (see **Figure 1** for the experimental design). The first two phases, the *animation* and *ostensive-labeling* phases, taught participants a novel word. In the animation phase, adults watched a five-minute, narrated, animated video in which a monkey presented his two toys to several potential playmates. One toy was labeled ten times as the "deebo" (IPA: [diboʊ]) in sentences like, "This is my deebo. Would you like to play with it?" The word was pronounced with a highly consistent, distinctive intonation contour commonly found in speech to infants: either a rise-fall or a low fall (see **Figure 2** for spectrograms and pitch tracks of the two pitch contours). The other toy was present and talked about equally often, but never labeled, in sentences like, "This is my other toy. Would you like to play with it?" In the ostensive-labeling phase, each toy appeared independently on the screen. The *deebo* was labeled four times in each of three trials, for a total of twelve repetitions, in sentences like: "This is a deebo. Deebo. Look at the deebo. The deebo." The other toy was talked about, but not labeled, in sentences like: "Look at this toy. Isn't it pretty? Would you like to play with it?"

The third phase, the *test*, contained 18 critical trials. In these trials, the two toys appeared side by side. In eight *trained-pronunciation* trials, participants were asked to locate the "deebo," in sentences like, "Where's the deebo? Can you find it?" In the other ten trials, adults heard a word that differed from the taught pronunciation in one of two ways. In five *vowel-change* trials, participants heard "dahbo" (IPA: [dɑboʊ]) with the original pitch contour; in five *pitch-change* trials, they heard "deebo" with a different pitch contour. Half the participants were originally taught the word *deebo* with a rise-fall contour (which changed to the low fall on pitch-change trials), and the other half were taught *deebo* with a low fall contour (which then changed to the rise-fall on pitch-change trials). In addition to these 18 critical trials, 69 familiar-word trials were interspersed throughout the ostensive-labeling and test phases. These familiar-word trials presented two familiar objects and asked adults to orient to one of them, in sentences like, "Look at the shoe. That's pretty." Target words in the familiar-word trials were produced with natural intonation and no segmental mispronunciations. These familiar-word trials, along with 20

short, attention-getting animations, were intended to distract adults from the purpose of the experiment and also to prevent boredom and sleepiness.

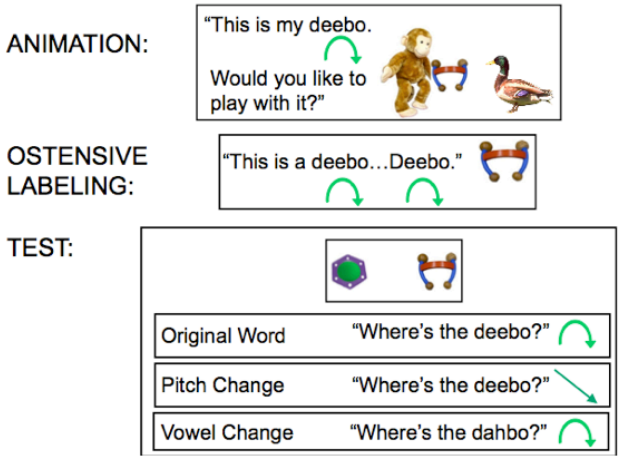

Figure 1: Experimental design. In the *animation* phase, participants heard the word "deebo" spoken with the same intonation contour (the rise-fall is used in this example) ten times in a story. Next, in the *ostensive-labeling* phase, the *deebo* was labeled directly twelve times. In *test* trials, the *deebo* and distracter objects were presented side-by-side. Adults heard eight trained-pronunciation (original-word) trials and five trials of each change type. Children heard the original word in eight trials and *either* the pitch change or the vowel change in the other eight trials. Finally, participants were asked to point to and name the objects (not pictured).

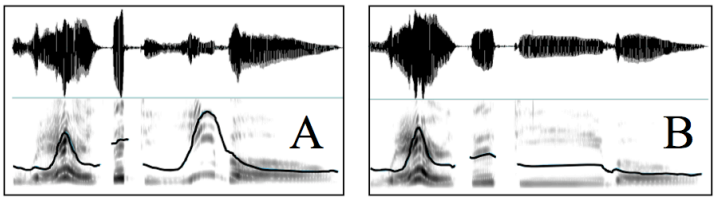

Figure 2: Intonation contours. Waveform, spectrogram, and pitch contour for the rise-fall contour (**A**) and low fall contour (**B**), in the sentence "Where's the deebo?"

Finally, in the *pointing-and-naming* phase, adults were asked to point to and name the objects. In *pointing* trials, both novel objects appeared on the screen and participants were asked to "Point to the [deebo]." The word was pronounced with the trained pronunciation and each of the changed pronunciations from the test phase, for a total of three trials. In *naming* trials, each object appeared separately on the screen next to a picture of the Sesame Street character Elmo, and participants heard, "Elmo doesn't know what that is. Tell Elmo what that is!"

After the experiment, adults filled out a questionnaire. The questions evaluated whether each participant had correctly learned the word-object pairing for *deebo*; whether she had noticed the pitch and vowel changes; and

whether she had interpreted the word "dahbo" as a label for the distracter object, or merely as a mispronunciation of "deebo."

Auditory Stimuli

A native English speaker (the first author) recorded auditory stimuli in clear child-directed speech, with exaggerated, infant-directed prosody and at a normal speaking volume. The animation sentences were embedded in a narration, similar to a storybook (e.g., "This is my deebo. Would you like to play with it?"). They accompanied an animated movie, meant to familiarize participants with the pairing of the word "deebo" and the object. "Deebo" was always spoken with a consistent intonation pattern: either a rising then falling contour (referred to as rise-fall) or a level, medium pitch followed by falling pitch (referred to as low fall; see **Figure 2** for spectrograms and pitch tracks of the two pitch contours). We chose pitch contours that could be interpreted either as lexical pitch or as phrasal intonation, because we wanted to avoid pushing participants into one interpretation or the other.

Ostensive-labeling sentences directly labeled the *deebo* object (e.g., "This is a deebo. That's right. Look at the deebo. The deebo."). In the animation and ostensive-labeling sentences, the word "deebo" was always spoken with the same intonation contour, though tokens were allowed to vary somewhat in length, absolute pitch, and amplitude (this variation helped them sound natural in context). In test sentences, participants were asked either "Where's the [deebo/dahbo]?" or "Which one is the [deebo/dahbo]?" The duration, pitch contour, and amplitude of test words were controlled carefully. Pointing sentences were comparable to test sentences, but asked participants to "Point to the [deebo/dahbo]." In all sentences, the word "deebo" (or "dahbo") occurred at the end of the sentence, where the pitch contours and duration sounded most natural. Sentences were always naturally produced, but in some cases the length or amplitude of the word was modified slightly using Praat sound-editing software (Boersma & Weenink, 2008). See **Appendix 1** for duration, maximum pitch, and mean pitch of each word token.

