Perceptual Beginnings to Language Acquisition

Janet F. Werker

University of British Columbia

Author Note

Preparation of this paper, and the research from my laboratory described therein, was supported by grants from SSHRC and NSERC, and by my Canada Research Chair and Canadian Institutes for Advanced Research Fellowship. Special thanks to Savannah Nijeboer for editing and for creating the figures, and to Dr. Padmapriya Kandhadai for editing advice. And finally, I want to express my gratitude to Martha Crago for her remarkable leadership editing this journal, and for inviting me to write the target article for the final issue under her editorship.

Correspondence concerning this article should be addressed to Janet F. Werker, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, Canada, V6T 1Z4. Email: jwerker@psych.ubc.ca

Abstract

In this article, I present a selective review of research on speech perception development and its relation to reference, word learning, and other aspects of language acquisition, focusing on the empirical and theoretical contributions that have come from my laboratory over the years. Discussed are the biases infants have at birth for processing speech, the mechanisms by which universal speech perception becomes attuned to the properties of the native language, and the extent to which changing speech perception sensitivities contribute to language learning. These issues are reviewed from the perspective of both monolingual and bilingual learning infants. Two foci will distinguish this from my previous reviews: first and foremost is the extent to which contrastive meaning and referential intent are not just shaped by, but in fact also shape, changing speech perception sensitivities, and second is the extent to which infant speech perception is multisensory and its implications for both theory and methodology.

Keywords: language acquisition, speech perception, phonetic development, multisensory, infancy

Introduction

The regularity and near universality in the steps children take from very early in life toward the acquisition of the native language show how deeply embedded language is in our biology. Yet, the vast ways in which experience and learning contribute to and shape acquisition reveal language to be, simultaneously, perhaps our most experientially sensitive and responsive cognitive attribute. Perception, cognition, learning, social understanding, and culture all come together in contributing to the emergence of this quintessential human capability.

While prepared by evolution to acquire any language, it is of course the native language we must ultimately acquire; as such, we acquire language in interaction with other members of our community. The first window infants have into that acquisition comes via their perceptual systems – in listening to, watching, and perhaps imitating language long before they know any words. My research program traces the perceptual foundations of language acquisition, starting in infancy and even prenatal development. A question that has driven this work almost from its inception is an attempt to understand when and how experience with the native language changes perceptual sensitivities, and just how these changing speech perception sensitivities set the stage for, and intersect with, successful language learning. While I have always argued that there are bidirectional influences between perceptual development and the achievement of various milestones in language development (see Figure 1), the emphasis in much of my writing has been on how attunement to the perceptual properties of the native language guides and bootstraps acquisition (see Wojcik, de la Cruz Pavia, & Werker, in press, for the most recent such focus). The current review will highlight as well the ways in which language function and use also drive perceptual attunement.

To this end, I will first present descriptive findings on the character of early speech perception, the neural organization that gets us started, and the timing of how speech perception changes across the first year of life. Both monolingual and bilingual infants will be considered. I will then describe research in which we have attempted to identify the ‘mechanisms of change’ – learning or otherwise – that underlie the movement from language general to language specific phonetic perception, and the relation between speech perception and language understanding. In the process I will also outline broader theoretical frameworks we have proposed over the years to account for these changes, while also providing the reader with a sense of how these frameworks have been modified with the accrual of new data. Woven into the review will be the question of whether there are ‘critical’ or ‘sensitive’ periods in development during which input is most effective and why, as well as the question of whether speech perception is best characterized and studied as an acoustic-only phenomenon or whether to fully understand the development of speech perception we need to focus on speech as a multisensory phenomenon. The theme of the current review is to consider the extent to which speech perception development not only sets the stage for, but is also influenced by, the acquisition of meaning.

Perceptual Foundations at Birth

By the time they are born, human infants are already well prepared for language acquisition. At birth, babies show a preference for speech over other types of complex sounds (Vouloumanos & Werker, 2007), and use distinct brain systems when processing speech sounds (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; May, Gervain, Carreiras, & Werker, in press; May, Byers-Heinlein, Gervain, & Werker, 2011; Minagawa-Kawai et al., 2011; Peña et al., 2003), with some support for greater involvement of the left hemisphere (Dehaene-Lambertz, 2000; Peña et al., 2003; Perani et al., 2011). The boundaries on infants’ perceptual preference for language are revealing: they extend beyond human speech to other primate vocalizations (Vouloumanos, Hauser, Werker, & Martin, 2010), but not to non-communicative vocalizations such as coughs, yawns, or sneezes (Shultz, Vouloumanos, Bennett, & Pelphy, 2014), leading to the suggestion that these areas of the human brain are broadly specialized for processing “communicative” sounds. However, in our work, neural activation in these same areas is also seen for natural sounds (e.g. all kinds of moving water) that have a similar fractal structure to speech (Gervain, Werker, Black, & Geffen, 2016), but not to either forward or backward Silbo Gomero, a whistled language that is used by shepherds in the Canary Islands (May et al., in press). Silbo Gomero, a ‘surrogate’ language that shares properties with its Spanish base, is never acquired as a first language – but once acquired, it is decidedly communicative, whereas water is not. Thus, appealing as the distinction is, it is not merely the communicative versus non-communicative nature of speech that underlies its neural specialization, but something as well about the match between its properties and the resolving capabilities of the related brain structures. By three months of age, English-learning infants no longer include other primate calls in the class of vocalizations they prefer (Shultz et al., 2014; Vouloumanos et al., 2010), showing a rapid increase in the precision of the type of vocal communication to which they attend.

Also by birth, infants show evidence of experiential influences on both speech perception and its neural foundations. Neonates prefer their mother’s voice (DeCasper & Fifer, 1980) and the language heard in utero (Mehler et al., 1988; Moon, Cooper, & Fifer, 1993), with neonates from a bilingual environment demonstrating a preference for both of their native languages (Byers-Heinlein, Burns, & Werker, 2010). Moreover, although the temporal and inferior frontal areas in the neonate brain are more activated to forward versus backward speech, the differential activation to forward speech is more pronounced for (and in some comparisons, only seen for) the language heard in utero (May, Byers-Heinlein, Gervain, & Werker, 2011; May et al., in press; Minagawa-Kawai, et al., 2011). Thus, from the earliest ages tested, it is apparent that babies at birth are both prepared to learn language – any language – but that they have already begun the journey toward becoming a specialized listener to, and learner of, their native language.

Newborn infants have a number of other perceptual sensitivities that prepare them for language acquisition in general. They discriminate languages from different rhythmical classes (Mehler et al., 1988); show a preference for, and specialized processing of, well-formed syllables (Gómez et al., 2014); and perceptually discriminate the two major grammatical classes of words – content versus function words (Shi, Werker, & Morgan, 1999). All of these abilities become further attuned to the properties of the native language over the first months of life. For example, infants become better able to discriminate their own language from another language from within the same rhythmical class by 4 months of age if bilingual or 5 months of age if monolingual (Nazzi, Jusczyk, & Johnson, 2000); develop a preference for the phonotactic regularities (Jusczyk, Luce, & Charles-Luce, 1994) and stress patterns (Jusczyk, Cutler, & Redanz, 1993) used in their native language by 8-9 months of age; and become experts by 8-9 months as well at precisely which grammatical words occur in their native language (Shi, Werker, & Cutler, 2006) and where (e.g. phrase initial or phrase final; Gervain, Nespor, Mazuka, Horie, & Mehler, 2008) – and further, use this knowledge to segment words (Shi & Lepage, 2008) and phrases (Gervain et al., 2008; Gervain & Werker, 2013) in a language-specific manner. By the time they are 18 months old and beginning to learn words, infants selectively associate with objects only those forms that sound like words in their native language (e.g. MacKenzie, Graham, & Curtin, 2011; May & Werker, 2014).

From early in life, listening to speech selectively enhances – and is enhanced by – communicative and referential functions, with attunement in this process as well. In the presence of speech sounds, but not tones, infants as young as 3 months detect the category structure of a set of pictures (e.g. detect the category dinosaurs; Ferry, Hespos, & Waxman, 2013). At this young age, rhesus monkey calls also facilitate object categorization, but no longer do so by 4 months (Perszyk & Waxman, 2016). Similarly, at 4-months infants infants orient more rapidly to visual objects when they hear speech than when they hear non-speech or silence, suggesting that something about speech triggers a readiness for reference (Marno, et al, 2015). By 5 months, infants look longer at pictures of human faces when hearing speech and at monkey faces when hearing primate calls, but do not look longer at duck faces when hearing duck sounds (Vouloumanos, Druhen, Hauser, Huizink, 2009), which indicates they have some expectation of the source of potentially meaningful vocalizations. By 12 months, they show expectations that speech, but not non-speech, can communicate unobservable intentions (Vouloumanos, Onishi, & Pogue, 2012) and potential ways to interact with objects (Martin, Onishi, & Vouloumanos, 2012); they also show an expectation for information from speakers of their native language (Begus, Gliga, & Southgate, 2016). Thus, in addition to the prefence infants show for speech from early in life, they also seem more readily prepared to link speech – over other kinds of sounds – to meaning and communicative intent.

