Speech Perception in Infants

Elizabeth K. Johnson

Summary

The study of infant speech perception began a half century ago, with the discovery of categorical perception in young infants. In the decades that followed, researchers grappled with big questions, including whether speech perception abilities were 'special' to humans, and how the infant brain might be innately wired with phoneme category knowledge. To this day, researchers still debate these issues, but the general consensus is that the human infant capitalizes on biologically pre-disposed perceptual and cognitive abilities to rapidly acquire a communication system in a way that far outperforms even the best artificial intelligence systems of our day. Precisely how infants achieve this feat is still far from fully understood. As researchers continue to develop new techniques and strategies for probing the mechanisms underlying infant speech development, increased focus on the processing of fluent naturalistic speech have revealed a prolonged trajectory of development that is strongly influenced by the nature of each individual child's day-to-day linguistic environment. As the field moves forward, an increased focus on understanding infant speech perception develops in real communicative contexts, as well as in diverse learning environments, will lead to better more comprehensive theories of spoken language development.

Keywords

infant studies, speech perception, language acquisition, child development, phonological development, lexical growth, perceptual learning

1. Introduction

The scientific study of infant speech perception traces its roots back to the 1970s, with the discovery that English-learning one-month-olds – like adults – appear to perceive voice onset time (VOT) speech continua in a categorical manner (Eimas, Siqueland, Jusczyk, & Vigorito 1971). This landmark finding bolstered the view that human language depends on innately endowed abilities and provided strong evidence against the view that children learn to hear phoneme contrasts only after they begin producing the contrasts themselves (see Jusczyk, 1997, for review). Soon after the discovery that English-learning infants perceive stop voicing contrasts in a way that resembles adults' perception, additional pioneering research showed that infants (unlike adults) were even sensitive to speech contrasts that did not exist in their native language (e.g., Trehub, 1976), but sensitivity to these contrasts waned as infants approached their first birthday (Werker & Tees, 1984). These findings gave rise to one of the most well-known hypotheses in the field of infant speech perception, discussed in most introductory textbooks on child development: children are born as citizens of the world, perceiving all possible speech contrasts in the human language. But by their first birthday, sensitivity to speech contrasts not used in the children's native language(s) declines, preparing the child to become a more efficient word learner. In the early 21st century, however, there has been a growing awareness that this textbook story is an oversimplification of how developmental speech perception actually works (e.g., Aslin, Werker, & Morgan, 2002; Benders & Alvater-Mackensen, 2017; Best, 2019; Narayan, 2019; Polka, Colantonio, & Sundara, 2001; see Section 3.1). Nonetheless, the general overall pattern of findings reported in the classic literature remain largely supported (e.g., Maurer & Werker, 2013; Sundara, Ngon, Skoruppa, et al., 2018).

As the field of infant speech perception has matured, researchers have expanded the focus of their research to include not only the discrimination of isolated speech sounds or words, but also the recognition of words and phrases in connected speech. This has been a crucially important shift in the field, since all-natural speech – including infant-directed speech – is rife with contextually-dependent variation (e.g., Buckler, Goy, & Johnson, 2018; Martin, Schultz, Versteegh, et al., 2015; Shockey & Bond, 1980). To gain a more complete picture of the acquisition process, researchers have also started to examine children's speech learning environment in a more holistic context – taking into account cognitive and social development as well as the acquisition of other levels of language structure – to understand how developmental speech perception works (e.g., Bergelson, 2020; Johnson, 2016; Johnson & White, 2020; Kuhl, 2012; Vouloumanos & Waxman, 2014). This article outlines the current state of the art in infant speech perception and summarizes several important theoretical debates in the field. The article closes with a discussion of how ecological validity and lack of diversity in our test populations are key considerations to take into account as the field moves forward, emphasizing how much more we still have to learn about infant speech perception in the real world.

2. A Chronological Overview

When does human speech development begin, and how long does it take? Although researchers once thought of birth as the onset of human experience, speech development begins in the womb and continues throughout the lifespan (Saffran, Werker, & Werner, 2007). In the remainder of this section, I will summarize what we currently know about infants' speech perception capabilities from the late fetal period through to the second birthday. A timeline for the first 12 months of development is provided in Figure 1.

[Insert Figure 1 about here]

2.1 The third trimester and neonatal period

Human infants are naturally attracted to the sound of the human voice, just as they are attracted to the human face. Although newborns find their mother's voice to be most interesting (DeCasper & Fifer, 1980), they also find the human voice in general more interesting than other complex acoustic stimuli (e.g., Vouloumanos & Werker, 2007). This may be because humans are innately wired to be drawn to the human voice, just as it has been argued that they are innately drawn to face-like configurations. Or this attraction to the human voice could also (at least partially) be due to auditory learning in the womb (see Faust, Carouso-Peck, Elson & Goldstein, 2020; Vernes, Janik, Fitch, & Slater, 2021; for related discussions in comparative development).

The human auditory system begins functioning in the third trimester of pregnancy, enabling the fetus to begin eavesdropping on their mother's conversations (e.g., Lecanuet & Schaal, 2002). The sound of the mother's voice is conveyed to the fetus vibro-acoustically, through both bone and tissue vibration as well as the audio channel. The sound that reaches the fetus through the uterine wall consists of the low frequency components of speech, emphasizing the vowel and rhythmic and melodic information. Postnatal behavioral studies have demonstrated that the human fetus is indeed learning from their auditory input during this period. Newborns not only prefer the sound of their mother's native tongue over other rhythmically-distinct languages (Moon et al., 1993; Nazzi et al., 1998), they also show a preference for songs and stories heard while they were still in the womb (DeCasper & Spence, 1986; Partanen et al., 2013). Researchers have even suggested that infants may begin learning about the language-specific vowel structure of their mother's tongue while they are still in the womb (Moon, Lagercrantz, Kuhl, 2013). Other studies have shown that neonates appear to be born with the capability of perceiving the acoustic correlates of phrase boundaries (Christophe, A., Mehler, J. & Sebastián-Gallés, 2001), and can distinguish between content and function words (Shi, Werker, & Morgan, 1999) – abilities that have been hypothesized to help infants bootstrap their way into language acquisition (e.g., Gutman, Dautriche, Crabbe, & Christophe, 2015). And perhaps the most surprising evidence for prenatal attunement to the speech signal is evidence that newborns appear to cry in a melody reflective of their mother's native language (Mampe, Friederici, Christophe, & Wermke, 2009).

2.2 From 1 to 6 months

Between 1 and 6 months of age, infants acquire an impressive amount of language-specific knowledge. They begin to learning about the sound structure of their native language (Choi, Cutler, & Broersma, 2017; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), start segmenting some word forms from fluent speech (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Johnson, Seidl, & Tyler, 2014; Shukla, Nespor, & Mehler, 2011), and begin to link some frequent word forms to their intended referent (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012). At this age, infants also distinguish their own variety of their native language from other accent varieties (Butler, Floccia, Goslin, & Panneton, 2011; Kitamura, Panneton, & Best, 2013; Nazzi, Jusczyk, & Johnson, 1999), and show a preference for speakers who have the same language background as themselves (Kinzler, Dupoux, & Spelke, 2007). By 5 to 6 months of age, infants also show evidence for integrating auditory and visual aspects of speech (e.g., Rosenblum, Schmuckler, & Johnson, 1995), and even imitate the production of speech sounds produced by adult interlocuter (Kuhl & Meltzoff, 1996).

Despite all of these impressive accomplishments, infants' speech processing abilities are still far from adult-like by 6 months of age. For example, although 6-month-olds show knowledge of the native language sound structure by displaying enhanced sensitivity to vowel contrasts in the native language, they have not yet tuned into the consonant or tone contrasts specific to their language (i.e., they still do not exhibit language-specific sensitivity to consonant and tone contrasts; e.g., Mattock & Burnham, 2006; Werker & Tees, 1984). And the ability to extract words from natural speech is largely limited to word forms occurring at the edges of major prosodic units (Johnson et al., 2014; see, however, Bortfeld et al., 2005). Moreover, although 6-month-olds listen longer to familiar word forms like their name, they seem insensitive to some types of mispronunciations (Bouchon, Floccia, Fux, Adda-Decker, & Nazzi, 2015; Bergelson & Swingley, 2018), and mixed results have been reported in studies examining how semantically broad infants' early referential knowledge might be (Bergelson & Aslin, 2017; Jusczyk & Aslin, 1999). Perception of speech in noise also continues to be a challenge for infants during this stage of development (Erickson & Newman, 2017).

Interestingly, however, as rudimentary as infant's word recognition abilities are by the end of the first half of year of life, their ability to tell apart talkers and deal with talker variability is much more impressive than studies with older children have suggested (Levi & Schwartz, 2013; Mann, Diamond, & Carey, 1979). For instance, by 4.5 months, infants readily tell apart the voices of unfamiliar females (Fecher & Johnson, 2019a), and there is some evidence that by 6 months infants succeed at cross-gender word recognition (i.e., words familiarized in a female voice are subsequently recognized when presented in a male voice; Johnson et al., 2014; Van Heugten & Johnson, 2012). The use of linguistic knowledge to recognize talkers is also observed at this early age (Fecher & Johnson, 2019a). Indeed, an emerging view in the field is that the development of talker recognition and general speech perception abilities are highly inter-dependent and must be studied together.