Visual Stimuli

Visual stimuli were displayed on a rectangular plasma video screen measuring 37 by 21 inches. In the animation phase, these stimuli consisted of photographs of objects, moving around in front of a painted scene of a grassy hill. A plush toy monkey moved around the scene, manipulating two novel toys and playing with other animals. Visual stimuli in the ostensive-labeling and test phases consisted primarily of photographs of objects on gray backgrounds. In ostensive-labeling trials, novel toys from the animation appeared on the screen alone, while in test trials, the two toys were displayed side by side. At the beginning of each ostensive-labeling and test trial, the deebo and/or distracter objects hopped or twisted on the screen (this was intended to get children's attention in Experiment 2), after which they remained still. All photos were edited to balance their salience by roughly equating brightness and size. The two novel toys were a purple-and-green

plastic disk (subsequently referred to as the *purple disk*) and a red-and-blue knobby wooden object (subsequently referred to as the *red knobs*; see **Figure 3**). The particular object that was labeled the "deebo" varied across participants, and was crossed with which pitch pattern they heard during the teaching.

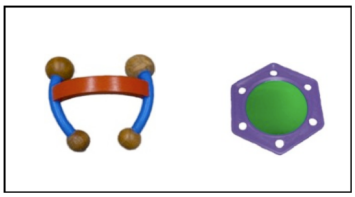

Figure 3: The two objects used in teaching and testing. On the left is the *red-knobs* object, and on the right is the *purple-disk* object. For each participant, one of these objects was labeled the "deebo" and the other was present equally often but never labeled.

Coding

After testing, trained coders, blind to target side, coded the direction and timing (beginning and end) of every eye movement a participant initiated during each trial. Eye movements were coded frame-by-frame with 33-millisecond resolution using the *SuperCoder* software (Hollich, 2005). Alignment of the timing of eye-movement events with auditory and visual stimulus events was ensured using a custom hardware unit that placed visible signals into the recorded video stream of the participant's face.

For each participant in each trial, we calculated the proportion of the time he or she fixated the *deebo* object (the amount of time spent looking at the *deebo* divided by the total time looking at either picture). We calculated this *deebo* fixation proportion over a specified time window after the onset of the target word: 200 to 2000 ms post–noun-onset. This time window is similar to the window commonly used with young children, 367 to 2000 ms (see Experiment 2 for an explanation for the time window used with children), but begins earlier because adults are known to respond more quickly than toddlers in this procedure (e.g., Swingley, 2009).

Results and Discussion

Adults provided four types of responses: looking times to each picture, elicited pointing and naming of the pictures, and questionnaire responses. Looking times provide a gradient measure of interpretation of the auditory stimulus, while pointing and naming force participants to make a discrete and conscious choice. Naming responses also allow us to probe for encoding of pitch and segmental information. Finally, questionnaire responses allow us to determine participants' final interpretation of the stimuli.

The pronunciation of test words (trained pronunciation, pitch change, or vowel change) exerted a significant effect on adults' fixation of the *deebo* in an analysis of variance ($F(2,69) = 77.16$, $p < .001$). There were no main effects, or interactions with trial type, of which object was the *deebo*, which pitch contour

was taught, or which type of change was presented first in the test phase. Planned comparisons thus further investigated only the effect of condition (pronunciation of the word) on *deebo* fixation.

When they heard the trained pronunciation of the word, adults fixated the *deebo* object significantly above chance, or 50% (mean, 91.8%; paired t(23) = 31.38; *p*(all tests 2-tailed) < .001). Participants also fixated the *deebo* above chance in response to the pitch change (mean, 89.3%; paired t(23) = 17.89; $p < .001$), and their accuracy did not differ significantly from their accuracy in response to the trained pronunciation.

In response to the vowel change, participants actually fixated the *deebo* below chance (this difference approached significance; mean, 39.7%; paired t(23) = -1.98; $p = 0.06$), and significantly less than in trained-pronunciation trials (paired t(23) = 10.13; $p < .001$) or pitch-change trials (paired t(23) = 9.29; $p < .001$). Every participant fixated the *deebo* less in vowel-change trials than in trained-pronunciation trials (see **Figure 4**). Eighteen of the 24 participants (75%) fixated the *deebo* less than 50% of the time in response to the vowel change, suggesting they used a mutual-exclusivity strategy (Markman & Wachtel, 1988), interpreting "dahbo" as a label for the distracter object. In pitch-change trials, by contrast, no participants fixated the *deebo* less than 50% of the time, and exactly half of the participants (12/24) fixated the *deebo* less in pitch-change trials than in trained-pronunciation trials.

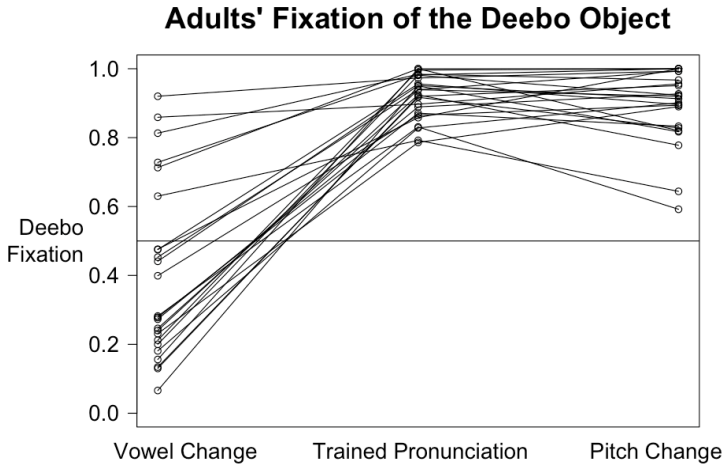

Figure 4: Adults' fixation of the *deebo* object in each trial type. The horizontal line indicates chance fixation, or 50%. Adults' fixation of the *deebo* object showed a large effect of the vowel change. All 24 participants fixated the *deebo* less in vowel-change trials than in trained-pronunciation trials, and 75% fixated the *deebo* less than 50% of the time in vowel-change trials. In contrast, adults showed no effect of the pitch change; only half of participants fixated the *deebo* less in pitch-change trials than in trained-pronunciation trials, and no participants fixated the *deebo* less than 50% of the time in pitch-change trials.