An important caveat to much of the above-mentioned work on the unique role speech sounds play in facilitating categorization and word learning is that, in the presence of referential cues, infants will treat non-speech or non-native speech as functionally linguistic. For example, if first familiarized to a video of two people conversing, but tones rather than words appear to come from the mouth of one of the two interlocuters, then infants will successfully detect the category structure of a set of objects when presented with tones in a categorization task soon thereafter (Ferguson & Waxman, 2016). As another example, if first familiarized with three known word-object pairings, which serve to indicate that the current task is one of object labeling, 18-month-old infants will better link non-native word forms to objects (May & Werker, 2014). Thus, while the human brain may be specialized to detect speech and to use speech most naturally for categorization and labeling, if provided with evidence that the context of a sound presentation is the exchange of meaning, infants will accept other kinds of sounds as speech as well. This is another example of the two-way interaction between form and meaning: while learning about the form provides infants with candidate sounds and/or words for learning meanings, the search for meaning also influences the way that infants process and treat different kinds of sounds. Thus, again, higher-order aspects of language use can "reach down" into perception.

Perceptual Attunement to Phonetic Categories

Since the early 1970s, we have known that infants can discriminate syllables differing in single phonetic features – whether they be in the initial consonant; in place, manner, or voicing; or in the height or roundedness of the vowel (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; see Kuhl, 2004 and Werker & Curtin, 2005, for reviews). These findings have been strengthened by more recent work using ERP, MEG, and fNIRS measures for phonetic discrimination (see Gervain & Mehler, 2010; Kuhl, 2010 for reviews). Discrimination is most robust when acoustic-phonetic variations cross, rather than come from within, known natural language phoneme distinctions (Eimas et al., 1971; Werker & Lalonde, 1988) and involves the same left hemisphere temporal areas activated in adult phonetic discrimination (Dehaene-Lambertz & Gliga, 2004).

It has also been known since the 1970s that, in infants younger than 6 months of age, perceptual sensitivity extends to non-native phonetic contrasts (e.g. Werker, Gilbert, Humphrey, & Tees, 1981; Streeter, 1976), including acoustically-similar non-native distinctions that adults who are not speakers of that language have difficulty discriminating (see Werker, 1989 for a review). This pattern of broad-based initial sensitivities, which supports discrimination of both native and non-native consonant distinctions in very young infants, changes across the first year of life such that infants show a decline in discrimination of non-native phonetic distinctions (e.g. Werker & Tees, 1984) and improved discrimination of native ones (e.g. Kuhl et al., 2006; Narayan, Werker, & Beddor, 2010) . Referred to as perceptual attunement or perceptual narrowing, this pattern of decline in non-native and improvement in native discrimination has been shown as well for the perception of vowel distinctions (e.g. Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994), lexical tone differences (e.g. Mattock & Burnham, 2006; Yeung, Chen, & Werker, 2013), and even categorical perception of hand sign (Baker, Golinkoff, & Petitto, 2006; Palmer, Fais, Golinkoff, & Werker, 2012).

Studies with infants growing up bilingual show a similar pattern of change, but with (perhaps) some variation. While some studies indicate that during the period of perceptual attunement, bilingual infants go through a stage where they discriminate best the phoneme distinctions that are most common across their two languages and collapse less frequent ones (Bosch & Sebastian-Galles, 2003), other studies (Sundara, Polka, & Molnar, 2008) – and even some with the same population of bilingual infants tested by Bosch and Sebastian-Galles (2003) – indicate that if more sensitive testing paradigms are used, the developmental trajectory is the same in bilingual and monolingual infants (Albareda-Castellot, Pons, & Sebastian-Galles, 2011). While there are some differences reported across the first year of life, the data are nevertheless consistent in indicating that bilingual infants discriminate the phoneme contrasts of each of their native languages by the end of the first year of life (e.g. Burns, Yoshida, Hill, & Werker, 2007; Bosch & Sebastian-Galles, 2003).

Perceptual attunement – or narrowing of phonetic discrimination from broad-based to language-specific – is the most common pattern of reorganization seen across the first year of life in phonetic discrimination (see Maurer & Werker, 2014). But there are other patterns of change as well. Discrimination of some distinctions, such as the contrast between /ng/ and /n/ in initial position, is not seen in young infants, and indeed does not emerge until later in infancy, following several months of listening experience with that distinction in one's native language (Narayan, Werker, & Beddor, 2010; see also Kuhl et al., 2006 and Sato, Sogabe, & Mazuka, 2010). Similarly, some non-native and never (or seldom) heard distinctions remain discriminable across the lifetime, failing to show a pattern of narrowing or decline. The best known example here is Zulu clicks, which adult non-click speakers can continue to discriminate across the lifetime (Best, McRoberts, & Sithole, 1997). A more recent example is rising versus falling (T2-T4) Mandarin tonal distinctions. Dutch-learning infants show a decline in discrimination of this distinction between 4-7 and 8-11 months of age, but do not fully stop discriminating, and then show an increase again after that age (Liu & Kager, 2014). Even in behavioral studies, earlier influences of the ambient language are apparent, for example on the internal structure of native phonetic vowel categories (Kuhl, 1991), in preference for syllables containing native lexical tones among tone-language learning infants (see Yeung, Chen, & Werker, 2013), and even for vowels heard in utero (Moon, Lagerkrantz & Kuhl, 2013). As well, the process of decline in discrimination of non-native phonetic distinctions may appear earlier for vowels than for consonants (Polka & Werker, 1994)

Mechanisms of Attunement

Timing

A continuing question in the literature, and one we have attempted to address as well, is what mechanisms underlie the establishment of native phonological categories. At the neurophysiological level, our initial findings (e.g. Werker & Tees, 1984) were taken as potential evidence for pruning of exuberant connectivity, with initial preparation for discrimination of both native and non-native distinctions and subsequent refinement to just native ones. While this may be the case at some level of analysis, in the years following, animal model work has yielded increasingly sophisticated information about the number of different ways circuits can be kept open and closed (see Werker & Tees, 2005 for a review of some of these). One line of work in my lab – which was reviewed in-depth in Werker and Hensch (2015) – examines the question of whether there are critical periods in speech perception development. Working within a model developed by Hensch (e.g. Hensch, 2005; Takesian & Hensch, 2013) delineating the molecular, neurophysiological, and experiential events that support the onset and closing of periods of plasticity in the brain, we have investigated the effect of different types of exposure and experience on the timing of speech perception development. This work indicates that while premature birth does not accelerate the timing of phonetic attunement (it corresponds to gestational rather than chronological age; see Peña, Werker, & Dehaene, 2012), exposure to certain classes of drugs may accelerate the onset of plasticity while experience with (untreated) maternal depression may delay it (Weikum, Oberlander, Hensch, & Werker, 2012). Bilingual babies also seem to show openness longer (Petitto et al., 2012; Pi Casaus, Sebastian-Galles, Werker, & Bonatti, 2015; see also Sebastian-Galles, Albareda, Weikum, & Werker, 2012 for visual language discrimination) but it is not known whether this reflects a timing delay at the circuit level, presumably caused by less exposure to/experience with each language, or an attentional advantage from simultaneously acquiring two languages. Given that this work was reviewed fairly comprehensively in Werker and Hensch (2015), it will not be the focus of the current paper.

Process

Another question we and others have addressed is what is the process by which infants become attuned to the phonetic categories that are used in their native languages? Do they learn them through simple passive learning? Through mapping sound on to meaning? And what types of interactional contexts best support such learning?