But how can we reconcile reports that infants possess somewhat sophisticated talker identification skills with reports that older children still struggle with talker recognition? Some have suggested that infants' ability to identify voices may develop along a U-shaped curve, with performance being reasonably good in infancy, only to become worse for a few years in early childhood, and then eventually reaching adult-like levels of competency later in childhood (Creel & Quam, 2015). However, others have argued that the appearance of a U-shaped curve in children's ability to identify talkers is more likely due to differences in the methodologies used to test infants versus children than any actual loss of ability during early childhood (Fecher, Paquette-Smith, & Johnson, 2019).

2.3 From 6 months to 12 months

In the second half of the first year of life, infants' rudimentary word recognition and speech processing abilities grow far more sophisticated, becoming tightly tuned to the native language sound structure (Jusczyk, 1999; Werker & Curtin, 2005). This attunement coincides with the maturation of other cognitive skills, allowing infants to add new words to their proto-lexicon in a much more efficient manner. At 7 to 8 months of age, infants begin extracting words from speech in all utterance positions (i.e., phrase-medial position; e.g., Mattys & Jusczyk, 2001a). By 9 to 12 months of age, infants have tuned into both the tone and consonant structure of their native language (Mattock & Burnham, 2016; Werker & Tees, 1984; see Werker & Curtin, 2005, for review).

As infants' ability to process the speech signal improves, infants' receptive vocabularies steadily expand. Between 6 and 9 months of age infants have been shown to associate word forms with the appropriate referent for an impressive number of common objects, and the ability to recognize these words appears equally robust regardless of whether the labels are provided by their own mother or a stranger (Bergelson & Swingley, 2018). By 9 to 12 months of age, there is evidence that infants' babbling begins to drift in the direction of the language(s) they are learning (e.g., Boysson-Bardies, Halle, Sagart, & Durand, 1989; Boysson-Bardies & Vihman, 1991; Whalen, Levitt, & Goldstein, 2007), and some indication that first words may be selected from the sound patterns infants have mastered through babbling (e.g., Keren-Portnoy, Majorano, & Vihman, 2009). Even though caregivers are typically focused on children's first word productions around the first birthday, the average one-year-old's receptive lexicon far out-sizes their productive lexicon (Swingley, 2009). Infants' growing receptive lexicons are reflected behaviorally in preferential listening to familiar (or familiarized) word forms when they are properly pronounced, but not necessarily when they are mispronounced, suggesting that infants at this age have much more detailed representations of words than younger infants who fail to notice similar mispronunciations (e.g., Johnson, 2005; Swingley 2005). During this period of development, children also display increasingly impressive talker recognition abilities. For example, by 8 months infants link novel voices to novel faces (Orena & Werker, 2021), and by 9 months monolingual (but not bilingual) infants show better attunement to voice distinctions in the native language than an unfamiliar language (Fecher & Johnson, 2019b; Fecher & Johnson, 2022).

2.4 From 12 to 18 months

By the time infants reach 12 months of age, they typically produce their first recognizable words, giving their caregivers concrete proof of their emerging language knowledge. Indeed, by their first birthday, infants know so much about the sound structure of their native language that it might be tempting to conclude that the only task left for them to do is to add more words to their vocabulary and work out the grammatical structure of their native language(s). Indeed, 12- to 18-month-olds' perception of the speech signal is clearly tuned into the language(s) they are learning (see Johnson, 2016, for review), and efficient enough to support the development of an impressive receptive vocabulary (Swingley, 2009). Children can learn new words quickly (e.g., Schafer & Plunkett, 1998), and their representations of known words appear to be well-specified (Swingley and Aslin, 2002) – even when words are presented in an unfamiliar accent (Van Heugten, Paquette-Smith, Krieger, & Johnson, 2018). During this period of development, toddlers also exhibit increasingly efficient recognition of spoken words (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998).

Yet at the same time, recent research has shown that 13- to 18-months-olds still struggle with many aspects of spoken language processing, revealing they still have a lot to learn before they can even begin to approach adult-like processing of the speech signal. For example, although infants begin segmenting consonant-initial words from fluent speech by 6 months of age, they fail to do the same for vowel-initial words until much later (Seidl & Johnson, 2008; Sundara & Kim, 2014). And although 13- to 18-month-olds appears to be sensitive to many subtle mispronunciations in familiar words in eye-tracking studies, this sensitivity is not always evident in basic word-referent pairing tasks (Stager & Werker, 1997; see, however, Yoshida, Fennell, Swingley, & Werker, 2009) or for certain types of mispronunciations (Alvater-Mackensen, Van der Feest, & Fikkert, 2013). And perhaps most strikingly, this age group appears very limited in their ability to recognize words presented in unfamiliar accents. For example, infants younger than 19 months of age readily listen longer to familiar words spoken in a familiar accent, but this same age group fails to show any listening preference for familiar words spoken in an unfamiliar accent (Best, Tyler, Gooding, Orlando, & Quann, 2009; Van Heugten & Johnson, 2014). Eye-tracking studies have reported similar results (Mulak, Best, Tyler, Kitamura, & Irwin, 2013; Van Heugten, Krieger, Johnson, 2015). Some have argued that infants' poor performance in these studies may be at least partially due to performance demands in particular lab studies more than actual inabilities to cope with accent variation in the real world (e.g., Van Heugten & Johnson, 2014). But other lines of research taking into account children's real-world exposure to multiple varieties of their native language have suggested that accent variation continues to present a substantial challenge to young infants past 18 months of age (e.g., Floccia, Delle Luche, Durrant, Butler, & Gosslin, 2012; see section 4.1).

2.5 From 18 to 24 months

As children near their second birthday, they go through a period of rapid vocabulary expansion (see Swingley, 2009, for review). Many have suggested that this rapid growth in the vocabulary, accompanied by increasingly efficient recognition of words, is triggered by qualitative changes in the way children process the speech signal at about 18 to 19 months of age. For example, it has been argued that infants shift from processing words in an acoustic-phonetic manner to a more adult-like phonological manner, and that this developmental milestone allows for rapid vocabulary expansion (e.g., Nazzi & Bertoncini, 2003). Similarly, others have suggested that rapid vocabulary growth is linked with the development of phonological constancy (e.g., Mulak et al., 2013). Children's rapid improvements in their ability to process other-accented speech at approximately

19 months of age supports the notion that a qualitative change in speech processing happens at this age; however, it is also possible that the abstract speech processing skills are present in younger children, but the task demands associated with laboratory tests of speech perception (and/or the artificial nature of these tests) mask these abilities in children under 19 months, at least to some extent (Paquette-Smith, Cooper, & Johnson, 2021; Van Heugten & Johnson, 2014). That is, the dramatic changes we see in infants' speech perception performance in the lab as they near their second birthday may be explained by quantitative rather than qualitative changes in their abilities (Johnson, Van Heugten, & Buckler, submitted). Regardless of how one categorizes the changes we see in children's speech processing abilities during this period of development, the dramatic improvements see in children's abilities during this time period are indisputable.

2.6 Beyond infancy

Developmental improvements in native language attunement and speech processing efficiency are far from complete at 24 months of age. For example, young children to struggle more than adults with difficult listening conditions, such as understanding foreign-accented speech and coping with noisy listening environments (e.g., Bent, 2018), for many years beyond 24 months. And some basic aspects of speech perception – such as fine-tuning phoneme discrimination abilities (Nittrouer, 2002; Walley, 1993), recognizing talkers (Creel & Quam, 2015), understanding emotional speech (Quam & Swingley, 2012; though see Paquette-Smith & Johnson, 2016), and some sociolinguistic competencies – may continue well into the school years. There is also evidence that major changes in the relationship between children's speech perception and speech production continue to develop past the second birthday (e.g., MacDonald, Johnson, Forsythe, Plante, & Munhall, 2012).

3 Active Debates in Infant Speech Perception

In the past half a century we have learned a lot about how speech perception develops, but as is always the case in any scientific field of study, there is still much we do not yet understand. For example, how do children transition from newborn infants with very limited speech processing skills into 2-year-olds who are surprisingly sophisticated in their language abilities? Despite all we know about *when* infants tune into the phoneme inventory of their native language, we still know little about *how* they do this. And although we know that infants rapidly learn language-specific cues to word boundaries, we don't fully understand how infants acquire this knowledge. In addition, infant speech perception researchers still debate what sort of speech representations infants must possess to successfully cope with the variation in the speech signal, and what role the articulatory motor system plays in early speech processing. Here, I briefly consider each of these topics in turn.

3.1 Acquisition of the Native Language Phoneme Inventory

In the early days of infant research, the finding that neonates were sensitive to many speech contrasts – even if those contrasts did not exist in their own native language(s) – led researchers to propose that infants might be born equipped with innate linguistic categories for all possible phonemic categories used in human language. The finding that infants then lose sensitivity to contrasts that are not used in their native language by about 10 months of age led to the proposal that by this age, infants' brains begin to lose the ability to perceive speech contrasts unused by the native language (or languages; see Werker & Tees, 2005, for review).