Adults' pointing, naming, and questionnaire responses provide additional insight into their interpretations of the pitch and vowel changes. **Tables 1** and **2** display adults' pointing and naming responses, respectively. When asked to "point to the deebo," regardless of the pitch contour used, all 24 adults pointed to the *deebo*. In contrast, responses to "point to the dahbo" were more varied: 19/24 participants pointed to the distracter object (though four of those showed uncertainty, assessed informally, either through their facial expression, their words, or rising intonation), while the other five participants pointed to the *deebo*. When asked to label the *deebo*, 22/24 participants said "deebo," while the other two did not name it. When asked to label the distracter object, 15/24 participants said "dahbo," (five of whom showed uncertainty), seven did not label it, one said "deebo," and one said "doba." (The latter two participants wrote on the questionnaire that they interpreted "dahbo" as a label for the distracter, but they incorrectly reproduced "dahbo" as "dubbo" and "doba," respectively, suggesting they were having trouble remembering or reproducing the /a/ vowel.) The pitch characteristics of adults' labeling responses did not reflect the pitch contour used in teaching; analyses of variance predicting the f0 maximum and f0 mean of labeling responses from the interaction of taught pitch (rise-fall or low fall) and which object participants were labeling (*deebo* or distracter) showed no significant effects.

Table 1: Adults' pointing responses. Points to the *deebo* / number of adults pointing (percentage pointing to *deebo*), for each condition.

Trained pronunciation		Pitch change		Vowel change	
24 / 24	(100%)	24 / 24	(100%)	5 / 24	(21%)

Table 2: Adults' naming responses. Number of responses / number of adults (percentage giving the particular response), for each object.

	Viewing *deebo*		Viewing distracter	
Said "deebo"	22 / 24	(92%)	1 / 24	(4%)
Said "dahbo"	0 / 24	(0%)	15 / 24	(63%)
Did not name / Used different vowel	2 / 24	(8%)	8 / 24	(33%)

Though the acoustic measurements did not reveal differences between adults' productions depending on which pitch contour they were taught, it could be that human judges would be more sensitive to subtle differences not captured by the acoustic measurements we used. With this in mind, ten new adult judges were trained to identify rise-fall or low fall contours. They were first given training exemplars taken from the training and test phases of the original experiment, and then were tested on classification of twelve more

exemplars. Only one adult made an error during this phase, on one of the twelve trials. The judges were then asked to categorize the experimental participants' productions as rise-fall or low fall contours. Adult productions were mixed in with child productions and presented in random order; classifications of the child productions are reported in the **Experiment 2 Results**. The judges' classifications of the adults' productions did not reflect the pitch contour participants were taught ($F(1,31) = .81$, $p = .38$), the object they were labeling ($F(1,31) = .09$, $p = .77$), or their interaction ($F(1,31) = 1.10$, $p = .30$) in an analysis of variance using the number of rise-fall classifications for each utterance (out of a possible ten) as the dependent variable. Judges assigned the "rise-fall" classification to participants' labels of the *deebo* object at similar rates regardless of which contour was taught (taught rise-fall, mean 4.56, SE 0.69; taught low fall, mean 4.55, SE 0.65). "Rise-fall" classification of participants' labels for the distracter object were also not significantly related to the taught contour (taught rise-fall, mean 3.57, SE 1.00; taught low fall, mean 5.00, SE 0.57). Participants' failure to imitate the taught pitch contour in their own productions suggests they did not consider the pitch pattern to be a relevant component of the word's sound.

In questionnaire responses, all 24 adults reported noticing the vowel change, and 17/24 reported having learned both "deebo" and "dahbo" as object labels. In contrast, only 12/24 participants reported noticing the pitch change. Eight of the twelve participants who did not report the change did remember it after prompting, either when the experimenter asked, "Did you notice any other changes in the word?" or when the experimenter reproduced the pitch contrast for them. No participants reported learning two words that contrasted in pitch.

Adults' responses across our measures of their learning were fairly consistent. Similar numbers of participants demonstrated learning of "dahbo" on each measure. Eighteen participants looked more at the distracter, and 19 pointed to the distracter, in response to "dahbo"; 15 labeled the distracter "dahbo"; and 17 reported learning the word "dahbo." Still, individual participants were not always wholly consistent. Thirteen participants showed all the behaviors consistent with learning the word "dahbo" (looking more to, pointing to, and labeling the distracter; and reporting having learned both words), and three participants showed *no* evidence of learning the word "dahbo." But eight participants exhibited some but not all behaviors associated with learning "dahbo," suggesting they did not commit to one single interpretation of the vowel change.

To summarize, adults universally showed no effect of the pitch change, fixating the *deebo* object equally in response to the trained pronunciation and the pitch change. They also universally showed sensitivity to the vowel change; all participants fixated the *deebo* less in response to the vowel change than in response to the trained pronunciation. Though we expected that adults might consistently interpret the large vowel change (from /i/ to /a/) as signaling a new word, we found instead that adults were fairly variable in their interpretations. This was true both across participants and, sometimes, within

individuals. All participants *noticed* the vowel change, as evidenced both by their questionnaire responses and their decreased fixation of the *deebo* in response to the vowel change. Detection and interpretation, however, are distinct.

Experiment 2

We next tested 2.5-year-olds in the same experiment, asking whether their interpretations of the pitch and vowel changes would be adult-like, reflecting their native phonology, or not yet fully developed. Children's responses could differ from the adult standard in two ways: children could treat the pitch change as lexically relevant, or they could fail to show sensitivity to the segmental change. Sensitivity to the pitch change would be consistent with evidence that young children are more open-minded than older listeners in interpreting new words (e.g., Namy, 2001; Namy & Waxman, 1998; and Woodward & Hoyne, 1999), and with evidence of a protracted developmental course for correct interpretation of pitch at other levels (e.g., pitch cues to emotions; Quam & Swingley, 2009). Lack of sensitivity to the vowel change is less likely, since 30-month-olds should be more sensitive to segmental changes than the younger children tested in previous experiments (e.g., 14-month-olds in Stager & Werker, 1997; 1.5-year-olds in Swingley & Aslin, 2007 and White & Morgan, 2008). We chose 30-month-olds for this reason, since we wanted the phonologically relevant change in the vowel to serve as a baseline for comparison with interpretations of the pitch-contour change. Still, children appear to be less sensitive to vowel changes than to consonant changes (Nazzi, 2005, testing 20-month-olds), and the pitch consistency in our teaching phase could also dampen children's sensitivity to the vowel change, given that increased variability in talker's voice (Rost & McMurray, 2009) and in affect (Singh, 2008) improve children's sensitivity to subtle contrasts.