Prior to the explosion of research showing sophisticated perceptual sensitivities to phonetic distinctions in young infants, it was assumed that a phonological system was built through establishing a lexicon. Here it was assumed that as children accrued a sufficient vocabulary, they would begin to establish lexical representations in which the words were highly similar phonologically, and that, through "contrast", they would establish the repertoire of native phonemes (Trubetskoy, 1969). The necessary lexicon was thought to include, for example, actual minimal pair words such as 'goat' and 'coat' for an English-learning infant (Barton, 1980), but also sufficient items that differed in distributed, individual features across words to eventually point to a matrix. However, with the advent of studies showing that infants begin life discriminating speech according to phonetic category boundaries, and that phonetic perceptual development includes narrowing and sharpening as well as building, this view began to fall into disfavor (see Brown & Matthews, 1997). Although many of us felt it important to make a distinction between phonetic category learning and phonological contrast (Pater, Stager, & Werker, 2004; Werker & Logan, 1985), the fact that native phonetic categories could be established prior to the establishment of a lexicon was deemed to require a new explanation. Thus, with the advent of findings from the infant speech perception field, researchers began to look for explanations that could occur prior to the establishment of a lexicon. Peter Jusczyk, in his word recognition and phonetic structure acquisition (WRAPSA) model, discussed the shrinking and stretching of perceptual space, in loose connection with familiarization with word forms (Jusczyk, 1993). Patricia Kuhl proposed a ‘magnet effect’, whereby highly familiar items – akin to category prototypes – would serve to attract similar variants and reduce discrimination among them (see Kuhl,1993; Kuhl et al., 2008). Catherine Best put forward similar notions, but based on ‘perceptual assimilation’ to similar articulatory gestures rather than acoustic distinctiveness (Best, 1994). But while useful descriptive frameworks for the observed patterns of results, these frameworks lacked empirical operationalization and prediction.

At around this time, Saffran, Aslin, and Newport (1996) published their seminal paper showing that infants are superb at tracking statistics in the input – in their case, transitional probabilities between syllables – as a means of pulling out words. Building on work she had initially completed with adults (Maye & Gerken, 2001), when she joined my lab, Jessica Maye asked whether infants could similarly use statistical regularities in the speech they hear – even without knowing how those speech sounds related to meaning – to pull out the phonetic categories of their native language (Maye, Werker, & Gerken, 2002). We reasoned that infants growing up in a language that distinguished a particular contrast would likely hear instances of those phones in a bimodal distribution, whereas infants growing up in a language that did not distinguish a contrast would likely hear those phones in something closer to a unimodal distribution (see Figure 2 a and b for an example). We modeled this in a ‘distributional learning’ paradigm wherein infants aged 6-8 months were familiarized to a d-t continuum that had acoustic cues unlike those used in English, and then tested their ability to discriminate the two categories. While all the infants heard all eight steps along this continuum, half were familiarized to a bimodal distribution in which there were more exemplars near the end points (in Figure 2b, steps 2 and 7), and half were familiarized to a unimodal distribution in which there were more exemplars of the two mid points (in Figure 2b, steps 4 and 5). Following familiarization, both groups were tested with stimuli that both groups had experienced with the same frequency during the familiarization phase (in Figure 2b, steps 1 and 8). Only infants familiarized to the bimodal distribution succeeded. In subsequent work, Maye, Weiss, and Aslin (2008) showed that not only could infants maintain or learn to collapse a phonetic category on the basis of distributional frequency, but they could also generalize what they had learned to a new phonetic contrast in which the same phonetic dimension (voice-onset time) was the critical distinguishing cue. Enhancement of initial sensitivity via distributional learning has also been shown for (a retroflex but not an alveolar-palatal) fricative discrimination in young infants (Cristia, McGuire, Seidl, & Francis, 2011), and for an acoustically-impoverished ba/da contrast, but here only with the inclusion as well of visual cues (Teinonen, Aslin, Alku, & Csibra, 2008).

The studies reviewed above provided support for the hypothesis that prior to establishing a lexicon to help with contrast, infants can track statistical regularities in the input (McMurray, Aslin, & Toscano, 2009). By the time they are 9-10 months of age and have begun to attune to the phonetic categories used in their native language, passive statistical learning is less effective. By this age, four rather than two minutes of prefamiliarization is required to lead to changes in category structure (Yoshida, Pons, Maye, & Werker, 2010), and even then, the results are not as robust or as generalizable as those reported with younger infants. By 18 months of age (Maye, personal communication) and in adulthood (Maye & Gerken, 2001; Hayes-Harb, 2007), while distributional learning can modify perceptual discrimination, listeners still fail overall to use these cues to discriminate non-native phonetic distinctions. By 9-10 months, learning appears more robust when there is contingent social interaction – even without an explicitly-presented bimodal versus unimodal massed presentation (Kuhl, Tsao, & Liu, 2003; Meltzoff & Kuhl, 2016) – and/or when the phonetic category bifurcation or collapse co-occurs with the presentation of distinct objects (Yeung & Werker, 2009). Still working within a statistical learning frame, we first conceptualized this latter finding as a form of ‘acquired distinctiveness’. Building on the animal perceptual learning literature (see Kluender, Lotto, Holt, & Bloedel, 1998; Lawrence, 1949), we suggested that the pairing of two perceptually similar syllables with two more obviously distinct objects helped English-learning infants attend to, and pull apart, the two non-native (Hindi) syllables and thus more successfully discriminate them (see Figure 2c). Subsequent work reviewed below, however, suggested that pairing two non-native syllables with two different objects may actually involve more abstract phonological knowledge guiding perceptual learning.

Two subsequent studies involved testing 9-10 month old infants on distinctions that are not used for lexical contrast (to distinguish two words) in their native language, but that none-the-less occur as paralinguistic cues. In the first study, we tested English-learning infants on their sensitivity to a rising versus flat tone difference as used in Mandarin and Cantonese (Yeung, Chen, & Werker, 2014). As noted earlier, by 9-10 months, English-learning infants have difficulty discriminating this difference (Mattock & Burnham, 2006; Mattock, Molnar, Polka & Burnham, 2008; Yeung, Chen, & Werker, 2014). In English, differences in tone are used functionally – for example, to signal a question versus a statement, to relinquish the floor in a conversation, for emphasis, and so on – but are not used to distinguish two words as they are in tone languages such as Mandarin. Thus, English-learning infants may have actively learned to ignore tone differences as signals of lexical contrast. And indeed, with the two syllables differing only in rising versus level tone, simple pairing of the two syllables with two different objects was not sufficient to improve discrimination. Instead, the infants had to first be given some evidence that the two different syllables referred to objects. Specifically, they had to first be primed with three known word-object pairings (e.g. hearing the word dog when shown a picture of a dog) and then presented with the new word-object pairings. Only with the addition of these referential trials was there an improvement in tone discrimination (see Fennell & Waxman, 2010 for the genesis of the referential manipulation).

In the second study, Yeung and Nazzi (2014) used a similar paradigm and extended these findings to lexical stress. In French, stress is not used phonemically to contrast the meaning of individual words, and French infants aged 9-10 months do not discriminate stress differences as well as do infants from languages in which it is used to distinguish meaning (Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009; Skoruppa et al., 2009). Like tone in English, stress is used only pragmatically in French, and so, using the same logic as before, the infants may have already learned to ignore stress on words as indicative of contrast. Here again the addition of cues signaling a referential context enabled infants to learn from the consistent word-object pairings and to improve in discrimination and lexical use of stress (Yeung & Nazzi, 2014). Importantly, in this work, the referential manipulation included both presentation of known word-object pairs before the pairing of new words with unknown objects (as in Yeung, et al, 2014) as well as ostensive pointing cues (see Wu, Gopnik, Richardson, & Kirkham, 2011), thus employing both reference and social interaction.

The Link Between Changing Perceptual Sensitivities and Later Language

An increasing body of work shows a link between changing speech perception sensitivities and their subsequent use in language processing and language learning. Word recognition tasks indicate that infants can recognize known words and distinguish them from minimal pair distinctions as used in their native language by as young as 10-12 months of age (e.g. Mani & Plunkett, 2010). Word learning tasks indicate that infants have difficulty using language-specific phonological knowledge to guide word learning in the initial stages of word learning (e.g. Stager & Werker, 1997; Swingley & Aslin, 2007), but that by 18 months of age, phonological knowledge does guide word learning across a variety of tasks (Swingley, 2007; Swingley & Aslin, 2000; Werker, Fennell, Corcoran, & Stager, 2002). One task we used to assess minimal pair word learning is the 'Switch' task. In this procedure, infants are habituated to two word-object pairings: object A with nonsense word A, and object B with nonsense word B. Following habituation, they are tested on whether they have learned the word by being shown, in sequence, a word-object pairing that matches their experience in the habituation phase and then a word-object pairing that is mismatched (e.g. object A with nonsense word B). Longer looking to the 'switch' trial indicates that infants have learned the link. Our initial work indicated that while 14-month-old infants can perform reliabily in this task when the nonsense syllables are phonologically distinct items such as "lif" and "neem", they have difficulty with minimal pair items such as "bih" and "dih" (Stager & Werker, 1997; Werker, Cohen, Lloyd, Stager, & Casasola, 1998), even though they can distinguish already known words (for a review, see Swingley, 2009).