But subsequent research suggested these views needed to be modified. First, research with non-speech auditory stimuli, as well as research with non-human animals, suggest that the way in which very young infants perceive speech contrasts is due to the way the human auditory system works (i.e., initial speech processing abilities were not specifically linguistic; see Jusczyk, 1997, for review). Second, a closer analysis of infants' sensitivity to speech contrasts revealed that very young infants did not necessarily perceive *all* speech contrasts used in human language with equal ease (see Benders & Alvater-Mackensen, 2018; Narayan, 2019; for review), and that infants never lose sensitivity to some speech contrasts that are not used in the native language (e.g., Best, McRoberts, & Sithhole, 1988; Best & McRoberts, 2003; Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995). Moreover, there is also some evidence that the contrasts that adults appear to lose sensitivity to can be re-learned to at least some level of competency through intensive training (demonstrating the brain has not literally lost the ability to perceive unused speech contrasts; e.g., Logan, Lively, & Pisoni, 1991; Singh, Liederman, Mierzejewski, & Barnes, 2011; see Lillard & Erisir, 2011, for a related discussion on cortical plasticity). Interestingly, some contrasts are far easier to learn than others, depending on the learner's native language phonology (e.g., Best, Goldstein, Nam, & Tyler, 2016). And finally, attunement to the native language phonology in the first year of life appears to involve not only a loss of the ability to perceive non-native contrasts, but an improvement in the ability to perceive contrasts in the native language (e.g., Kuhl, Stevens, Hayashi, Deguchi, Kirtani, & Iverson, 2006; Narayan, Werker, & Beddor 2010; Polka, Colantonio, & Sundara, 2001). For all of these reasons, classic notions regarding how infants learn about the sound structure of their native language have had to undergo some revision. Although contemporary models of infant speech perception often assume an important role of innate perceptual or learning biases at birth that make the child 'ready to learn' linguistic structure, they typically do not assume innate linguistic categories themselves at birth, and many models allow for variability in how easily infants initially find it to perceive (or learn to perceive) different types of speech contrasts.

Explaining how infants tune into the vowel and consonant contrasts of their native language(s) has been (at least partially) more difficult than determining which sound contrasts infants are sensitive to at birth and identifying when infants begin showing language-specific perception of speech contrasts. We know infants are born able to perceive most contrasts used in the world's languages, and that by the end of the first year of life they are tuned into the contrasts used in their native language. But how do infants do this? Researchers have long since ruled out the notion that infants need to learn minimal pairs before they can determine which speech sounds are contrastive, since infants tune into the phoneme inventory of their native language(s) well before they know many (if any) minimal pairs (i.e., English-learning infants do not need to learn minimal pairs like 'dip' and 'tip' before realizing that voicing is contrastive in word-initial alveolar stops; see Johnson & White, 2019, for review). Instead, bottom-up solutions have been proposed, where infants track the distributional structure of their input, with distinct peaks in the distribution of acoustic-phonetic dimensions indicating likely contrasts (Maye, Werker, & Gerken, 2002; see Lisker & Abramson, 1964, for evidence that peaks and troughs exist in input). For example, English-learning infants could notice that in their input, the bilabial stops 'p' and 'b' they hear in word-initial position have a voice onset time (VOT) that fall into two categories: over 30 ms for "voiceless" 'p' and under or around 15 ms for "voiced" 'b'. At the same time, they may also notice that these two peaks in the distribution are separated by a trough in the distribution, with few VOTs falling midway between these two peaks. Thus, this could allow them to infer that this acoustic-phonetic dimension was important to pay attention to in their native language. Infants learning other languages, with for example a three-way contrast in VOTs, would notice three peaks in their distribution rather than two.

Strong evidence has been provided to support this distributional explanation for how infants tune into the phoneme inventory of their native language. For example, children tend to learn the most frequent contrasts in their language faster, suggesting input statistics play an important role in sound category acquisition (Anderson, Morgan, & White, 2003). And artificial language studies have shown that infants rapidly extract this type of distributional information from speech streams in the lab, and that exposure to different speech sound distributions affects infants' perception of the speech signal (Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002). Other studies have carried out acoustic-phonetic analyses of corpus data, to determine whether this information is present in naturalistic infant-directed speech. Some of these studies suggest that the necessary distributional information is readily available in children's input (e.g., Werker, Pons, Dietrich, Kajikawa, Fais, & Amano, 2007), whereas other studies suggest that this information might not be as readily available as one might hope (e.g., Martin et al., 2015). In an attempt to produce a more ecologically valid explanation for how infants might acquire the phoneme inventory of their native language(s), some have suggested that knowledge integration with other levels of linguistic structure might be key. For example, tracking the realization of vowels in early-acquired words like 'hand' and 'foot' might help children map out the vowel space of their native language (Martin, Peperkamp, & Dupoux, 2013). By paying attention to how a vowel varies in different tokens of an early learned word like 'foot', this could help infants begin to work out the phoneme inventory of their native language before acquiring a large vocabulary (e.g., see also Feldman, Griffiths, Goldwater, Morgan, 2013; Feldman, Myers, White, Griffiths, & Morgan, 2013; Thiessen, 2007; Yeung, Chen, & Werker, 2014; for related proposals). In short, a currently popular view is that the acquisition of the native language sound structure may be synergistic with the acquisition of a lexicon, with word learning helping phoneme learning and vice versa.

As researchers continue to grapple with understanding how infants acquire the phoneme inventory of their language, an additional layer of complexity that researchers will need to explore further is how infants simultaneously work out the sound structure of multiple languages or language variants at once. For example, what does a child do if their primary caregiver speaks with a non-native accent, porting elements of their L1 language into their L2 and thus not providing the same distributional evidence to native language contrasts as other speakers in the community provide? And what does a child do if some talkers in their environment maintain a phonological contrast, and others do not? I revisit this question in section 3.3.

3.2 Segmentation of Words from Fluent Speech

The word segmentation problem refers to the difficulties that arise from the lack of fully reliable cues to word boundaries in spoken language. Overcoming the word segmentation problem is critical to successful language development, as infants are addressed with predominantly multi-word utterances (e.g., Johnson, Lahey, Ernestus, & Cutler, 2013). Indeed, learning to extract words from speech is such a foundational skill that the speed with which children accomplish this feat predicts children's language abilities between 2 and 6 years of age (e.g., Kooijman, Junge, Johnson, Hagoort, & Cutler, 2013; Newman, Bernstein Ratner, Jusczyk, Jusczyk, & Dow, 2006; Singh, Reznick, Xuehua, 2012). But how do infants solve the word segmentation problem?

We know adults solve the word segmentation problem by relying on language-specific cues to word boundaries (Cutler, 2012). For example, content words in English are predominantly stress-initial, and English-speaking adults capitalize on this pattern by being biased to perceive stressed syllables as word onsets. Early in life, infants begin using the same sort of language-specific segmentation strategies as adult users of their native language (see Junge, 2017, for review). For example, by 7.5 months, English learners readily extract stress-initial (or trochaic) but not stress-final (or iambic) words from fluent speech (Johnson & Jusczyk, 2001; Jusczyk, Houston, & Newsome, 1999). Soon thereafter, infants begin using other language-specific segmentation strategies, such as attending to probabilistic phonotactics and subphonemic cues to word boundaries (e.g., Johnson, 2008; Mattys & Jusczyk, 2001b). But given that the cues to word boundaries differ across languages, and that infants are addressed with predominantly multi-word utterances, how do infants learn language-specific strategies for segmenting words from speech?

Tracking transitional probabilities between syllables is perhaps the most well-known proposed solution to how infants solve the word segmentation problem (e.g., Aslin, Saffran, & Newport, 1998; Romberg & Saffran, 2010; Saffran, Aslin, & Newport, 1996). According to this view, infants note both the baseline frequency of syllables in their input as well as the likelihood of those syllables occurring together with any other syllables in the input. Words are identified as sequences of syllables that are statistically most likely to co-occur (e.g., the syllables 'pre' and 'tty' frequenty co-occur because they belong to the word 'pretty', but the syllables 'tty' and 'ba' from the phrase 'pretty baby' will not co-occur as often because they span a word boundary). According to most versions of the statistical learning hypothesis, infants can learn the phonological and prosodic characteristics of words in their language by analyzing the sound properties of the words extracted from speech by tracking transitional probabilities. For example, English-learning infants could learn that most content words begin with a stressed syllable, and thus could then use this knowledge to extract additional words from speech (Thiessen & Saffran, 2007). Language-specific phonotactic patterns could be learned in a similar manner (Sahni, Seidenberg, & Saffran, 2010). Evidence to support this distributional learning hypothesis include artificial language studies showing that 6- to 8-month-olds can use transitional probabilities to find re-occurring word-like units in a stream of syllables (Saffran et al., 1996; Thiessen & Saffran, 2003), corpus analyses (Ngon, Martin, Dupoux, Cabrol, Dutat, & Peperkamp, 2013; Swingley, 2005), and studies presenting infants with carefully controlled natural speech materials (Pelucchi, Hay, & Saffran, 2009).