Method

The design, apparatus, and stimuli were comparable to Experiment 1. Children saw the same *animation* and *ostensive-labeling* phases as in Experiment 1. The other two phases differed slightly from the adult version. The *test* phase had three important modifications. First, because of children's more limited attention spans, each child heard *either* the vowel or the pitch change in the test trials, not both. The experiment contained eight *trained-pronunciation* trials, either eight *pitch-change* trials or eight *vowel-change* trials, and only ten *familiar-word* trials (instead of the 69 included in the adult experiment). Finally, children also participated in the *pointing-and-naming* phase, but heard only two pointing trials, corresponding to the trained pronunciation and the pronunciation change the child heard in test. As in Experiment 1, there were two naming trials, one for each toy. In each pointing or naming trial, if the child did not point or speak, the trial was replayed and the parent and experimenter encouraged the child to respond without biasing her response. Parents kept their eyes closed in both the test and pointing-and-naming phases to avoid biasing the child's responses. Within a week of the test

date, parents completed the MacArthur Communicative Development Inventory of Words and Sentences (Fenson et al., 1994), which measured their child's productive vocabulary.

Participants

Forty-eight children between the ages of 29 months, 3 days and 32 months, 8 days were included in the analysis. All participants were learning English as their dominant language and hearing it at least 2/3 of the time, as reported by their caregivers. Twenty-four children, 13 male, were included in the *vowel-change* condition (mean age 30 months, 19 days, *SD* = 24 days; mean productive vocabulary 512 words, *SD* = 154 words); and 24 children, 13 male, were included in the *pitch-change* condition (mean age 30 months, 17 days, *SD* = 30 days; mean productive vocabulary 468 words, *SD* = 181 words).

Fifteen more children participated but were excluded (four from the pitch condition, eleven from the vowel condition) for having fewer than six usable trials (including the point trial) in any of the trial types (familiar-word, trained-pronunciation, or changed-pronunciation trials). Trials were only included as usable if the child fixated the pictures for at least 10 frames during the analysis window, out of a possible 50.

Results and Discussion

We calculated children's fixation of the *deebo* over a specified time window after the onset of the target word (beginning slightly later than the window used with adults): 367 to 2000 ms after noun onset. Before 367 ms, children are unlikely to be responding to the target word (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Swingley & Aslin, 2000). After 2000 ms, they are likely to have completed their response and moved their attention elsewhere.

Before asking whether children responded to changes in the word's pronunciation, we had to determine whether they learned the word at all, by comparing children's *deebo* fixation to chance fixation, or 50%. In trials where "deebo" was spoken with the trained pronunciation, both groups' *deebo* fixation was significantly above chance (vowel-change group: mean, 67.4%; paired $t(23) = 5.73$; p(all t-tests 2-tailed) < .001; pitch-change group: mean, 66.3%; paired $t(23) = 5.60$; $p < .001$).

Next, we considered whether either the pitch change or the vowel change significantly affected children's fixation of the *deebo* object. **Figure 5** displays *deebo*-fixation proportions for each group in trained-pronunciation and change trials. Trial type (trained vs. changed pronunciation) interacted significantly with condition (pitch vs. vowel) in an analysis of variance ($F(1,92) = 11.57$, $p < .001$). The vowel change caused a significant decrease in *deebo* fixation compared with responses to the trained pronunciation (mean decrease, 15.0%; paired $t(23) = -3.50$; $p < .005$), exhibited by 20/24 participants (binomial $p < .001$). Additionally, 11/24 participants actually fixated the *deebo* less than 50% of the time in response to the vowel change (compared with only 2/24 children who did so in response to the pitch change), suggesting they may have

used a mutual exclusivity strategy to map the word "dahbo" onto the distracter object (Markman and Wachtel, 1988). Overall, looking to the *deebo* in response to the vowel change did not differ from chance (mean, 52.8%; paired $t(23) = 0.77$; $p = 0.45$).

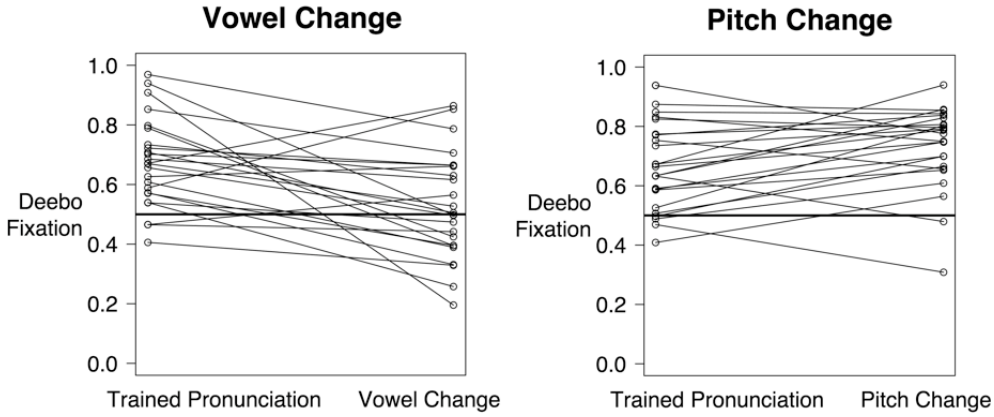

Figure 5: Thirty-month-old children's fixation of the *deebo* object in each trial type. ***Left:*** Each vowel-change participant's fixation of the target object (the *deebo*) in response to the trained pronunciation and the vowel change; the horizontal line indicates chance fixation, or 50%. The vowel change caused a significant decrease in *deebo* fixation (15% on average) compared with responses to the trained pronunciation. ***Right:*** Each pitch-change participant's *deebo* fixation in response to the trained pronunciation and the pitch change. The pitch change actually caused a significant, though smaller, *increase* in target fixation (6.6% on average), perhaps because its novelty increased children's attentiveness.

Children exhibited a much different response to the pitch-contour change. Instead of a decrease in *deebo* fixation, we found a small *increase* compared with responses to the trained pronunciation (mean increase, 6.6%; paired $t(23) = 2.40$; $p < .05$). This effect of the pitch change was less than half the size of the vowel-change effect, and less consistent: 16/24 participants fixated the *deebo* more in response to the changed pitch (binomial $p > .05$). Still, this effect was unexpected. We speculate that the pitch change, after a long familiarization with one consistent pitch contour, may have made children more attentive and thus more successful at orienting to the target. Overall, looking to the *deebo* in response to the pitch change was significantly above chance (mean, 72.9%; paired $t(23) = 8.16$; $p < .001$).