These early studies indicated that infants fail to reliably learn minimal pair distinctions in the switch task until 18-20 months of age (Werker, Fennell, Corcoran, & Stager, 2002), and can continue to have difficulty until an even later age if they are growing up bilingual (Fennell, Byers-Heinlein, & Werker, 2007). From a number of studies that replicated and then extended these results, we now know that if the computational demands of the task are simplified, infants can learn minimal pair words at a younger age. For example, prefamiliarizing monolingual infants with either the word forms (Fennell & Werker, 2003) or objects (Fennell, 2012) allows them to succeed at 14 months, as does testing them in a two-choice, side by side looking task where they can compare the two objects and look to the one that most closely matches the word with which it was previously paired (Yoshida, Fennell, Swingley, & Werker, 2009). They can also succeed at this younger age if the minimally-different syllables are presented in distinct syllabic contexts (Thiessen, 2007), presented in stressed syllables (Archer, Ference, & Curtin, 2014), or if multiple speakers produce the words during the familiarization phase (Rost & McMurray, 2009). Similarly, we now know that bilingual-learning infants can pass these tasks as early as monolingual-learning infants if the nonsense words match better the phonetic form they are typically hearing in their day-to-day language environment (Fennell & Byers-Heinlein, 2014; Mattock, Polka, Rvachew, & Krehm, 2010). That is, if they are typically hearing speech from a bilingual adult, bilingual infants perform best when the nonsense words are pronounced by a bilingual, whereas monolingual infants perform best when the nonsense words are pronounced by a monolingual speaker of their language (and show difficulty if the words are produced by a bilingual speaker). By 18-24 months, monolingual and bilingual infants appear to perform similarly (Singh, Hui, Chan, & Golinkoff, 2014).

In attempting to understand the difficulty that 14- but not 18-month-old infants have in using native phonetic categories to guide word learning when they are not given extra support by contextually relevant cues, a number of us proposed frameworks for understanding the link between speech perception development and phonological/lexical acquisition. In our ‘PRIMIR’ model (Curtin & Werker, 2007; Werker & Curtin, 2005), Suzanne Curtin and I built on the evidence showing that while young infants are sensitive to phonetic contrast, they are also sensitive to a number of other acoustic-phonetic cues in speech. They discriminate pitch (Nazzi, Floccia, & Bertoncini, 1998), stress (Sansavini, Bertoncini, & Giovanelli, 1997), and individual voices (as evident, for example, in their preference for the mother's voice; DeCasper & Fifer, 1980). All of these aspects of the speech signal are important as all of them are informative: they signal differences in emphasis, carry information about speaker identity, and so on. In many languages, however – including English – these cues are strictly paralinguistic and do not signal a difference in meaning between two words. In PRIMIR we suggest that 14-month-old infants might weigh these paralinguistic cues as heavily as they do lexically contrastive information, and thus can only attend specifically to the phonetic differences if they are highlighted and/or if the computational demands of the word learning situation are minimized (see Werker & Fennell, 2004, for an earlier version of this as a 'computational resources' explanation). We reasoned that by 18 months, when they have a sufficiently large vocabulary, 'phonological' categories would likely be prioritized, and so when encountering a word and an object together, infants would treat the word as a label and hence attend to the phonetic features rather than to other acoustic cues. This in turn would enable them to more easily use phonological contrast to guide their word learning. Proposals focusing on a shift to phonological contrast have been put forward by several other authors as well, including those coming from a lexical perspective (Heitner, 2004; Swingley, 2009) to a more distributional perspective (Beckman & Edwards, 2000; Thiessen, Kronstein, & Hufnagle, 2013).

In further support, working jointly with Swingley, we found that at 18 months of age it is indeed phonological categories rather than just language-specific phonetic categories that guide word learning in the switch task. We compared English and Dutch learning infants on their ability to link words differing in vowel color (/tam/ vs. /tem/) and words differing in short versus long vowels (/tam/ vs. /taam/) to two different objects. Importantly, the vowel color distinction is phonemic in both English and Dutch, whereas the vowel duration distinction is phonemic in Dutch but not in English – yet English infants this age can discriminate vowel length differences in a nonsense syllable discrimination task (Mugitani, Pons, Fais, Werker, & Amano, 2009). As predicted, both the English and Dutch infants succeeded at learning the word-object association when the words differed in vowel color but only the Dutch infants succeeded when the words differed in vowel duration (Dietrich, Swingley & Werker, 2007).

Building from these findings, Fennel and Waxman (2010) proposed that if 14-month-old infants were put in a situation wherein it was clear that the word-object pairings were about reference, then even at this younger age they should be able to use minimal pair phonetic differences to guide word learning. To test this question, infant performance following two ‘referential’ manipulations was compared to their performance in the standard task. In one referential manipulation, prior to habituation to the two minimally-different words pair with two different objects, infants were first shown three known word-object pairings in an attempt to ‘inform’ the infants that this task was about reference. This manipulation was sufficient to boost infants’ performance, enabling them to learn the two phonetically similar words. A second referential manipulation wherein infants were presented the test words in very short sentences rather than in isolation also boosted performance (Fennell & Waxman, 2010). In our lab we found similar results: when the referent was indicated by a pointing interlocutor, 14-month-old infants also showed better performance (Fais et al., 2012). Thus, we have again an example of the bidirectional relationship between phonetic and lexical development.

The minimal pair word learning task requires infants to use native phonological categories to guide word learning. As such, we reasoned it might provide a sensitive diagnostic for children at-risk for a phonological delay. Thus, in two longitudinal studies we investigated whether performance in the standard minimal pair Switch word-learning task in the toddler years would predict later phonological knowledge and pre-literacy. In the first study, with only 16 infants, we found a very strong relationship (Bernhardt, Kemp, & Werker, 2007). However, in a second, more tightly-controlled longitudinal study with a larger sample – including children from more diverse SES backgrounds than in our usual studies – this predictive relation did not replicate, with the only hint of a relation being found among the poorest-performing toddlers who also had low vocabulary (Kemp et al., 2017). Although this raises the possibility that the task could help identify those slow talkers who might have an actual language delay, the finding applied to only four children and was discovered during exploratory posthoc analyses. Thus, at best, the link must be explored further.

Is There Bidirectionality in Perceptual and Language Development?

At the time that we were initially conducting these speech perception and word learning studies, consensus knowledge indicated that infants only begin to recognize words toward the end of the first year of life (Fenson et al., 1994). While maternal reports (from studies using CDI vocabulary checklists) sometimes indicated understanding of a few very common words by 8-10 months of age, the numbers were small and in many cases no words at all were checked off as understood. Similarly, although a few experimental studies provided evidence that infants as young as 6 months old can look reliably more to very common referents – specifically feet versus hands and other body parts (Tincoff & Jusczyk, 1999 and 2012, respectively) in the presence of the relevant word, and that infants as young as 4 months old can respond to their own names (Mandel, Jusczyk, & Pisoni, 1995), these instances were seen as high frequency exceptions. More recently, a growing number of studies have indicated that infants detect the match between words and objects for far more referents and at a far earlier age than anyone had expected. In a widely-cited study, Bergelson and Swingley (2012) reported that if highly familiar objects are used, infants as young as 6 months old show evidence of looking longer to matching than to non-matching objects. This effect is particularly pronounced if familiar voices are used to produce the words (Bergelson & Swingley, in press). Such precocious word recognition finds a plausible mechanism in studies showing that as early as 3 months of age, neural responses indicate rapid learning of the association of words and objects (Friedrich & Friederici, 2017). What is still unknown is the extent to which these early word-object associative linkings carry the same referential weight as does lexical understanding in older infants, or whether they reveal at best ‘proto-words’.

With the increasing evidence that the foundations for referential word learning are in place far earlier than once expected, we can again consider the idea that establishment of native speech sound categories – even in the earliest stages of infant development – involves acquisition of sound categories *within the context* of the earliest appreciation of differential meaning. That is, rather than perceptual attunement being a prerequisite to word meaning, with the phonological categories of the native language subsequently being used to direct word learning, it seems increasingly likely that the appreciation of what words mean and of what phonetic distinctions are phonological in the native language emerge in parallel. In his writings, Heitner (2004) credits Roger Brown with this insight in his 1958 “Word Game”. In my lab, it was an insight Henny Yeung also independently introduced in his (unpublished) master’s thesis (2006), arguing – as does Heitner – that just as the presence of distinct words guides category learning (a la

Waxman), so, too, does the presence of distinct objects guide speech sound categorization (see also Swingley, 2009, for similar arguments). It was this hypothesis that guided the design and implementation of Yeung’s empirical work, reviewed above, showing that the presence of two distinct objects can help infants as young as 9 months old continue to distinguish non-native contrasts (Yeung, Chen, & Werker, 2014; Yeung & Nazzi, 2014; Yeung & Werker, 2009).