However, other researchers have questioned the ecological validity of this hypothesis, arguing that infants' ability to segment words from an artificial language (or from an artificially simplified sample of real language) might not scale up to the challenge of natural language (Johnson, 2012; Johnson & Tyler, 2010; Sohail & Johnson, 2016; Yang, 2004). To begin with, the computational explosion associated with tracking and remembering all syllables heard in their input would also be daunting, especially for infants who have not yet learned the phoneme inventory of their native language. And natural speech syllables are highly variable in their acoustic-phonetic realization (e.g., Lahey & Ernestus, 2013). It would presumably be difficult for young infants to pull out neatly categorized syllables from natural running speech, especially in languages where syllable boundaries can be ambiguous. Moreover, tracking transitional probabilities may not be as effective a strategy for infants learning morphologically-rich languages (e.g., Hungarian), or languages with very different phonological structures than English (e.g., Cantonese). And finally, there is little evidence that adults rely heavily on transitional probabilities when extracting words from speech – indeed, by about 8 months of age infants already rely more heavily on speech cues to word boundaries (e.g., stress and coarticulation) than transitional probabilities, even in situations where the reliability of transitional probabilities cues is greatly strengthened relative to their reliability in natural language (e.g., Johnson, 2008; Johnson & Jusczyk, 2001; Johnson & Seidl, 2009). So even in languages where transitional probabilities are more informative regarding the location of word boundaries, it seems like an enormous effort for infants to extract transitional probabilities between syllables only to discard this information a few months later in development.

An alternative to the transitional probability tracking solution has been proposed, whereby infants use acoustically salient prosodic edges (e.g., utterance and phrase boundaries) to learn language-specific cues to word boundaries (Aslin, Woodward, LaMendola, & Bever, 1996; Seidl & Johnson, 2006; see Johnson, 2016, for review). Since the edges of utterance and phrase boundaries are by necessity also the edges of word boundaries, by attending to these acoustically salient boundaries in speech, infants could begin to learn how word boundaries sound in the native language (Johnson et al., 2014). Proponents of this view have argued that attending to the sound properties of utterance and phrase boundaries would be more efficient and less computationally taxing on infants than tracking transitional probabilities between syllables. Support for this Edge Hypothesis has been provided by both natural and artificial language studies (Johnson et al., 2014; Seidl & Johnson, 2006; 2008; Sohail & Johnson, 2016; Shukla, White, & Aslin, 2011), as well as corpus analyses (Johnson, Lahey, Ernestus, & Cutler, 2013; Johnson et al., 2014). But of course, attention to phrase and utterance boundaries also has its challenges, and other sources of information are undoubtedly needed to learn to segment words from speech. In addition to the Edge Hypothesis, additional strategies that infants could use to facilitate the segmentation of words from speech include use of highly frequent words such as the word 'mommy' to break up longer utterances (Brent & Siskind, 2001; Mersad & Nazzi, 2012), and the Possible Word Constraint where infants rely on universals such as the necessity that each word contain a vowel to help them find words in speech (Johnson, Jusczyk, Cutler, & Norris, 2003). In the end, it seems clear that no single word segmentation strategy can on its own enable infants to solve the word segmentation problem (e.g., Brent, 1999). A likely possibility is that infants rely on several word segmentation strategies, to differing degrees depending on the infant's developmental status and the structural properties of their native language(s).

3.3.Dealing with Speech Variability

Spoken language is highly variable, with no word ever produced in an identical fashion twice. And yet, foundational work in developmental speech perception suggested that infants under one year of life genuinely struggle to cope with this variation – with infants under 10.5 months reportedly unable to recognize familiarized words across changes in talker gender, talker accent, or emotional affect (Houston & Jusczyk, 2000; Singh, Morgan, & White, 2004; see White, 2017, for review). These early findings had a huge influence on how the field viewed infant speech perception, supporting assertions that early representations were purely exemplar-based, and that infants could not handle variation in the speech signal efficiently until they had experienced exposure to additional variation in realization of word forms (e.g., Werker & Curtin, 2005; see Goldinger, 1996, for related views on adult speech perception). But if speech is naturally variable, and infants cannot handle this variation, then how do their speech perception skills develop so quickly? Do these limitations mean that when infants learn a new word from a female caregiver, they do not recognize this same word when it is spoken by a male caregiver? And that infants who have exposure to relatively few talkers are slower to develop speech processing skills than those who have routine exposure to more talkers?

Experimental studies designed to address precisely these questions have shown that (contrary to initial reports) 7.5-month-olds are capable of recognizing words they learn from a female caregiver when spoken by a male caregiver (Van Heugten & Johnson, 2012), and that the number of speakers an infant normally interacts with has no measurable effect on children's attunement to the sound structure of their native language (Bergmann & Cristia, 2018). Indeed, more recent studies have shown that infants readily succeed at cross-gender word recognition even with unfamiliar speakers (i.e., not just with their primary caregivers; Bergelson & Swingley, 2018; Johnson et al., 2014), and that infants may not struggle with cross-affect word recognition as much as initially thought (Singh, Nestor, & Bortfeld, 2008). These more recent findings are further strengthened by classic studies showing that 6-month-olds readily handle talker variation in a speech sound discrimination task (Kuhl, 1979). Nonetheless, there is evidence that other-accented speech remains a challenge for infants long after they show success coping with talker and affect variation (e.g., Best et al., 2009), prompting some to propose that infants may use different strategies (requiring different types of linguistic knowledge; e.g., Kriengwatana, Terry, Chládková, & Escudero, 2016) to cope with talker and emotional affect variation than they need to cope with accent variation.

The fact that infants appear to struggle with accent variation long after they demonstrate some competency handling talker and affect variation brings us back to a debate introduced in section 2.4 – how adept are infants under 19 months at coping with different accents? Although initial studies in this area suggested that infants failed to recognize words produced in other accents prior to about 19 months of age (e.g., Best et al., 2009; Mulak et al., 2013), these studies did not allow infants much opportunity to adjust to unfamiliar accents, and isolated word tokens spoken in an unfamiliar accent were presented to infants without visual or communicative context. More recent studies have addressed this limitation by giving 15-month-olds a few minutes to adapt to an unfamiliar accent in the lab, either through watching a video of a woman reading a familiar story before test (Van Heugten & Johnson, 2014), or through live interaction with an experimenter speaking the accent before test (Paquette-Smith et al., 2021). The former of these studies, and to some extent the latter study, demonstrate that in more naturalistic situations where infants are given a chance to adapt to an accent, they are capable of adapting to unfamiliar accents earlier than previously thought (see also White & Aslin, 2011, for evidence for targeted accent adaptation in 19-month-olds). These findings have been used to argue that infants likely possess abstract speech sound representations well before they show the rapid vocabulary growth seen in children after 18 months of age (i.e., at least by 15 months of age, if not earlier). More work is needed in this area though, in particular to determine the mechanism underlying infants' ability to adapt to unfamiliar accents, and to explore whether differences in infants' real-world experience with accent variation might be linked with individual differences with accent adaptation in the lab (see section 4.1).

3.4 The role of the motor system

When Eimas and colleagues (1971) demonstrated that 1-month-olds already perceive speech contrasts in a manner that resembles adult listeners (i.e., in what appeared to be a categorical manner), it seemed one could no longer easily argue that infants need to produce speech contrasts themselves in order for them to perceive them (see Jusczyk, 1997, for discussion). As a result, for many years acoustic-based theories of infant speech perception predominated in laboratory work, while (at least in the developmental world) speculations over the causal link between articulation and perception took a back seat. But in the past 10 years, the link between infants' articulatory skills and speech perception capabilities has received increasing attention. In one recent study, infants under 6 months of age struggle to perceive speech contrasts when their tongue is prevented from moving into positions needed to articulate that contrast (Bruderer, Danielson, Kandhadai, & Werker, 2015; Choi, Bruderer, & Werker, 2019). Other studies have used the motor cortex response seen in infants in response to speech as support for articulatory-based explanations for speech perception (Kuhl, Ramirez, Bosseler, Lin, & Imada, 2014), and have shown that auditory exposure to other language speech sounds in early infancy can improve the ability to produce these contrasts in adulthood (Choi, Cutler, & Broersma, 2017). There is even some evidence that infants perceive the contrasts they can babble better than contrasts they do not babble (Marinella, Vihman, & DePaolis, 2014), and preliminary work suggests that toddlers with cleft palate might struggle to perceive contrasts they have difficulties producing (Fikkert & Lammertink, 2016). In combination, these studies provide tantalizing evidence that speech motor development may play a crucial role in infant speech development. Additional support for the view that articulatory information (or at least visual aspects of articulation) plays a role in speech perception has been provided by studies showing that infants read visual speech information from the face from an early age (Patterson & Werker, 2003), and are perhaps more likely to look to the mouth during challenging listening conditions (e.g., Pons, Bosch, & Lewkowicz, 2015).