When comparing children's fixation of the *deebo* object to chance, we face the risk that children might be biased to look at one picture or the other, making 50% an inadequate baseline. To alleviate this concern, we conducted analogous tests in which we subtracted *deebo* fixation before the word's onset from *deebo* fixation in the 367-2000 ms window. We then compared this difference score to chance, or 0%. These tests yielded the same pattern of significance as the tests reported above. Children increased their *deebo* looking

upon hearing the *trained pronunciation* in both the pitch-change group (mean increase, 13.1%; paired $t(23) = 3.32$; $p < .005$) and the vowel-change group (mean increase, 14.8%; paired $t(23) = 3.39$; $p < .005$). Children also increased their *deebo* looking in response to the pitch-change pronunciation (mean increase, 20.7%; paired $t(23) = 5.70$; $p < .001$). In response to the vowel-change pronunciation, in contrast, children's increase in *deebo* fixation did not differ from chance (mean increase, 2.2%; paired $t(23) = 0.73$; $p = 0.47$).

Next, we asked whether participants' age would affect their sensitivity to either change in pronunciation. We computed an analysis of covariance (ANCOVA) using each child's difference in *deebo* fixation between familiar and changed pronunciations as the dependent variable, and condition (pitch-contour change or vowel change), age in days, and their interaction as predictors. The effect of condition was significant ($t(44) = 2.11$, $p < .05$), as was the interaction of age and condition ($t(44) = 2.20$, $p < .05$). The effect of the interaction term arose because sensitivity to the vowel change was positively correlated with age ($r = 0.57$, $p < .005$), but there was essentially no correlation between age and children's sensitivity to the pitch change ($r = -0.17$, ns). Prior studies testing children's sensitivity to changes in the pronunciations of familiar words have, in most cases, failed to find a relationship between children's age and the magnitude of the effects of pronunciation changes (e.g., Swingley & Aslin, 2000; Bailey & Plunkett, 2002). In a shorter-term word-learning situation like the current one, however, older children may be better able to encode the vowel information than younger children, or may be more likely to consider the vowel change relevant to identification of the referent. An analogous ANCOVA testing effects of productive vocabulary size, rather than age, yielded no significant effects or interactions involving vocabulary size. However, a ceiling effect may have reduced the predictive power of the Communicative Development Inventory (the vocabulary checklist); over half of children (25/48) were reported to produce more than 80% of the words on the form. In analyses of variance, neither gender, the pitch contour used in teaching (rise-fall or low fall), nor the object used as the *deebo* (*red knobs* or *purple disk*) interacted with the effects of either mispronunciation.

Children's pointing and naming responses provided a useful supplement to the eyegaze data. Eyegaze, while a sensitive measure of word recognition, does not necessarily reliably index children's conscious interpretation of the utterance. For example, reduced looking to the *deebo* object upon hearing the vowel change could mean only that the changed pronunciation was not prototypical (and thus an inferior cue to the target), thus delaying or interfering with recognition. Pointing and naming responses involve discrete choices, and measure children's ultimate interpretation of the spoken words. Here, we found that children's pointing and naming responses were consistent with the results of the eye-movement analyses. **Table 3** shows pointing responses for children in each condition in response to the trained pronunciation and the changed pronunciation. Only pointing responses for children who responded in both trials are included (vowel change, n = 11; pitch change, n = 12). Children in

the vowel-change condition pointed much more often to the *deebo* (as opposed to the distracter object) when they heard "deebo" than when they heard "dahbo." Children in the pitch-change condition, by contrast, pointed more to the *deebo* than to the distracter in both trained-pronunciation trials and pitch-change trials. Pitch-change children pointed significantly more to the *deebo* object than would be expected by chance in response to the pitch change (binomial $p < .05$), and showed a trend in the same direction in response to the trained pronunciation (binomial $p = .146$). Vowel-change children pointed to the *deebo* above chance in response to the trained pronunciation (binomial $p < .001$), but not in response to the vowel change (binomial $p = 1$).

Table 3: Children's pointing responses. Points to target / number of children pointing (percentage of points to target), for each combination of condition and trial type.

	Trained pronunciation		Changed pronunciation	
Pitch-change children	9 / 12	(75%)	10 / 12	(83%)
Vowel-change children	11 / 11	(100%)	6 / 11	(55%)

Children's pointing responses to the trained pronunciation and the change could take four forms: pointing to the target for both pronunciations (abbreviated TT), pointing to the distracter for both (DD), pointing to the target for the trained pronunciation and to the distracter for the change (TD), or vice versa (DT). Children's distribution over these categories varied with mispronunciation type (X^2 (3, n = 23) = 8.57, $p < .05$), reflecting the fact that the vowel change caused children to point more to the distracter (TT = 6, **TD = 5**, DT = 0, DD = 0), while the pitch change did not (TT = 9, **TD = 0**, DT = 1, DD = 2). The pointing results indicate that children in the pitch-change condition considered both pronunciations good matches to the *deebo* object, while for children in the vowel-change condition, "dahbo" was a worse match.

In naming trials, children were asked by a recorded voice to label both the *deebo* and distracter objects. We do not have responses from many children, either because they refused to respond, they said something other than a label for the object (e.g., "Elmo"), or they did not participate in the trials. Children were not always able to correctly pronounce all the sounds of the word (e.g., they sometimes said "teenbo" or "deedo" instead of "deebo"), so we scored productions for whether the first syllable contained the /i/ vowel (as in "deebo") or the /a/ vowel (as in "dahbo"). **Table 4** displays children's use of these vowels in their labeling of the objects. All children who produced either vowel are included, whether or not they responded in both naming trials. When asked to label the *deebo* object, both groups produced more /i/ vowels (vowel-change group: 15; pitch-change group: 14) than /a/ vowels (vowel-change group: 0; pitch-change group: 1). Children were more reluctant to label the distracter object, but the data we have are consistent with the looking and pointing responses: vowel-change participants labeled the distracter object

with an /a/ vowel (5 responses) slightly more than with an /i/ vowel (1 response). In the pitch-change group, we expected children to have no name for the distracter object, and their responses are consistent with that: only two children produced /i/ vowels, and no children produced /a/ vowels. Like adults' productions, children's labeling of the *deebo* did not reflect the pitch contour they were taught; analyses of variance predicting f0 maximum and f0 mean, respectively, from taught pitch (rise-fall or low fall) showed no significant effects. (Since only seven children labeled the distracter object, we did not include *object* as a predictor, instead excluding trials where the child was labeling the distracter object.)

Table 4: Children's naming responses. Responses with the /i/ vowel / responses with either vowel (percentage using /i/ vowel).

	Viewing *deebo* object		Viewing distracter object	
Pitch-change children	15 / 15	(100%)	2 / 2	(100%)
Vowel-change children	14 / 15	(93%)	1 / 6	(17%)

Recall from Experiment 1 that ten adult judges, trained to identify rise-fall and low fall contours in our stimulus materials, categorized participants' productions of our test words. Judges' classifications of children's productions as having rise-fall or low fall contours revealed no effect of taught pitch ($F(1,26) = .47$, $p = .50$) in an analysis of variance (again, there were too few instances of distracter-labeling to include *object* as a predictor). Judges assigned the "rise-fall" classification at similar rates for productions from children who were taught the rise-fall (and were labeling the *deebo* object; mean 6.33, SE 0.49); and those taught the low fall (mean 5.94, SE 0.37). Children's failure to imitate the taught pitch contour in their own productions suggests that they did not treat the pitch pattern as relevant for reproducing the word.