Importantly, the suggestion that word-object co-occurrences can also guide speech perception development is different from the original Trubetskoy (1969) notion, which argued that children must first establish a lexicon and only thereafter could a system of contrast be constructed. In the current approach, the establishment of object categories and of speech sound categories develop together in a mutually informative fashion. With the evidence that the presence of words drives categorization from early in life (Waxman, 2004; Ferry, Hespos, & Waxman, 2013), the evidence that vocabulary learning begins far earlier than we previously expected (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999; 2012), and the emerging evidence that the presence of distinct objects drives speech sound categorization (Yeung, Chen, & Werker, 2014; Yeung & Nazzi, 2014; Yeung & Werker, 2009), we are well-positioned to begin looking earlier in infancy – in the period during which perceptual attunement takes place – for the co-emergence of object and speech sound categorization. This is not to argue that statistical learning doesn’t also play an important role. Indeed, distributional learning of language-specific phonetic (rather than phonological) categories could yield exactly those kinds of stable percepts on which a cyclical “word game” could then operate. But rather than having to necessarily come first, statistical learning of native phonetic categories could also occur concurrently with semantically-driven acquisition, with the statistical regularities and object mappings providing stimultaneous converging sources of confirmation (also see Johnson, 2016).

In this way, listening to speech and noting its functional significance could influence and be influenced by the development of native language phonetic categories, the emergence of native language phonological categories, and a conventionalized understanding of the kinds of semantic distinctions that are encoded in the native language.

A Final Consideration: Speech Perception is Multisensory

Visual Influences

As noted at the outset, infant speech perception is not only auditory, but is, like adult speech perception, richly multisensory. Perceptual attunement has been shown to apply to speech perception in other sensory modalities (e.g. vision) as well as to multisensory speech perception (e.g. audio-visual integration). For example, young but not older infants can discriminate rhythmically-distinct languages just by watching silent talking faces. When habituated to multiple speakers producing sentences in one language but with the sound turned off, and then shown those same speakers producing new sentences in either the old language or in a rhythmically-distinct language, infants 4 and 6 months of age show a recovery in looking time to the language switch (Weikum et al., 2007). By 8 months of age, mononlingual-learning infants no longer do so, whereas bilingual-learning infants maintain the ability to discriminate the two languages visually (Sebastian-Galles, Albareda, Weikum, & Werker, 2012; Weikum et al., 2007).

We have also known for some time that from as early as 4 months old (Kuhl & Meltzoff, 1982) and even 2 months old (Patterson & Werker, 2003), infants can match heard with seen phonetic information. For example, when shown two side by side faces of the same person, one producing an /i/ and the other an /a/, infants will look to the side that matches what they hear when presented with an auditory /i/ or /a/. Preferential looking to the match in the first 6 months of life is seen not only for native speech, but also for non-native speech sounds (Danielson, Bruderer, Kandhadai, Vatikiotis-Bateson, & Werker, 2017; Pons, Lewkowicz, Soto-Faraco, & Sebastian-Galles, 2009; Kubicek et al., 2014).

Just as auditory phonetic perception attunes across the first year of life, so does audio-visual (AV) match. Spanish lacks a /v/ phone and hence lacks a ba-va phonetic distinction. Yet young Spanish infants will look preferentially to a picture of a face articulation of /va/ or /ba/ when hearing the matching syllable, but stop doing so by 9-10 months of age (Pons et al., 2009). We see related results for non-matching speech: at 6 but not 11 months of age, infants look differentially to the face when viewing mismatching, non-native AV speech displays than they do to matching ones (Danielson et al., 2017). This provides another example of AV perceptual attunement. Moreover, the presentation of matching AV speech seems to facilitate discrimination of an otherwise difficult native distinction at 6-8 months of age (Teinonen et al., 2008) and of an otherwise non-discriminable non-native vowel difference at 9-11 months of age (Ter Schure, Junge, & Boersma, 2016; but see Danielson et al., 2017, for a lack of facilitation in non-native consonant discrimination).

AV interactions continue to influence speech processing as infants move into word recognition and word learning. At 12-13 months of age, infants detect auditory mispronounciations of familiar words in both auditory-only and AV matched conditions, but fail to detect mispronunciations when the AV display is mismatched (Weatherhead & White, 2017). By 18 months of age, if first taught a word-object match with only auditory information, infants will look correctly to the match in both an auditory-only and in a visual-only test condition (Havy, Foroud, Fais, & Werker, in press). Infants of this age cannot, however, learn the pairing when given only visual facial information in the training phase, whereas adults can. In this study, we first presented infants with a frontal view of a woman pronouncing two nonsense words – but importantly, with no referent object. This brief AV exposure was followed by a training period wherein the woman turned toward an object and pointed at it while she produced the word. Two word-object pairings were taught. In the A-only training condition, her pointing arm covered the lower half of her face, preventing access to most of the visual phonetic information. In the V-only training condition, her pointing arm was below the face, thus showing the infants the articulatory movements, but the sound was turned off. The subsequent test phase assessed word recognition in a preferential looking task wherein the two objects were presented side by side, and the infant either heard (A-only) or saw (V-only) the woman presenting one of the words. Word learning was evident by longer looking to the match. As noted above, toddlers could learn the words in the A-only training condition (they subsequently demonstrated word recognition in both the A-only and V-only testing conditions), but they could not learn the words in the V-only training condition, whereas adults could.

While not the focus of this article, it should be noted that visual influences may also play a role in the perceptual bootstrapping of syntactic structure. Young infants can use both word frequency (Gervain et al., 2008) and prosody (Bernard & Gervain, 2012) to parse an artificial language into constituent phrases, and bilingual learning infants can use both simultaneously (Gervain & Werker, 2013). Current work in my lab is investigating the extent to which visual cues in talking faces also convey these prosodic differences. To date we have found that adults can use cues such as head nods to parse language streams into syntactic phrases (de la Cruz Pavia, Werker, Vatikiotis-Bateson & Gervain, 2016) and are currently investigating this same question with infants.

Oral Motor Influences

A growing body of research indicates that oral-motor movements, or the somatosensory feedback from them, also impact speech perception. The first evidence came from a number of naturalistic observations, showing, for example, that infants are better able to discriminate sounds that are in their babbling repertoire (DePaolis, Vihman, & Portnoy, 2011), and are more likely to imitate sounds when they both hear the sounds and see the faces moving than when only listening (Coulon, Hemimou, & Streri, 2013; for a full review, see Guellai, Streri, & Yeung, 2014). While in my lab, Henny Yeung conceptualized and designed the first study to test this empirically, by investigating the influence of oral-motor movements on auditory-visual matching (Yeung & Werker, 2013). This work with infants 4-5 months old revealed a robust influence – in this case as a contrast effect. When watching two side by side faces, one articulating /i/ and the other /u/, and presented with the heard vowel of one or the other, the infants looked to the matching side. However, when a pacifier (or the parent’s finger-tip) was held between the infants’ lips (resulting in a rounded lip configuration much like that required for producing an /u/), they looked to the visual /i/ when hearing the /u/, and when a teether (or the long side of the parent’s finger) was placed between the infants’ lips (resulting in a stretched lip configuration much like that required for producing an /i/), they looked more to the /u/ face when hearing the /i/ sound (Yeung & Werker, 2013). Control conditions confirmed this effect only when there was shared information between the oral-motor movement and the corresponding sound heard and seen.

More recently, we have found an oral-motor influence not just on AV matching, but on auditory-only speech perception as well. As discussed earlier, at 6 months of age, English-learning infants are able to discriminate the non-native distinction between the dental /da/ and retroflex /Da/ phones produced in Hindi, even though they have not heard them in canonical syllable form in their language input (e.g. Werker & Tees, 1984). We have recently shown that this discrimination is impaired if 6-month-old infants are prevented from moving their tongue tips in ways that correspond to the adult production of a dental /da/ versus a retroflex /Da/ (Bruderer, Danielson, Kandhadai, & Werker, 2015). Specifically, when a flat teether was held in their mouths such that it interfered with tongue tip movement, infants no longer showed evidence of discrimination, whereas when a gummy teether was held in their mouths such that it still allowed tongue tip movement, discrimination remained intact. The fact that this effect was seen in 6-month-old infants even for a non-native distinction, and even though infants this age are not yet babbling consonant-vowel syllables – and hence not yet fully imitating the sounds heard in their native language – suggests preparation prior to specific experience for linking heard speech and related self-produced motor movements (see Choi, Kandhadai, Danielson, Bruderer, & Werker, in press, for a discussion of the events in prenatal and early postnatal development that might support the mapping between heard speech and self-produced oral-motor movements even prior to babbling).

How Does Multisensory Speech Perception Fit into the Overall Picture?