Taken together, these findings give researchers good reason to carefully reconsider the potential link between infants' articulatory abilities and their speech perception abilities, and to pay heightened attention to theories emphasizing the importance of articulatory knowledge (or capacities) for perceptual developmental (e.g., Best, Goldstein, Nam, & Tyler, 2016 Kuhl et al., 2014). An important question for future research will be to determine whether there is an actual causal link between the development of articulatory abilities in infants and their ability to perceive speech information, or if the linking between perception and production abilities are more associative. Keeping an eye on current controversies in the adult literature, where there is currently substantial debate over the relationship between perception and production (e.g., Eger & Reinisch, 2019), might also be useful to understanding how perception and production are related in early development. The growing body of literature on children’s processing of their own speech, as well as other children’s speech, may also shed some light on this issue (e.g., Bernier & White, 2019; Cooper, Fecher, & Johnson, 2018; Masapollo, Polka, & Menard, 2016; Shuster, 1998).

4 Critical Analysis of Scholarship

Like all fields of scientific inquiry, the field of infant speech perception has areas that could benefit from further attention and/or improvement. Some of these areas are not unique to infant speech perception – such as the need to promote increasingly rigorous standards of statistical testing, improving the sensitivity and reliability of testing methodologies, overcoming the silo effect where researchers are overly focused on their own field of study and ignore key insights from other fields, ensuring that test samples represent the diversity in human experience around the world, and promoting the consideration of ecological validity in experimental studies. All of these issues important, but the last of these two, the need to ensure diversity of the tested populations (in particular the need to increase the linguistic diversity of test populations, since speech perception is shaped by the linguistic experiences an individual has) and considerations of ecological validity, are particularly relevant to the study of infant speech perception. This section briefly discusses both of these issues and their importance.

‘

4.1 Tackling the Linguistic Diversity Issue

The need to study speech processing in infant populations learning languages other than English has long been recognized in the field of infant speech perception (e.g., Singh, Burnham, Hay, Liu, & Mattock, 2019), because without examining development in infants learning different languages, it is virtually impossible to study the role of language input on speech perception abilities. Thus, the field has always pushed to expand the field beyond the English- and French-learning populations that were initially the focus of most developmental speech research. But a focus on infants learning more than one language has only emerged in the last few decades (Curtin, Byers-Heinlein, & Werker, 2011). In this relatively short period of time, work with bilinguals has shown that speech development in bilingual infants can differ in some important ways from speech development in monolingual infants. For example, bilingual infants may sometimes show a different developmental trajectory than monolingual infants when tuning into the phoneme inventory of their native languages (Byers-Heinlein & Fennell, 2014). Bilingual infants may also rely more on, or be more proficient in using, visual speech cues (Pons et al., 2015; Sebastian-Galles, Albareda-Castellot, Weikum, & Werker, 2012). Indeed, there is even evidence that exposure to multiple languages from birth affects the development of non-speech capabilities, including general learning and cognitive abilities (Kovacs & Mehler, 2009; Fecher & Johnson, 2019b; Singh, Fu, Meaney, Rifkin-Graboi, 2015).

The explosion of work on bilingualism is a great improvement over the field's historical focus on monolingual infants, especially given how common it is for children to learn more than one language at once. But the field still has a long way to go. For example, a potentially important issue that hasn't received as much attention in the field – and is particularly relevant to the study of infant speech perception – is the fact that children can also experience exposure to predominantly non-native varieties of their first language, and/or be exposed to more than one variety of a single native language (Johnson, 2018). Could exposure to non-native and/or multiple varieties of the native language affect speech development in a meaningful way, and if so, how would this type of experience compare to exposure to multiple languages? Is it possible that work looking at bilingual speech development has confounded effects of bilingualism with effects of hearing multiple accents, and/or that differences in speech perception performance reported in various labs might be related to how much accent variation their infant population is typically exposed to? And how might these factors affect infant speech perception studies in populations where exposure to more than one accent or dialect is associated with socio-economic status and/or racial background? Clearly, greater attention to accent and dialect exposure is necessary, especially given the growing body of research showing how much infants initially struggle to comprehend unfamiliar accents (see section 2.4).

To date, there have only been a handful of studies exploring how exposure to primarily non-native and/or multiple varieties of a language affects infant speech perception. There is some indication that exposure to non-native varieties of speech may complicate some early aspects of infant speech development. For example, infants may initially perceive non-native accents as more like unfamiliar languages than accented varieties of their native language (Paquette-Smith & Johnson, 2015). And a study examining how multi-accent exposure affects development has suggested that 20-month-old multi-accent British English learning children preferentially acquire the socially dominant accent in their community instead of the 'other' accent spoken by their parents (Floccia, Delle Luche, Durrant, Butler, & Goslin, 2012). Research from this same population has suggested that multi-accent 20-month-olds possess less detailed representations of words than their age-matched mono-accent peers (Durrant, Delle Luche, Cattani, & Floccia, 2015).

Note that these studies worked with only monolingual children exposed to different regional varieties of their native language. In another word recognition study with Canadian English learning 24-month-olds, researchers examined how exposure to non-native varieties of a language impacted early word recognition. Children routinely exposed to both native and non-native varieties of the native language were shown to recognize words more slowly than their age-matched peers who had little if any exposure to other accents other than the locally dominant native variety of English (Buckler, Oczak-Arsic, Siddiqui, & Johnson, 2017). Comparable results have been reported with Canadian English learning 12-month-old infants exposed to multiple varieties of their native language (both native and regional varieties) who, unlike their age-matched mono-accent peers, fail to show any recognition of familiar word forms (Van Heugten & Johnson, 2017). Then again, at the same time, other studies have suggested that multi-accent exposure would potentially have a positive impact on speech processing (i.e., multi-accent Dutch-learning 24-month-olds outperformed their mono-accent peers in their ability to recognize words spoken in different varieties of Dutch; Van der Feest & Johnson, 2016).

To summarize, although the picture is still a bit murky regarding the effects of multi-accent exposure early in life, the overwhelming indication to date appears to be that exposure to non-native and/or multiple varieties of the native language definitely impacts infant speech perception. Thus, the infant speech perception field has a lot to gain by further investigating the impact of multi-accent exposure on early development. Work in this area will help address the lack of diversity issue currently plaguing infant speech work, which in turn will help researchers test (and hopefully strengthen) current theories of infant speech and language development. Understanding how multi-accent exposure affects development may also have important implications for more applied areas, such as speech therapy and early childhood education.

4.2 Bridging the Divide between the Lab and the Real World

Another aspect of developmental speech perception research that has improved greatly in recent years – but still has much further to go – is bridging the great divide between lab work and the real world. Foundational studies in the field were necessarily based on infants' perception of artificially manipulated isolated syllables. Natural speech variation was almost entirely (if not completely) eliminated from most of these studies. Constraints in early infant testing methodologies limited researchers to asking yes or no questions like 'can infants tell these two speech sounds apart'. And historically, the cost of setting up infant labs and the intensive training needed to test infants also prohibited researchers from working with a wide variety of language learning populations, with most early work disproportionally focused on English and French learning infants in middle-class to wealthy western-culture communities. Thus, early work tended to study only simple speech perception behaviors in just a few language learning populations. Although studies of this sort helped build a solid foundation for the future study of infant speech perception (see Jusczyk, 1997, for review), these early classic studies were not necessarily strong on ecological validity or generalizability to other populations.

But over the years, infant testing methodologies and audio-visual recording and editing technologies have been improved to the point where we can begin to ask much more nuanced questions of infants. For example, eye-tracking technologies not only provide much more informative measures of infant word recognition than preferential looking data, it has advanced to the point where it can be used as a portable technology, facilitating use in a wider variety of settings and communities. And classic infant habituation and preferential looking paradigms are becoming easier to set up on a small budget, thanks to advances in computer programming and audio-visual technology (e.g., Newman, Shroads, Johnson, Kamdar, Morini, Onishi, Smith, & Tincoff, 2021). Brain measures such as ERP and NIRS also hold great promise for helping researchers unlock the early mysteries of infant speech perception (e.g., Kuhl & Rivera-Gaxiola, 2008), as do data dense corpus studies enabled by advances in speech recording and analysis technology (e.g., Cristia, Lavechin, Scaff, et al., 2021). Indeed, a promising approach for future work will be to combine all of these different technologies in creative ways, to ask new questions about how individual experiences by infants relate to individual differences in the development of speech processing skills (e.g., Kooijman, Johnson, & Cutler, 2008).

Another aspect of ecological validity that the field needs to consider more carefully in the future is how infants' performance might differ in real-world contextualized situations versus laboratory-constructed decontextualized situations (as hinted at in our discussion of accent adaptation in section 2.4). For example, although it is now increasingly common for developmental labs to present infants with connected speech (an improvement over the days when all infant speech studies involved isolated syllables), this speech is often presented in the auditory modality only, with no consideration of the role of visual cues or interlocuter identity on speech perception. And even when contextual information such as visual cues are considered, they may not be considered in a way that acknowledges the importance of each child's unique home environment or everyday experience. For example, infants may all be presented with the same experimenter of a certain race speaking in a certain accent, despite differences in the race and accent exposure in the infant population being tested. Recent research has suggested this might be a crucial lurking variable in many studies (e.g., infants growing up in a Caucasian-dominant community are more likely to 'listen through' a speech error when the speaker is non-Caucasian than when they are Caucasian; Weatherhead & White, 2018; see St. Pierre, White, & Johnson, under review, for a related discussion). Findings such as this underscore some potentially very important overlooked elements in the field of infant speech perception and child language acquisition.