To summarize, our findings from the pointing and naming trials are consistent with our eye-movement result that children treated the vowel change—but not the pitch change—as relevant. Children pointed predominantly to the *deebo* when they heard both the trained pronunciation of the word and the pitch change, but pointed roughly equally to the *deebo* and the distracter object in response to the vowel change. In their naming of the objects, both groups of children used the /i/ vowel (as in "deebo") more often than the /a/ vowel (as in "dahbo") to label the *deebo* object. Children who had heard the word "dahbo" were slightly more likely to use the /a/ vowel than the /i/ vowel to label the distracter object, while pitch-change children were not.

General Discussion

We addressed the development of interpretation of nonphonemic, but consistently realized, dimensions of the sounds of words by teaching 2.5-year-

olds and adults a novel word, "deebo," which was always produced with a consistent, salient pitch contour. In test, we changed either the pitch contour or the vowel (from /i/ to /a/). All of the 22 tokens participants heard in the teaching phase had the same vowel and the same pitch contour. If participants were storing each exemplar of this new word without selective emphasis on the native-language dimensions of contrast (as predicted by Goldinger's 1998 "extreme" model), they would be expected to treat both changes as equally relevant in word recognition. We found instead that both children and adults interpreted these changes in accordance with English phonology, reacting to the segmental change but not to the pitch change. Even 2.5-year-olds were able to override the consistency of the teaching exemplars to assign the pitch variation to the appropriate level, possibly interpreting it as phrasal intonation rather than as part of the word.

At both ages, we saw individual variation in participants' interpretations of the vowel change. Adults' and children's interpretations may have varied partly because of tension between their phonological knowledge and the pragmatics of the experiment. Participants' phonological knowledge may tell them that a change from /i/ to /a/ signals a new word. Consistent with that knowledge, 18/24 adults and 11/24 children fixated the *deebo* less than 50% percent of the time in response to "dahbo," suggesting they hypothesized that "dahbo" was a new word referring to the previously unlabeled distracter object. The pragmatics of the experiment, however, may support the alternative interpretation that "dahbo" is simply a mispronunciation of "deebo." In vowel-change trials, the *deebo* object was on the screen (with a distracter object), and participants heard a word that differed from "deebo" in only one segment. In the real world, interlocutors occasionally mispronounce words, requiring listeners to accommodate some variation. When an object is present and a speaker produces a word differing from that object's label in only one segment, this variant may well be a mispronunciation rather than a new word. Consistent with this interpretation, 6/24 adults and 13/24 children fixated the *deebo* more than 50% of the time in response to "dahbo," suggesting they hypothesized that "dahbo" was simply a mispronunciation of "deebo." The tension between English phonology and the pragmatics of the experiment may explain why many adults were inconsistent in their treatment of "dahbo" across different measures, apparently unable to settle on one interpretation or the other.

Pitting children's experience with a word against their phonology

Our finding that children do not treat all dimensions alike when representing and recognizing a new word is relevant to an ongoing debate over the abstractness of young children's—and even adults'—word representations. Psychological speech-recognition models have typically assumed that representations of words are composed of abstract phonemes (cf. Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986; and Norris, 1994), but experimental evidence suggests that adults' word representations are highly detailed. In word recognition, adults are sensitive to subphonemic information (Andruski, Blumstein, & Burton, 1994; Dahan, Magnuson, Tanenhaus, &

Hogan, 2001; McMurray, Tanenhaus, & Aslin, 2002; Salverda, Dahan, & McQueen, 2003; Salverda et al., 2007) and to characteristics of the speaker's voice (Palmieri, Goldinger, & Pisoni, 1993; Goldinger, 1996; Luce & Lyons, 1998). And they are better at recalling a list of words spoken by one talker, at one speaking rate, than a list spoken by different talkers or at different speaking rates (Nygaard, Sommers, & Pisoni, 1995). Pronunciation of words in speech also reflects knowledge of word frequencies, information not available in abstract phonological representations. For example, speakers are more likely to reduce high-frequency words in production than low-frequency words (for reviews, see Pierrehumbert, 2001; Bybee, 2001a, 2007), and words that are used frequently together are more susceptible to liaison (Bybee, 2001b).

This evidence for nonphonemic information in word representations has led to the development of exemplar theories of speech-sound learning (Jusczyk, 1993), perception (Johnson, 1997), and production (Pierrehumbert, 2002). According to exemplar theories, word and speech-sound categories emerge from the storage of many detailed exemplars of the category. In word recognition, a word form activates the stored exemplars, and that pattern of activation is used to categorize the new token. Through the incorporation of attention weights (Johnson, 1997; Jusczyk, 1993), exemplar models can selectively emphasize certain acoustic or phonetic dimensions over others. Jusczyk (1993) proposed that phonological development proceeds by fine-tuning attention weights to emphasize dimensions relevant in the native phonology. Because less-relevant dimensions are not completely deweighted, even adults show sensitivity to variation on these dimensions in implicit tasks, but their word recognition is not impaired. In contrast, young children are much more sensitive to episodic details, failing to recognize a word when the fundamental frequency (Singh, White, & Morgan, 2008), talker's voice (Houston & Jusczyk, 2000), or affect (Singh, Morgan, & White, 2004) has changed between familiarization and test. Presumably, infants are more sensitive to these dimensions in word recognition because they are still fine-tuning the weights of acoustic dimensions to match their native phonology.

Our results could be consistent with either the Jusczyk-style (1993) exemplar perspective or the abstraction view. The abstraction view is transparently consistent with our finding that English-learning children disregard lexical pitch. According to the abstraction perspective, children categorize new words as sequences of consonants and vowels, and do not store information like pitch in the lexical representation if it is not phonologically distinctive.

If viewed from the exemplar perspective, our results could be seen as evidence that 2.5-year-olds have already tuned their weights of acoustic dimensions to match the phonology of English, so that pitch information is downweighted sufficiently to not impact word recognition. This characterization is not typical of exemplar models, which were designed to account for listeners' retention of noncontrastive information (e.g., Goldinger, 1998). Still, this weaker version of exemplar models (e.g., Jusczyk, 1993) is consistent with our results.