There are many ways in which the study of multisensory speech perception could change our approach to understanding the perceptual foundations of language acquisition. Given that the speech percept is multisensory from early in life and that perceptual attunement is found for both visual-only and AV speech perception, perhaps work on critical/sensitive periods and influences on their timing should be focused on multisensory rather than auditory-only speech percepts. Issues concerning neural organization – and thus potential compensation, loss, and reorganization – come to the fore. Speech as a multisensory percept also has important implications for infants with different types of sensory impairments – a chapter unto itself.

Secondly, a multisensory speech percept could be part of what contributes to the privileged status of speech for categorization and reference. It has been suggested that multisensory binding is an important part of what makes a representation a candidate for abstraction (Xu, 2016). To the extent that the speech percept is multisensory from very early in life – or at least to the extent that the same phonetic content can be recognized across different sensory modalities – it has the qualities that favor abstraction. Thus, unlike other kinds of percepts where cross-modal recognition needs to be first established through experience and only then can the representation become a candidate for a symbol, the multimodal status of the speech percept may be part of what enables it to support categorization and potentially even reference from very early in life. This in turn may further support the bidirectional interface between phonetic attunement and phonological development.

Summary

In summary, language acquisition begins long before the production or comprehension of the first word. In this article, I have reviewed some aspects of the intellectual journey my work has taken as my students, postdoctoral fellows, collaborators and I have tried to understand the perceptual foundations of language acquisition. We have considered the perceptual biases infants have at birth that get them started, the mechanisms by which they become attuned to the phonetic and phonological categories of the native language, and the complex relation between speech and language. My current thinking is that the relation is a bidirectional one, with speech perception development guiding word learning and other aspects of language development, but with a preparation for reference and meaning also guiding speech perception development.

References

Albareda-Castellot, B., Pons, F. & Sebastian-Galles, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental Science,* 14, 395-401.

Archer, S., Ference, J., & Curtin, S. (2014). Now you hear it. Fourteen-month-olds succeed at learning minimal pairs in stressed syllables. *Journal of Cognition and Development,* 15, 110-122.

Baker, S. A., Golinkoff, R. M., & Petitto, L. A. (2006). New insights into old puzzles from infants' Categorical Discrimination of soundless phonetic units. *Language Learning and Development,* 2, 147-162.

Barton, D. (1980). Phonemic perception in children. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child Phonology Volume 2: Perception.* New York, NY: Academic Press.

Beckman, M. E., & Edwards, J. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development,* 71, 240–249.

Begus, K., Gliga, T., & Southgate, V. (2016). Infants' preferences for native speakers are associated with an expectation of information. *Proceedings of the National Academy of Sciences*, 113, 12397-12402.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109, 3253-3258.

Bergelson, E., & Swingley, D. (in press). Young infants' word comprehension: New talker versus wrong phoneme. *Child Development.*

Bernard, C. & Gervain, J. (2012). Prosodic cues to word order: What level of representation? *Frontiers in Psychology*, 3, 85-90.

Bernhardt, B. M., Kemp, N., & Werker, J. F. (2007). Early word-object associations and later language development. *First Language*, 27, 315-328.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words*. Cambridge, MA: MIT Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1997). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception & Performance,* 14, 345-360.

Bosch, L., & Sebastian-Galles, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, 65, 33-69.

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

Brown, C, & Matthews, D. (1997). The role of feature geometry in the development of phonemic contrasts. In S. J. Hannahs and M. Young-Scholten (Eds.), *Focus on Phonological Acquisition*. Philadelphia, PA: John Benjamins.

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences,* 112, 13531-13536.

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, *28*, 455-474.

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science*, 21, 343-348.

Choi, D., Kandhadai, P., Danielson, D. K., Bruderer, A. G., & Werker, J. F. (in press). Does early motor development contribute to speech perception? [Peer commentary on "Neonatal Imitation in Context: Sensory-Motor Development in the Perinatal Period" by N. Keven & K.A. Akins]. *Behavioral and Brain Sciences*.

Cristia, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics*, 39, 388-402.

Coloun, M., Hemimou, C., & Streri, A. (2013). Effects of seeing and hearing vowels on neonatal facial imitation. *Infancy*, 18, 782-796.

Curtin, S. & Werker, J. F. (2007). Perceptual Foundations of Phonological Development. In M. Gareth Gaskell, G.T.M Altmann, P.Bloom, A. Caramazza and P. Levelt (Eds.), *Oxford Handbook of Psycholinguistics.* New York, NY: Oxford University Press.

Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., & Werker, J. F. (in press; online March 6, 2017) The organization and reorganization of audiovisual speech perception in the first year of life. *Cognitive Development*.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science,* 208, 1174-1176.

Dehaene-Lambertz, G. (2000) Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience,* 12, 449-460.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, 298, 2013-2015.

Dehaene-Lambertz, G., & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience*, 16, 1375-1387.

de la Cruz-Pavía, I., Werker, J. F., Vatikiotis-Bateson, E., & Gervain, J. (2016, September 1). *Head nods influence perceived phrase segmentation: a study with adult monolinguals and bilinguals*. Poster presented by I. de la Cruz Pavia at the AMLAP16 Architectures and Mechanisms for Language Processing Psychology Workshop, September 1-3, Bilbao, Spain.

DePaolis R. A., Vihman M. M., & Keren-Portnoy T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development,* 34, 590-601.

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*, 104, 16027-16031.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science,* 171, 303-306.

Fais, L., Werker, J. F., Brown, C., Leibowich, J., Barbosa, A. V., & Vatikiotis-Bateson, E. (2012). Here's looking at you, baby: What gaze and movement reveal about minimal pair word-object association at 14 months. *Laboratory Phonology*, 3, 91-124.

Fennell, C. T., (2012). Object familiarity enhances infants' use of phonetic detail in novel words. *Infancy*, 17, 339-353.

Fennell, C. T. & Byers-Heinlein, K. (2014). You sound like Mommy: Bilingual and monolingual infants learn words best from speakers typical of their language environments. *International Journal of Behavioral Development,* 38, 309-316.

Fennell, C. T., Byers-Heinlein, K., & Werker, J. F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development*, 78, 1510-1525.

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, 81, 1376-1383.

Fennell, C. T., & Werker, J. F. (2003). Early word learners’ ability to access phonetic detail in well-known words. *Language and Speech*, 46, 245-264.

Fenson, L., Dale, P. S., Reznick, 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, 1-185.

Ferguson, B. & Waxman, S. R. (2016). What the [beep]? Six-month-olds link novel communicative signals to meaning. *Cognition,* 146, 185-189.

Ferry, A., Hespos, S., & Waxman, S. (2013). Non-human primate vocalizations support categorization in very young human infants. *Proceedings of the National Academy of Sciences,* 110, 15231-15235.

Friedrich, M., & Friederici, A. D. (2017). The origins of word learning: Brain responses of 3-month-olds indicate their rapid association of objects and words. *Developmental Science*, 20: e12357.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese-Italian cross-linguistic study. *Cognitive Psychology,* 57, 56-74.

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology*, 61, 191-218.

Gervain, J. & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, 4, 1-6.

Gervain, J., Werker, J. F., Black, A., & Geffen, M. N. (2016). The neural correlates of processing scale-invariant environmental sounds at birth. *NeuroImage*, 133, 144-150.

Gómez, D. M., Berent, I., Benavides-Varela, S., Bion, R. A. H., Cattarossi, L., Nespor, M., & Mehler, J. (2014). Language universals at birth. *Proceedings of the National Academy of Sciences*, 111, 5837-5841.

Guellaï, B., Streri, A., & Yeung, H. H. (2014). The development of sensorimotor influences in the audiovisual speech domain: Some critical questions. *Frontiers in Psychology*, 5, 1-7.

Havy, M., Foroud, A., Fais, L., & Werker, J. F. (in press; online January 26, 2017). The role of auditory and visual speech in word-learning at 18 months and in adulthood. *Child Development.*

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research, 23*, 1-31.

Heitner, R. (2004). The cyclical ontogeny of ontology: An integrated developmental account of object and speech categorization. *Philosophical Psychology*, 17, 45-57.

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nature Reviews Neuroscience*, 6, 877-888.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development,* 32, 262-274.

Johnson, E. K. (2016). Constructing a proto-lexicon: an integrative view of infant language development. *Annual Review of Linguistics*, *2,* 391-412.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64, 675-687.

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., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33, 630-645.

Kemp, N., Scott, J., Bernhardt, B. M., Johnson, C. E., Siegel, L. S., & Werker, J. F. (2017). Minimal pair word learning and vocabulary size: Links with later language skills. *Applied Psycholinguistics*, 38, 289-314.

Kluender, K. R., Lotto, A. J., Holt, L. L., & Bloedel, S. L. (1998). Role of experience for language-specific functional mappings of vowel sounds. *The Journal of the Acoustical Society of America*, 104, 3568-3582.