5 Summary and Future Directions

In the last half a century, we have learned a lot about how infants tune into the phonemic inventory of their language, extract words from speech, learn to recognize words in different contexts, and deal with variability in the speech signal. But there is still much we do not yet understand. To further unlock the mysteries as to how children develop speech processing expertise, I expect infant researchers will need to integrate hypotheses and methods from a range of diverse disciplines. And we will need to work with populations from a much broader range of linguistic backgrounds using a diverse array of testing technologies. With a strongly interdisciplinary approach and a commitment to working with diverse populations in methodologically innovative ways, the field will be on the way to unlocking the mysteries of infant speech perception, and in the process perhaps better understand why very young humans are so much better than artificial speech recognition systems at processing spoken language.

Further Reading

Best, C.T. (2017). Speech perception in infants: Propagating the effects of language experience. In: The Handbook of Psycholinguistics: E.M. Fernandez & H.S. Cairns (Eds.), 470-490. John Wiley and Sons.

Cristia, A. (2020). Language input and outcome variation as a test of theory plausibility: the case of early phonological acquisition. *Developmental Review*, 57,

Hohle, B., Bijeljac-babic, R., & Nazzi, T. (2020). Variability and stability in early language acquisition: comparing monolingual and bilingual infants' speech perception and word recognition. *Bilingualism: Language and Cognition*, 23, 56-71.

Johnson, E.K., Van Heugten, M., & Buckler, H. (2022). Navigating accent variation: a developmental perspective. *Annual Review of Linguistics*.

Swingley, D. (2021). Infants' learning of speech sounds and word forms. In Papafragou, Trueswell, & Gleitman (Eds), Oxford Handbook of the Mental Lexicon.

Quam, C., & Creel, S.C. (2021). Impacts of acoustic-phonetic variability on perceptual development for spoken language: a review. *Wily Interdisciplinary Reviews in Cognitive Science*, e1558.

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

References

Alvater-Mackensen, N., Van der Feest, S.V.H., & Fikkert, P. (2013). Asymmetries in early word recognition: the case of stops and fricatives. *Language Learning and Development*, 10, 149-178.

Anderson, J.L., Morgan, J.L., & White, K.S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, 46, 155-182.

Aslin, R.N., Saffran, J.R., & Newport, E.L. (1998). Computation of conditional probability statistics by 8-month-olds. *Psychological Science*, 9, 321-324.

Aslin, R.N., Werker, J.F., & Morgan, J.L. (2002). Innate phonetic boundaries re-visited. *The Journal of the Acoustical Society of America*, 112, 1257-1260.

Aslin, R.N., Woodward, J.Z., LaMendola, N.P., & Bever, T.G. (1996). Models of word segmentation in fluent maternal speech to infants. In K. Demuth & J.L. Morgan (Eds.), *Signal to Syntax* (pp. 117-134). Mahwah, NJ: Erlbaum.

Benders, T., Alvater-Mackensen, N. (2017). Before the word: acquiring a phoneme inventory. In G. Westermann & N. Mani (Eds), *Developmental Psychology: Early Word Learning,* New York: Psychology, pp. 1-14.

Bent, T. (2018). Development of unfamiliar accent comprehension continues through adolescence. *Journal of Child Language*, 45, 1-12.

Bergelson, E. (2020). The comprehension boost in early word learning: older infants are better learners. *Child Development Perspectives*, 14, 142-149.

Bergelson, E., & Aslin, R.N. (2017). Nature and origins of the lexicon in 6-month-olds. *PNAS*, 114, 12916-12921.

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

Bergelson, E., & Swingley, D. (2018). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Development*, 89, 1567-1576.

Bergmann, C., Cristia, A. (2018). Environmental influences on infants' native vowel discrimination. The case of talker number in daily life. *Infancy*, 23, 484-501.

Bernier, D.E., White, K.S. (2019). Toddlers' sensitivity to phonetic detail in child speech. *Journal of Experimental Psychology*, 185, 128-147.

Best, C.T. (2019). Experiential effects in speech perception: do they arise from a level playing field? *The Journal of the Acoustical Society of America*, 146, 2760-2760.

Best, C.T., Goldstein, L.M., Nam, H., & Tyler, M.D. (2016). Articulating what infants attune to in native speech. *Ecological Psychology*, 28, 216-261.

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

Best, C. T., McRoberts, G. W. & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance,* ***14****,* 45-60.

Best, C.T., & McRoberts, G.W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, 46, 183-216.

Best, C.T., Tyler, M.D., Gooding, T.N., Orlando, C.B., & Quann, C.A. (2009). Development of phonological constancy: toddlers' perception of native- and Jamaican-accented words. *Psychological Science*, 20, 539-542.

Bortfeld, H., Morgan, J.L., Golinkoff, R.M., & Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychological Science*, 16, 298-304.

Brent, M.R. (1999). Speech segmentation and word discovery: a computational perspective. *Trends in Cognitive Science*, 3, 294-301.

Brent, M.R., & Siskind, J.M. (1999). The role of exposure to isolated words in early vocabulary development.. *Cognition*, B33-B44.

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

Buckler, H., Goy, H., & Johnson, E.K (2018). What infant-directed speech tells us about the development of compensation for assimilation. *Journal of Phonetics*, 66, 45-62.

Buckler, H., Oczak-Arsic, S., Siddiqui, N., & Johnson, E.K. (2017). Input matters: speed of word recognition in two-year-olds exposed to multiple accents. *Journal of Experimental Child Psychology*, 164, 87-100.

Butler, J., Floccia, C., Goslin, J., & Panneton, R. (2011). Infants' discrimination of familiar and unfamiliar accents in speech. *Infancy*, 16, 392-417.

Byers-Heinlein, K., & Fennell, C.T. (2014). Perceptual narrowing in the context of increased variation: insights from bilingual infants. *Developmental Psychobiology*, 56, 274-291.

Christophe, A., Mehler, J., & Sebastian-Galles, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, 3, 385-394.

Choi, D., Bruderer, A.G., & Werker, J.F. (2019). Sensorimotor influences on speech perception in pre-babbling infants: replication and extension of Bruderer et al. (2015). *Psychonomic Bulletin and Review*, 26, 1388-1399.

Choi, J., Broersma, M., & Cutler, A. (2017). Early phonology revealed by international adoptees' birth language retention. *PNAS*, 114, 7307-7312.

Choi, J., Cutler, A., & Broersma, M. (2017). Early development of abstract language knowledge, evidence from perception-production transfer of birth-language memory. *Royal Society Open Science*, 4, 160660. doi:10.1098/rsos.160660.

Cooper, A., Fecher, N., & Johnson, E.K. (2018). Toddlers' comprehension of adult and child talkers: adult targets versus vocal tract similarity. *Cognition*, 173, 16-20.

Creel, S., & Quam, C. (2015). Apples and oranges: developmental discontinuities in spoken-language processing? *Trends in Cognitive Science*, 19, 713-716.

Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Rasanen, O., Bruce, J., & Bergelson, E. (2021). A thorough evaluation of the Language Environment Analysis

(LENA) system. *Behavior Research Methods*, 53, 467-486.

Curtin, S., Byers-Heinein, K., & Werker, J.F. (2011). Bilingual beginnings as a lens for theory development PRIMIR in focus. *Journal of Phonetics*, 39, 492-504.

Cutler, A. (2012). *Native listening: language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

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

Durrant, S., Delle Luche, C., Cattani, A., & Floccia, C. (2015). Monodialectal and multidialectal infants' representations of familiar words. *Journal of Child Language*, 42, 447-465.

Eimas, P.D., Siqueland, E.R., Jusczyk, P.W., Vigorito, J. Speech perception in infants. *Science*, 171, 303-306.

Eger, N.A., & Reinisch, E. (2019). The impact of one's own voice and production skills on word recognition in a second language. *Journal of Experimental Psychology: Learning Memory, and Cognition*, 45, 552-271.

Erickson, L.C., & Newman, R.S. (2017). Influences of background noise on infants and children. *Current Directions in Psychological Science*, 26, 451-557.

Faust, K.M., Carouso-Peck, S., Elson, M.R., & Goldstein, M.H. (2020). The origins of social knowledge in altricial species. *Annual Review of Developmental Psychology*, 2, 225-246.

Fecher, N., & Johnson, E.K. (2019a). By 4.5 months, linguistic experience already affects infants' talker processing abilities. *Child Development,* 90, 1535-1543.

Fecher, N., & Johnson, E.K. (2019b). Bilingual infants excel at foreign language-talker recognition. *Developmental Science*, 22, e12778.