Though simple exemplar models help account for effects of nonphonemic variation on word recognition, recall, and production, some questions remain. If people store nonphonemic detail about individual tokens of a word, how do we seem to make the phonologically normative interpretive decisions so consistently? Attention weights, which emphasize those dimensions on which sounds contrast, begin to suggest an answer, but they are an incomplete solution in two ways. As Francis and Nusbaum (2002) point out, attention weights that operate at the level of the entire dimension (following in the vein of Nosofsky's 1986 generalized context model; Jusczyk, 1993, 1994; Johnson, 1997) are insufficient. Mature interpretation of speech requires more than attending just to an entire relevant dimension (e.g., Iverson & Kuhl, 1995). Instead, it appears to require *localized* variation in attention along a dimension, in which differences near the category center are compressed, and differences near the category boundary are expanded (Goldstone, 1994; Guenther, Husain, Cohen, & Shinn-Cunningham, 1999).

More fundamentally, the demands of ordinary conversation require listeners to attend to word-level *and* utterance-level phonetic information, both of which are given in the very same signal. Rather than supposing that listeners attend to one level at the expense of the other, we argue that listeners construct a *model* of the utterance, based on linguistic knowledge, to estimate the most probable interpretation (e.g., Dahan, Drucker, & Scarborough, 2008). For a given phonetic attribute (whether it be pitch, duration, glottalization, etc.), responsibility for the value of that attribute may need to be partitioned among several factors. In the case of duration, the length of a vowel results from word-level characteristics (e.g., vowel identity, syllable position, identity of the following consonant) and utterance-level characteristics (e.g., speaking rate, location relative to prosodic boundaries), as shown in numerous phonetic studies (e.g., Klatt, 1973; van Santen, 1992). The child's task is to discover the linguistic model that aligns best with that of her community.

We have shown that 2.5-year-olds have settled on the correct linguistic model for interpretation of pitch variation at the lexical level. An important extension of the present research will be to investigate the developmental trajectory of the interpretation of pitch. This trajectory could take two forms. Children could start out disregarding pitch variation, and then learn, through exposure to their native language, to attend to pitch at the relevant levels. Alternately, children could start out treating pitch as potentially relevant (e.g., at the lexical level), and then learn to ignore it if their native language doesn't provide evidence of structure at that level. We find the latter trajectory more likely, because of evidence that children start out more open-minded about what can be a word, constraining their hypotheses over development (Namy, 2001; Namy & Waxman, 1998; Woodward & Hoyne, 1999). However, further research is required to pinpoint the precise developmental trajectory. The present work provides an important starting point by demonstrating that by 2.5 years, interpretation of lexical pitch is similar to the adult interpretation, at least under the conditions tested here.

Studies like the current one shed light on outstanding questions about the nature of the speech interpretation by providing evidence about the development of interpretation of perceptible, but nonphonemic, variation. We considered the interplay between the acoustic particulars of listeners' experience with a word and the constraints of their phonological system. From previous research, we know that adults show "echoes" of nonphonemic variation in word recognition, and infants often have even more trouble disregarding this variation. Young children often seem to struggle to interpret novel words through the lens of their native-language sound system. Yet we found that both children and adults could disregard consistency in the pitch contour of a novel word, recognizing a newly learned word even when the consistency of its pitch contour was violated. This result tells us that by 2.5 years, children do not treat all dimensions of the sounds of words equally, but instead interpret a nonphonological change in pitch contour differently from a phonological vowel change.

Acknowledgements

Thanks to members of the Swingley lab, especially Jane Park, Sara Clopton, Rebecca McCue, and Kristin Vindler Michaelson for help with participant recruitment, testing, and coding. Thanks also to members of the Institute for Research in Cognitive Science at the University of Pennsylvania, especially Delphine Dahan, John Trueswell, and Lila Gleitman, for their helpful feedback on the experiments. Finally, many thanks to the parents, children, and undergraduates who participated in the study. Funding was provided by NSF Graduate Research Fellowship and NSF IGERT Trainee Fellowship grants to C.Q., NSF grant HSD-0433567 to Delphine Dahan and D.S., and NIH grant R01-HD049681 to D.S.

References

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52,* 163-187.

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development, 17,* 1265-1282.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 9-54). Oxford: Oxford University Press.

Boersma, P., & Weenink, D. (2008). Praat: doing phonetics by computer (Version 5.0.30) [Computer program]. Retrieved from http://www.praat.org/

Bolinger, D. (1989). Intonation and its uses. *Stanford, CA: Stanford University Press.*

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast. *Language and Speech, 46*, 217-244.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new pussycat? On talking to babies and animals. *Science, 296,* p. 1435.

Bybee, J. (2001a). *Phonology and language use.* Cambridge: Cambridge University Press.

Bybee, J. (2001b). Frequency effects on French liaison. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 337-359). Amsterdam, Netherlands: John Benjamins Publishing Company.

Bybee, J. (2007). *Frequency of use and the organization of language.* Oxford: Oxford University Press.

Cutler, E. A., & Clifton, C. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Proceedings of the Tenth International Symposium on Attention and Performance* (pp. 183–196). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: adjusting the signal or the representations? *Cognition, 108,* 710-718.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16,* 507-534.

Demuth, K. (1995). The acquisition of tonal systems. In J. Archibald (Ed.), *Phonological Acquisition and Phonological Theory* (pp. 111-134). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Dietrich, C., Swingley, D., & Werker, J.F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences of the USA, 104*, 16027-16031.

Fennell, C. T. (2006). Infants of 14 months use phonetic detail in novel words embedded in naming phrases. In *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 178-189). Somerville, Massachusetts: Cascadilla Press.

Fennell, C. T., Waxman, S. R., & Weisleder, A. (2007). With referential cues, infants successfully use phonetic detail in word learning. In *Proceedings of the 31st Annual Boston University Conference on Language Development* (pp. 206-217). Somerville, Massachusetts: Cascadilla Press.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59,* i-185.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*, 181-195.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development, 60,* 1497–1510.

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papousek, U. Jurgens, & M. Papousek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 262–282). Cambridge: Cambridge University Press.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10,* 279–293.
Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the second year. *Psychological Science, 9,* 72-75.
Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 349-366.
Fry, D. (1958). Experiments in the perception of stress. *Language and Speech, 1,* 205-213.
Fulkerson, A. L., & Haaf, R. A. (2003). The influence of labels, non-labeling sounds, and source of auditory input on 9- and 15-month-olds' object categorization. *Infancy, 4,* 349–369.
Galligan, R. R. (1987). Intonation with single words: Purposive and grammatical use. *Journal of Child Language, 14*, 1-21.
Gandour, J. T. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 41-76). New York: Academic Press.
Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613-656.
Gauthier, B., Shi, R., & Xu, Y. (2007). Learning phonetic categories by tracking movements. *Cognition, 103,* 80-106.
Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22,* 1166–1183.
Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251-279.
Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*, 178-200.
Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America, 106*, 2900-2912.
Gussenhoven, C. (2004). *The phonology of tone and intonation.* Cambridge: Cambridge University Press.
Halle, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32,* 395-421.
Harrison, P. (2000). Acquiring the phonology of lexical tone in infancy. *Lingua, 110,* 581–616.
Hillenbrand, J.M., Clark, M.J., & Houde, R.A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America, 108,* 3013-3022.