Kubicek, C., de Boisferon, A. H., Dupierrix, E., Pascalis, O., Loevenbruck, H., Gervain, J., & Schwarzer, G. (2014). Cross-modal matching of audio-visual German and French fluent speech in infancy. *PLOS ONE*, 9: e89275.

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

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

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews,* 5, 831-843.

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron,* 67, 713-727.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363, 979-1000.

Kuhl, P. K., & Meltzoff, A N. (1982). The bimodal perception of speech in infancy. *Science,* 218, 1138-1141.

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

Kuhl, P. K., Tsao, F., & Liu, H. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences,* 100, 9096-9101.

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

Lawrence, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology,* 39, 770-784.

Liu, L., & Kager, R. W. J. (2014). Perception of tones by infants learning a non-tone language. *Cognition,* 133, 385-394.

MacKenzie, H., Graham, S. A., & Curtin, S. (2011). Twelve-month-olds privilege words over other linguistic sounds in an associative learning task. *Developmental Science*, 14, 399-410.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, 6, 314-317.

Mani, N., & Plunkett, K. (2010). Twelve-month-olds know their cups from their keps and tups. Infancy, 15, 445-470.

Marno H, Farroni T, Vidal Dos Santos Y, Ekramnia M, Nespor M, Mehler J. Can you see what I am talking about? Human speech triggers referential expectation in four-month-old infants. *Scientific Reports*. 5: 13594.

Martin, A., Onishi, K. H., & Vouloumanos, A. (2012). Understanding the abstract role of speech in communication at 12 months. *Cognition,*123, 50-60.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, 10, 241-265.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, 106, 1367-1381.

Mattock, K., Polka, L., Rvachew, S., & Krehm, M. (2010). The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science*, 13, 229-243.

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology,* 56, 154-178.

May, L., Byers-Heinlein, K., Gervain, J., & Werker, J. F. (2011). Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech? *Frontiers in Psychology*, 2, 19.

Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us? *Proceedings of the Annual Boston University Conference on Language Development*, 25, 480-490.

May, L., Gervain, J., Carreiras, M., & Werker, J. F. (in press). The Specificity of the Neural Response To Speech at Birth. *Developmental Science*.

May, L., & Werker, J. F. (2014). Can a Click be a Word?: Infants' Learning of Non-Native Words. *Infancy*, 19, 281-300.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11, 122-134.

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

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science,* 12, 369-378.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143-178.

Meltzoff, A. N. & Kuhl, P. K. (2016). Exploring the infant social brain: what's going on in there? *Zero to Three Journal,* 36, 2-9.

Minagawa-Kawai, Y., van der Lely, H., Ramus, F., Sato, Y., Mazuka, R., & Dupoux, E. (2011). Optical brain imaging reveals general auditory and language-specific processing in early infant development. *Cerebral Cortex*, 21. 254-261.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16, 495-500.

Moon, C., Lagercrantz, H. and Kuhl, P. K. (2013). Language experienced *in utero* affects vowel perception after birth: A two-country study. *Acta Paediatrica*, 102, 156-160.

Mugitani, R., Pons, F., Fais, L., Werker, J. F., & Amano, S. (2009). Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology*, 45, 236-47.

Narayan, C., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, 13, 407-420.

Nazzi, T., Floccia, C., Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, 21, 779-784.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, (43), 1-19.

Palmer, S. B., Fais, L., Golinkoff, R. M., & Werker, J. F. (2012). Perceptual narrowing of linguistic sign occurs in the 1st year of life. *Child Development*, 83, 543-553.

Pater, J., Stager, C., & Werker, J. F. (2004). The perceptual acquisition of phonological contrasts. *Language*, 80, 384-402.

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, 22, 237-247.

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science,* 6, 191-196.

Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences*, 100, 11702-11705.

Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., ... & Friederici, A. D. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences*, 108, 16056-16061.

Perszyk, D., & Waxman, S. R. (2016). Listening to the calls of the wild: The role of experience in linking language and cognition in young infants. *Cognition*, 153, 175-181.

Petitto, L. A., Berens, M. S., Kovelman, I., Dubins, M. H., Jasinska, K., Shalinsky, M. (2012). The “Perceptual Wedge Hypothesis” as the basis for bilingual babies phonetic processing advantage: New insights from fNIRS brain imaging. *Brain and Language,* 121, 130-143.

Pi Casaus, G., Sebastian-Galles, N., Werker, J. F., Bonatti, L. (2015, November 15). *The role of linguistic experience in percetual narrowing: the case of bilingual infants*. Talk presented by G. Pi at: Boston University Conference on Language Development, November 13-15, Boston University, Boston, MA, USA.

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.

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastian-Galles, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences*, 106, 10598-10602.

Rost, G.C. & McMurry, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12, 339-349.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science,* 274, 1926-1928.

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology,* 33, 3.

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology,* 46, 106-119.

Sebastian-Galles, N., Albareda, B., Weikum, W., Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science,* 23, 994-999.

Shi, R., & Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science*, 11, 407-413.

Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy,* 10, 187-198.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*(2), B11-21.

Shultz, S., Vouloumanos, A., Bennett, R. H., & Pelphy, K. (2014). Neural specialization for speech in the first months of life. *Developmental Science*, 72, 1-9.

Singh, L., Hui, T. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. *Developmental Science,* 17, 94-109.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastian- Galles, N., … & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, 12, 914-919.

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.

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

Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*, 108, 232-242.

Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. Developmental Psychology, 43, 454-464.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B*, 364, 3617-3622.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition***,** 76, 147-166.

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, 54, 99-132.

Takesian, A. E. & Hensch, T. (2013). Balancing stability/plasticity across brain development. *Progress in Brain Research*, 207, 3-34.

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, 108, 850-855.

Ter Schure, S., Junge, C., & Boersma, P. (2016). Discriminating non-native vowels on the basis of multimodal, auditory, or visual information: Effects on infants' looking patterns and discrimination. *Frontiers in Psychology*, 7, 1-11.

Thiessen E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory & Language*, 56, 16-34.

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin,* 139, 792-814.

Tincoff R. & Jusczyk P. W. (1999) Some beginnings of word comprehension in 6-month-olds. *Psychological Science,* 10, 172-175.

Tincoff R. & Jusczyk P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy,* 17, 432-444.

Trubetskoy, N. S. (1969). *Principles of Phonology.* (C. A. M. Baltaxe, Trans.) Berkeley/Los Angeles: University of California Press. (Original work published 1958)

Vouloumanos, A., Druhen, M. J., Hauser, M. D., & Huizink, A. T. (2009). Five-month-old infants' identification of the sources of vocalizations. *Proceedings of the National Academy of Sciences,* 106, 18867-18872.

Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Development,* 81, 517-527.

Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of Sciences,* 109, 12933-12937.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, 10, 159-64.

Waxman, S. R. (2004). Everything had a name, and each name gave birth to a new thought: Links between early word-learning and conceptual organization. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a Lexicon*. Boston, MA: MIT Press.

Weatherhead, D. & White, K. S. (2017). Read my lips: Visual speech influences word processing in infants. *Cognition*, 160, 103-109.

Weikum, W. M., Oberlander, T. F., Hensch, T. K., & Werker, J. F. (2012). Prenatal exposure to antidepressants and depressed maternal mood alter trajectory of infant speech perception. *Proceedings of the National Academy of Sciences*, 109, 17221-17227.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastian-Galles, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science*, 316, 1159-1159.

Werker, J. F. (1989). Becoming a native listener: A developmental perspective on human speech perception. *American Scientist*, 77, 54-59.

Werker, J. F., Cohen, L. B., Lloyd, V., Stager, C. L., & Casasola, M. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, 34, 1289-1309.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, 1, 197-234.

Werker, J. F., & Fennell, C. T. **(**2004). Listening to sounds versus listening to words: Early steps in word learning. In G. Hall & S. Waxman (Eds.), *Weaving a lexicon* (pp. 79-109). Cambridge, MA: MIT Press.

Werker, J. F., Fennell, C. T., Corcoran, K., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, 3, 1-30.

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

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, 66, 173-196.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology,* 24, 672-683.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics,* 37, 35-44.

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., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology,* 46, 233-251.

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, 21, 221-226.

Wojcik, E., de la Cruz Pavía, I., & Werker, J. F. (in press). Language Acquisition. *Oxford Research Encyclopedia of Psychology.* Oxford University Press.

Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology*, 47, 1220–1229.

Xu, F. (2016). Preliminary thoughts on a rational constructivist approach to cognitive development: Primitives, symbols, learning, and thinking. In D. Barner & A.S. Baron (Eds). *Core Knowledge and Conceptual Change*. New York, NY: Oxford University Press.