Fecher, N., & Johnson, E.K. (2022). Revisiting the talker recognition advantage in bilingual infants. *Journal of Experimental Child Psychology*, 214.

Fecher, N., Paquette-Smith, M., & Johnson, E.K. (2019). Resolving the (apparent) talker recognition paradox in developmental speech perception. *Infancy*, 24, 570-588.

Feldman, N.H., Griffiths, T.L., Goldwater, S., & Morgan, J.L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review*, 120, 751-778.

Feldman, N.H., Myers, E.B., White, K.S., Griffiths, T.L., & Morgan, J.L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition*, 127, 427-438.

Fernald, A., Pinto, J.P., Swingley, D., Weinberg, A, & McRoberts, G.W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year of life. *Psychological Science*, 9, 228-231.

Fikkert, P., & Lammertink, I. (2016). Speech perception in children with a cleft palate: preliminary data. Proceedings of the 40th Annual Boston University Conference on Language Development, Eds. J. Scott and D. Waughtal, pp 130-142. Somerville, MA: Cascadilla Press.

Floccia, C., Delle Luche, C., Durrant, S., Butler, J., & Goslin, J. (2012). Parent or community: where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition*, 124, 95-100.

Goldinger, S.D., 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn*. 22, 1166–1183.

Gutman, A., Dautriche, I., Crabbe, B., Christophe, A. (2015). Bootstrapping the prosodic bootstrapper: Probabilistic labelling of prosodic phrases. *Language Acquisition*, 22, 285-309.

Houston, D.M., & Jusczyk, P.W. (2000). The role of talker-specific information in word segmentation by infants. *J. Exp. Psychol: Human Perception and Performance*, 26, 1570-1582.

Johnson, E.K. (2005). English-learning infants' representations of word-forms with iambic stress. *Infancy*, 7, 95-105.

Johnson, E.K. (2008). Infants use prosodically conditioned acoustic-phonetic cues to extract words from speech. *Journal of the Acoustical Society of America*, 123, EL 144-148.

Johnson, E.K. (2012). Bootstrapping language: are infant statisticians up to the job? In *Statistical Learning and Language Acquisition*, ed. P. Rebuschat, J. Williams, pp. 55-90. Boston: de Gruyter.

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

Johnson, E.K. (2018). Putting the terms 'monolingual' and 'bilingual' under the microscope. *Applied Psycholinguistics*, 39, 753-756.

Johnson, E.K., Jusczyk, P.W., Cutler, A., Norris, D. (2003). Lexical viability constraints on speech segmentation by infants. *Cognitive Psychology*, 46, 65-97.

Johnson, E.K., Jusczyk, P.W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *Journal of Memory and Language*, 44, 548-567.

Johnson, E.K., Lahey, M., Ernestus, M., & Culter, A. (2013). A multimodal corpus of speech to infant and adult listeners. *The Journal of the Acoustical Society of America*, 134, EL534-EL540.

Johnson, E.K., Seidl, A., & Tyler, M.D. (2014). The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month-olds. *PLOS ONE*, e83546.

Johnson, E.K., & Tyler, M.D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, 13, 339-345.

Johnson, E.K., Van Heugten, M., & Buckler, H. (submitted). Navigating accent variation: a developmental perspective. *Annual Review of Linguistics*.

Johnson, E.K., & White, K. (2019). Six questions in infant speech and language development. In: *Human Language: from Genes and Brains to Behavior*, ed. P. Hagoort. pp. 99-112, Boston, MA: MIT Press.

Johnson, E.K., & White, K. (2020). Developmental sociolinguistics: children's acquisition of language variation. *Wiley Interdisciplinary Reviews in Cognitive Science*, 11.

Junge, C. (2017). The proto-lexicon: segmenting word-like units from the speech stream. In G. Westermann & N. Mani (Eds), *Developmental Psychology: Early Word Learning*. New York: Psychology Press, pp. 15-29.

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

Jusczyk, P.W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences,* 3, 323-328.

Jusczyk, P.W., Houston, D.M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159-207.

Keren-Portnoy, T., Majorano, M., & Vihman, M.M. (2009). From phonetics to phonology: the emergence of first words in Italian. *Journal of Child Language*, 36, 235-267.

Kim, Y. J., & Sundara, M. (2014). Segmentation of vowel-initial words is facilitated by function words. *Journal of Child Language*, 42, 709-733.

Kinzler, K.D., Dupoux, E., & Spelke, E.S. (2007). The native language of social cognition.

PNAS, 104, 12577-12580.

Kitamura, C., Panneton, R. & Best, C. T. (2013). The development of language constancy: Attention to native versus non-native accents. *Child Development*, *84,* 1686-1700.

Kooijman, V., Johnson, E.K., & Cutler, A. (2008). Reflections on reflections of infant word recognition. In Friederici, Thierry (ed.), in book *Early Language Development: Bridging Brain and Behavior*, in series Language Acquisition Research. (pp. 91-114). Amsterdam, The Netherlands: John Benjamins Publishing Compaany.

Kooijman, V., Junge, C., Johnson, E.K., Hagoort, P., & Culter A. (2013). Predictive brain signals of linguistic development. *Frontiers in Psychology*, 4.

Kovacs, A.M. & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *PNAS*, 106, 6556-6560.

Kriengwatana, B., Terry, J., Chládková, K., Escudero, P. (2016). Speaker and accent variation are handled differently: Evidence in native and non-native listeners. PLOS ONE 11, e0156870. https://doi.org/10.1371/journal.pone.0156870

Kuhl, P.K. (1979). Speech perception in early infancy: Perceptual constancy for perceptually dissimilar vowel categories. *Journal of the Acoustical Society of America*, 66, 1168-1179.

Kuhl, P.K. (2012). Social mechanisms in early language acquisition: understanding integrated brain systems supporting language. In J. Decety and J.T. Cacioppo (Eds.), *The Oxford Handbook of Social Neuroscience* (p. 649-667). Oxford, UK, Oxford University Press.

Kuhl., P.K., Conboy, B.T., Coffeey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory explanded (NLM-e). *Philos. Trans. R. Soc. B*, 363, 979-1000.

Kuhl, P.K., & Meltzoff, A.N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100, 2425-2438.

Kuhl, P.K., Ramirez, R.R., Bosseler, A., Lin, J-F.L., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *PNAS*, 111, 11238-11245.

Kuhl, P., & Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annual Review of Neuroscience*, 31, 511-534.

Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kirtani, 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-M, Liu, H-M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Science*, 255, 606-608.

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.

Lecanuet, J., Schaal, B. (2002). Sensory performances in the human foetus: a brief summary research. *Intellectica*, 1, 29-56.

Levi, S.W., & Schwarz, R.G. (2013). The development of language-specific and language-independent talker processing. *Journal of Speech Language, and Hearing Research*, 56, 913-920.

Lew-Williams, C., Peliucchi, B., Saffran, J.R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, 14, 1323-1329.

Lahey, M., & Ernestus, M. (2013). Pronunciation variation in infant-directed speech: phonetic reduction of two highly frequent words. *Language Learning and Development*, 10, 308-327.

Lillard, A.S., & Erisir, A. (2011). Old dogs learning new tricks: neuroplasticity beyond the juvenile period. *Developmental Review*, 31, 207-239.

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

Logan, J.S., Lively, S.E., Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/ and l/: a first report. *The Journal of the Acoustical Society of America*, 89, 874-886.

MacDonald, E.N., Johnson, E.K., Forsythe, J., Plante, P., & Munhall, K.G. (2012). Children's development of self-regulation in speech production. *Current Biology*, 22, 113-117.

Majorano, M., Vihman, M.M., & DePaolis, R.A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development,* 10, 179-204.

Mampe, B., Friederici, A.D., Christophe, A., & Wermke, K. (2009). Newborn's cry melody is shaped by their native language. *Current Biology*, 19, 1994-1997.

Mann, V.A., Diamond, R., & Carey, S. (1979). Development of voice recognition: parallels with face recognition. *Journal of Experimental Child Psychology*, 27, 153-165.

Mattys, S., & Jusczyk, P. (2001a). Do infants segment words or continuous recurring patterns? *Journal of Experimental Psychology: Human Perception and Performance*, 27, 644-655.

Mattys, S., & Jusczyk, P.W. (2001b). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78, 91-121.

Martin, A., Peperkamp, S., & Dupoux, E. (2013). Learning phonemes with a proto-lexicon.

Cognitive Science, 37, 103-124.

Martin, A., Schultz, T., Versteegh, M., Kiyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: a comprehensive test of the hyperarticulation hypothesis. *Psychological Science*, 26, 341-347.

Masapollo, M., Polka, L., & Menard, L. (2016). When infants talk, infants listen: pre-babbling infants prefer listening to speech with infant vocal properties. *Developmental Science*, 19, 318-328.

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

Maurer, D., & Werker, J.F. (2013). Perceptual narrowing during infancy: a comparison of language and faces. *Developmental Psychobiology*, 56, 154-178.

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.

Mersad, K., & Nazzi, T. (2012). When mommy comes to the rescue of statistics: infants combine top-down and buttom cues to segment speech. *Language Learning and Development*, 8, 303-315.