Hirschberg, J., & Ward, G. (1992). The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics, 20,* 241-251.
Hollich, G. (2005). Supercoder: A program for coding preferential looking (Version 1.5). [Computer Software]. West Lafayette: Purdue University.
Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology, 26,* 1570–1582.
Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language, 27,* 3-42.
Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97,* 553-562.
Johnson, K. (1997) Speech perception without speaker normalization: An exemplar model. In Johnson & Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145-165). San Diego: Academic Press.
Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA Model of how speech perception develops. *Journal of Phonetics, 21,* 3–28.
Jusczyk, P. W. (1994). Infant speech perception and the development of the mental lexicon. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 227-270). Cambridge, MA: The MIT Press.
Katz, G. S., Cohn, J. F., & Moore, C. A. (1996). A combination of vocal f0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. *Child Development, 67,* 205–217.
Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mothers' speech: Adjustments for age and sex in the first year. *Infancy, 4,* 85-110.
Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development, 24,* 372-392.
Klatt, D.H. (1973). Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America, 54,* 1102-1104.
Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period." *Language Learning and Development, 1,* 237–264.
Ladd, D. R. (1996). *Intonational Phonology.* Cambridge: Cambridge University Press.
Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America, 32*, 451-454.
Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology, 43,* 912-917.

Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory & Cognition, 26,* 708-715.
Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language, 57,* 252-272.
Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121-157.
Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10,* 241-265.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.
McMurray, B., Tanenhaus, M., & Aslin, R. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*, B33-B42.
Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In Metsala, J. L. & Ehri, L. C. (Eds.), *Word recognition in beginning literacy* (pp. 89-120). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
Moore, D. S., Spence, M. J., & Katz, G. S. (1997). Six-month-olds' categorization of natural infant-directed utterances. *Developmental Psychology, 33,* 980–989.
Namy, L. L. (2001). What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. *Infancy, 2,* 73–86.
Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. *Child Development*, *69,* 295–308.
Narayan, C. R. (2006). Acoustic-perceptual salience and developmental speech perception. Dissertation, University of Michigan.
Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition, 98,* 13–30.
Norris, D. (1994). Attention, similarity, and the identification-categorization relationship. *Cognition, 52,* 189-234
Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.
Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60,* 355-376
Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics, 57,* 989-1001.
Ota, M. (2003). The development of lexical pitch accent systems: An autosegmental analysis. *Canadian Journal of Linguistics, 48,* 357-383.
Palmieri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 19, 309–328.
Papousek, M., & Hwang, S. C. (1991). Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics, 12,* 481-504.
Papousek, M., Papousek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development, 14,* 415-440.
Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language, 80,* 384–402.
Pierrehumbert, J. (1980). The Phonology and Phonetics of English Intonation. Dissertation, MIT.
Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 337-359). Amsterdam, Netherlands: John Benjamins Publishing Company.
Pierrehumbert, J. B. (2002). Word-specific phonetics. In *Laboratory Phonology VII*, (pp. 101-139). Berlin: Mouton de Gruyter.
Pierrehumbert, J.B. (2006). The next toolkit. *Journal of Phonetics, 34*, 516-530.
Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 421–435.
Quam, C., Swingley, D., & Park, J. (2009). Developmental change in preschoolers' sensitivity to pitch as a cue to the speaker's emotions. *Society for Research in Child Development 2009 Biennial Meeting*, Denver, CO.
Quam, C., Yuan, J., & Swingley, D. (2008). Relating intonational pragmatics to the pitch realizations of highly frequent words in English speech to infants. *In* B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 217-222). Austin, TX: Cognitive Science Society.
Roberts, K. (1995). Categorical responding in 15-month-olds: Influence of the noun-category bias and the covariation between visual fixation and auditory input. *Cognitive Development, 10,* 21–41.
Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90,* 51-89.
Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically-modulated sub-phonemic variations on lexical competition. *Cognition, 105,* 466-476.
Singh, L., Morgan, J. L, & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language, 51,* 173-189.
Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early

spoken word recognition. *Language Learning and Development, 4*, 157-178.
Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*, 381–382.
Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language, 10,* 1-15.
Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language, 29*, 251-274.
Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology, 43*, 454-464.
Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language, 60,* 252-269.
Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech, 46*, 265-294.
Swingley, D., & Aslin, R. N. (2000). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science, 13,* 480–484.
Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology, 54*, 99-132.
Thiessen, E.D. (2007). The effect of distributional information on children's use of phonemic contrasts. Journal of Memory and Language, 56, 16-34.
Trehub, S.E., & Hannon, E.E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition, 100*, 73-99.
Turk, A.E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics, 28*, 397-440.
van Santen, J.P.H. (1992). Contextual effects on vowel duration. *Speech Communication, 11*, 513-546.
Vihman, M. (1996). *Phonological development: The origins of language in the child.* Malden, MA: Blackwell Publishing.
Vihman, M., & Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics, 45*, 683-725.
Ward, G., & Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language, 61,* 747-776.
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,* 49-63.
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,* 1–30.
Werker, J.F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1,* 197-234.
White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*, 114-132.

Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development*, *70,* 65–77.
Yoshida, K., Fennell, C., Swingley, D., & Werker, J.F. (2009). 14-month-old infants learn similar-sounding words. *Developmental Science*, *12*, 412-418.

Appendix 1: Acoustics of the teaching and test words. Mean and standard deviation of duration (in seconds), pitch maximum (in Hz), and pitch mean (in Hz) for each teaching and test word.

Word	Pitch	Phase	Duration (SD)	Pitch Max (SD)	Pitch Mean (SD)
Deebo	Rise-fall	Teaching	1.245 (0.076)	587.7 (56.2)	284.8 (15.5)
Deebo	Low fall	Teaching	1.370 (0.121)	264.1 (11.7)	215.1 (6.8)
Deebo	Rise-fall	Test	1.321 (0.038)	673.4 (26.3)	300.1 (2.7)
Deebo	Low fall	Test	1.292 (0.077)	283.9 (2.9)	232.7 (9.1)
Dahbo	Rise-fall	Test	1.326 (0.044)	757.4 (1.2)	295.1 (7.1)
Dahbo	Low fall	Test	1.283 (0.007)	274.3 (16.1)	221.3 (5.3)