Yeung, H. H. (2005). *Infants' understanding of signs: Linking sounds and concepts* (Unpublished MA thesis). The University of British Columbia, Vancouver.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language,* 68, 123-139.

Yeung, H. H., Chen, L. M., & Werker, J. F. (2014). Referential labeling can facilitate phonetic learning in infancy. *Child Development,* 85, 1036-1049.

Yeung, H. Henny & Nazzi, Thierry (2014). Object labeling influences infant phonetic learning and generalization. *Cognition*, 132, 151-163.

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, 113, 234-243.

Yeung H. H., & Werker J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science,* 24, 603-612.

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen month-old infants learn similar sounding words. *Developmental Science*, 12, 412-418.

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, 15, 420-433.

Figure 1

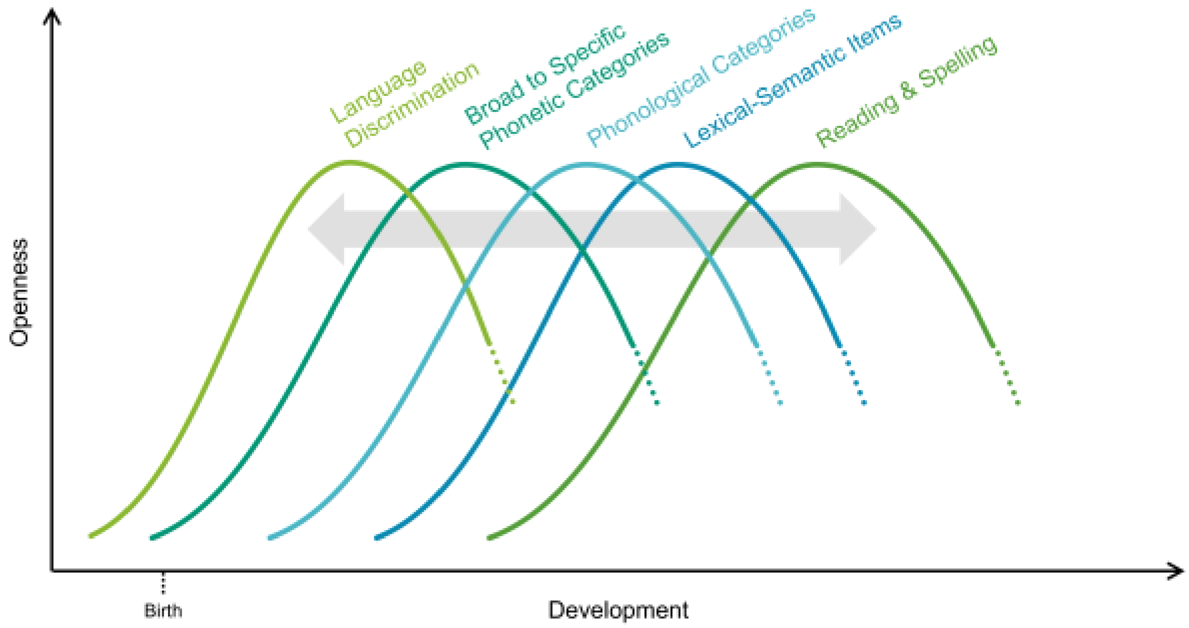

Figure 1. An illustration of the cascading and interacting nature of the steps in language acquisition in infancy. The curves indicate the opening, peak, and diminishment of critical or sensitive periods in the development and mastery of that particular milestone. The overlapping nature of the peaks points to their mutual influence and overlapping timing of emergence. For example, phonological categories and lexical items could emerge simultaneously and be mututally-influential. The arrow is mean to indicate that these influences are not merely bidirectional, but can also occur across levels. For example, learning two words (lexical-semantic items) that differ in only a single phonetic feature might also influence the attunement to native phonetic categories.

Figure 2

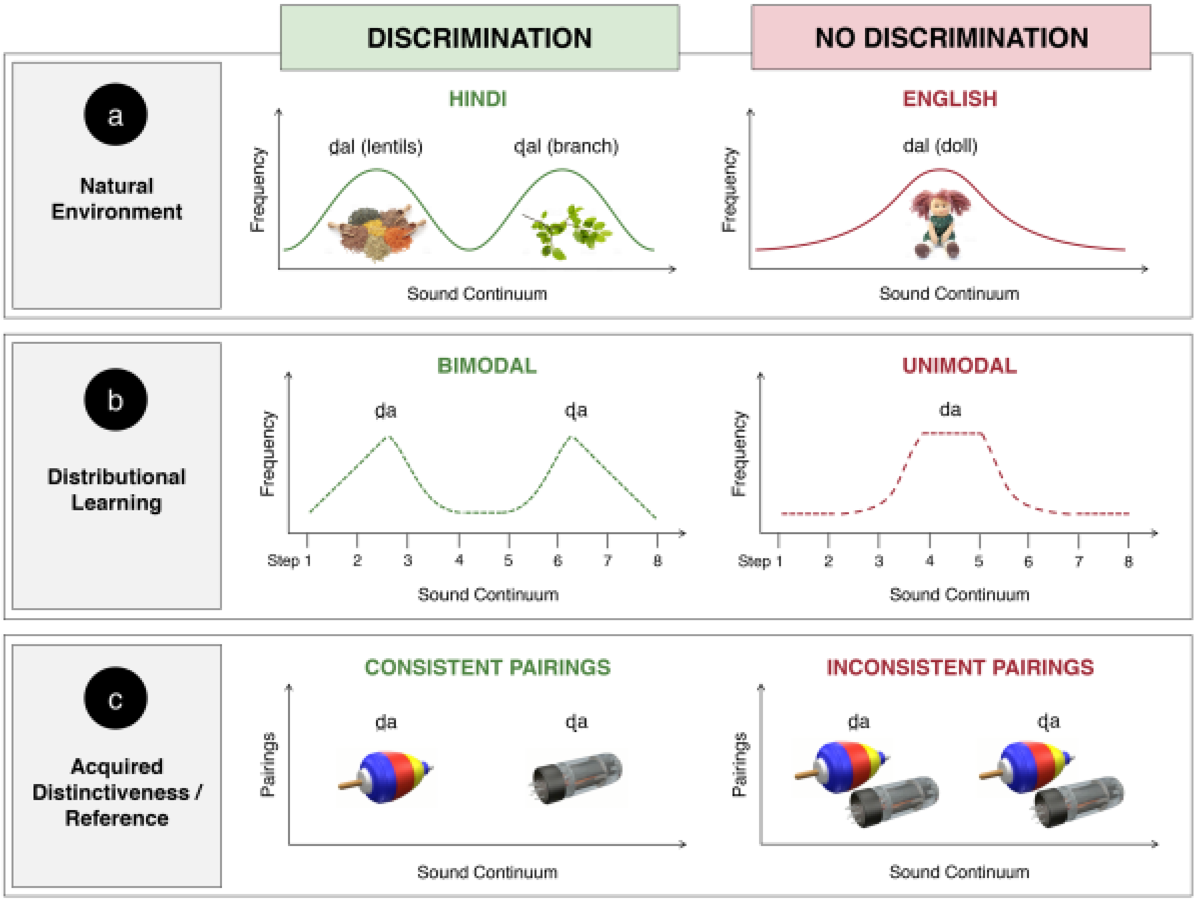

Figure 2. An illustration of speech contrast discrimination in infancy within the context of three “learning” environments: (a) natural, (b) distributional, and (c) acquired distinctiveness and/or reference. The top panel (a) illustrates the variability in the pronunciation of "da" an infant might hear in a natural language environment, depending on whether the infant is growing up learning Hindi (left), in which dental /d̪a/ and retroflex /ɖa/ are contrastive, or English (right), in which they are not. The middle panel (b) illustrates how the variability present in natural learning environments (Hindi-learning infants encounter /da/ phones in a bimodal distribution and English-learning infants encounter /da/ phones in a unimodal distribution) is modeled in studies of distributional learning, using relative frequencies of sounds that are created along an 8-step

continuum from /d̪a/ to /ɖa/. In such studies, English-learning infants will only discriminate /d̪a/ from /ɖa/ if they are first familiarized to a bimodal distribution. The bottom panel (c) illustrates how the same variability is modeled in studies of acquired distinctiveness: two consistent sound-object pairings are presented to highlight the distinction between /d̪a/ and /ɖa/ in comparison to inconsistent pairings. English-learning infants will only discriminate if they are first familiarized to the consistent pairings. For other non-native distinctions, ostensive or other cues to the referential nature of the word-object pairing are required in addition to acquired distinctiveness to enable discrimination. Figure and legend adapted from “How Do Infants Become Experts at Native-Speech Perception?” by J. F. Werker, H. H. Yeung, and K. A. Yoshida, 2012, *Current Directions in Psychological Science, 21*(4), 221-226.