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

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

Mulak, K.E., Best, C.T., Tyler, M.D., Kitamura, C., & Irwin, J.R. (2013). Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent. *Child Development*, 84, 2064-2078.

Narayan, C.R. (2019). An acoustic perspective on 45 years on infant speech perception, part 1: consonants. *Language and Linguistics Compass*, 13, e12352.

Narayan, C.R., 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., & Bertoncini, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Developmental Science*, 6, 36-42.

Nazzi, T., Bertoncini, J., Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 756-766.

Nazzi, T., Juszcyk, 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.

Newman, R.S., Ratner, N.B., Jusczyk, A.M., Jusczyk, P.W., & Dow, K.A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Developmental Psychology*, 42, 643-655.

Newman, R.S., Shroads, E.A., Johnson, E.K., Kamdar, J., Morini, G., Onishi, K., Smith, E., & Tincoff, R. (2021). Introducing BITTsy: the Behavioral Infant & Toddler Testing System. *Behavior Research Methods.*

Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Duter, M., & Peperkamp, S. (2013). (Non)words, (non)words, (non)words: evidence for a protolexicon during the first year of life.

Developmental Science, 16, 24-34.

Nittrouer, S. (2002). Learning to perceive speech: how fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America*, 112, 711-719.

Orena, A.J., & Werker, J.F. (2021). Infants use disambiguation to learn new voices. *Child Development*, 92, e1048-e1060.

Paquette-Smith, M., Cooper, A., Johnson, E.K. (2021). Targeted adaptation in infants following live exposure to an accented talker. *Journal of Child Language*, 48, 325–349.

Paquette-Smith, M., & Johnson, E.K. (2015). Spanish-accented English is Spanish to English-learning 5-month-olds. *Proceedings of the 18th International Congress of Phonetic Sciences,* 18.

Paquette-Smith, M., & Johnson, E.K. (2016). I don't like the tone of your voice: infants use vocal affect to socially evaluate others. *Infanc*y, 21, 104-121.

Partanen, E., Kujala, T., Tervaniemi, M., Huotilainen, M. (2013). Prenatal music exposure induces long-term neural effects.. *PLOS ONE*, e78946.

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

Pelucchi, B., Hay, J.F., & Saffran, J.R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, 80, 674-685.

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

Pons, F., Bosch, L, & Lewkowicz, D.J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science*, 26, 490-498.

Quam, C., & Swingley, D. (2012). Development of children's interpretation of pitch cues to emotions. *Child Development*, 83, 236-250.

Romberg, A.R., Saffran, J.R. Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews. Cognitive Science*, 1, 905-914.

Rosenblum, L.D., Schmuckler, M.A., Johnson, J. (1995). The McGurk effect in infants. *Attention Perception and Psychophysics*, 59, 347-357.

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

Saffran, J.R., Werker, J.F., & Werner, L.A. (2007). *The infant's auditory world: hearing, speech, and the beginnings of language*. In D. Kuhn, R.S. Siegler, W. Damon, & R. M. Lerner (Eds), Handbook of Child Psychology: Cognition, Perception, and language (58-108), John Wiley & Sons, Inc.

Sahni, S.D., Seidenberg, M.S., & Saffran, J.R. (2010). Connecting cues: overlapping regularities support cue discovery in infancy. *Child Development*, 81, 727-736.

Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development*, 69, 309-320.

Sebastian-Galles, N. (2010). Bilingual language acquisition: Where does the difference lie? *Human Development*, 53, 245-255.

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

Seidl, A., & Johnson, E.K. (2006). Infant word segmentation revisited: edge alignment facilitates target extraction. *Developmental Science*, 9, 565-573.

Seidl, A., & Johnson, E.K. (2008). Boundary alignment enables 11-month-olds to segment vowel initial words from speech. *Journal of Child Language*, 35, 1-24.

Shi, R., Werker, J. L., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, **72**, B11–B21. doi:10.1016/S0010-0277(99)-00047-5

Shockey, L., & Bond, Z. S. (1980). Phonological processes in speech addressed to children. *Phonetica*, *37*, 267–274.

Shukla, M., White, K.S., Aslin, R.N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-month-old infants. *PNAS*, 108, 6038-6043.

Shuster, L.I. (1998). The perception of correctly and incorrectly produced /r/. *Journal of Speech, Language, and Hearing Research*, 41, 941-950.

Singh, L., Burhham, D., Hay, J., Liu, L., Mattock, K. (2019). Lexical tone perception in infants and young children: empirical studies and theoretical perspectives. *Frontiers in Psychology,* 10, 1195.

Singh, L., Fu, C.S.L., Meaney, M.J., & Rifkin-Graboi, A. (2015). Back to basics: a bilingual advantage in infant visual habituation. *Child Development*, 86, 294-302.

Singh, L., Liederman, J., Mierzejewski, R., & Barnes, J. (2011). Rapid reacquisition of native phoneme contrasts after disuse: you do not always lose what you do not use. *Developmental Science*, 14, 949-959.

Singh, L., Morgan, J.L., & White, K.S. (2004). Preference and processing: the role of speech affects in early spoken word recognition. *Journal of Memory and Language*, 51, 173-189.

Singh, L., Nestor, S.S., & Bortfeld, H. (2008). Overcoming the effects of variation in infant speech segmentation: influences of word familiarity. *Infancy*, 13, 57-74.

Singh, L., Reznick, S.J., Xuehua, I. (2012). Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Developmental Science*, 15, 482-495.

Sohail, J., & Johnson, E.K. (2016). How transitional probabilities and the edge effect contribute to listeners' phonological bootstrapping success. *Language Learning and Development*.

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.

Sundara, M., Ngon, C., Skorppa, K., Feldman, N.H., Onario, G.M., Morgan, J.L., & Peperkamp, S. (2018). Young infants' discrimination of subtle phonetic contrasts. *Cognition*, 178, 57-66.

Swingley, D. (2005a). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, 8, 432-443.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, 50, 86-132.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philos. Trans. R. Soc. B*., 364, 3617-3632.

Swingley, D., & Aslin, R.N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13, 480-484.

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

Thiessen, E.D., & Saffran, J.R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7- and 9-month-olds. *Developmental Psychology*, 39, 706-716.

Thiessen, E.D., & Saffran, J.R. (2007). Learning to learn: infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, 3, 73-100.

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.

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

Van der Feest, S.V.H., & Johnson, E.K. (2016). Input-driven differences in toddlers' perception of a disappearing phonological contrast. *Language Acquisition*, 23, 89-111.

Van Heugten, M., & Johnson, E.K. (2012). Infants exposed to fluent natural speech succeed at cross- gender word recognition*. J. Speech Lang. Hear. Res.* 55, 554–560. https://doi.org/10.1044/1092-4388(2011/10-0347)

Van Heugten, M., & Johnson, E.K. (2014). Learning to contend with accents in infancy: benefits of brief speaker exposure. *Journal of Experimental Psychology: General*, 143, 340-350.

Van Heugten, M., Paquette-Smith, M., Krieger, D.R., & Johnson, E.K. (2015). Infants' recognition of foreign-accented words: flexible yet precise signal-to-word mapping strategies. Journal of Memory and Language, 100, 51-60.

Van Heugten, M., & Johnson, E.K. (2012). Infants exposed to fluent natural speech succeed at cross-gender word recognition. *Journal of Speech, Language, and Hearing Research*, 55, 554-560.

Van Heugten, M., & Johnson, E.K. (2017). Input matters: multi-accent language exposure affects word form recognition in infancy. *The Journal of the Acoustical Society of America Express Letters*, 142, EL196-EL200

Van Heugten, M., Paquette-Smith, M., Krieger, D.R., & Johnson, E.K. (2018). Infants' recognition of foreign-accented words: flexible yet precise signal-to-word mapping strategies. *Journal of Memory and Language*, 100, 51-60.

Vernes, S.C., Janik, V.M., Fitch, W.T., & Slater, P.J.B. (2021). Vocal learning in animals and humans. *Philosophical Transactions of the Royal Society of London B Biol Sciences*, 376: 20200234.

Vouloumanos, A., & Waxman, S.R. (2014). Listen up! Speech is for thinking during infancy. *Trends in Cognitive Science*, 18, 642-646.

Vouloumanos, A., & Werker, J.F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science*, 10, 159-164.

Walley, A.C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, 13, 286-350.

Weatherhead, D., & White, K.S. (2018). And then I saw her race: race-based expectations affect infants' word processing. *Cognition*, 177, 87-97.

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

Werker, J.F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese, *Cognition*, 147-162.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.

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

White, K.S. (2017). The impact of input variability on word learning. In G. Westermann & N. Mani (Eds), *Developmental Psychology: Early Word Learning*. New York: Psychology Press, pp. 83-95.

White, K.S., & Aslin, R.N. Adaptation to novel accents by toddlers. *Developmental Science*, 14, 372-384.

Yang, C.D. (2004). Universal grammar, statistics, or both? *Trends in Cognitive Science*, 8, 51-56.

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

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

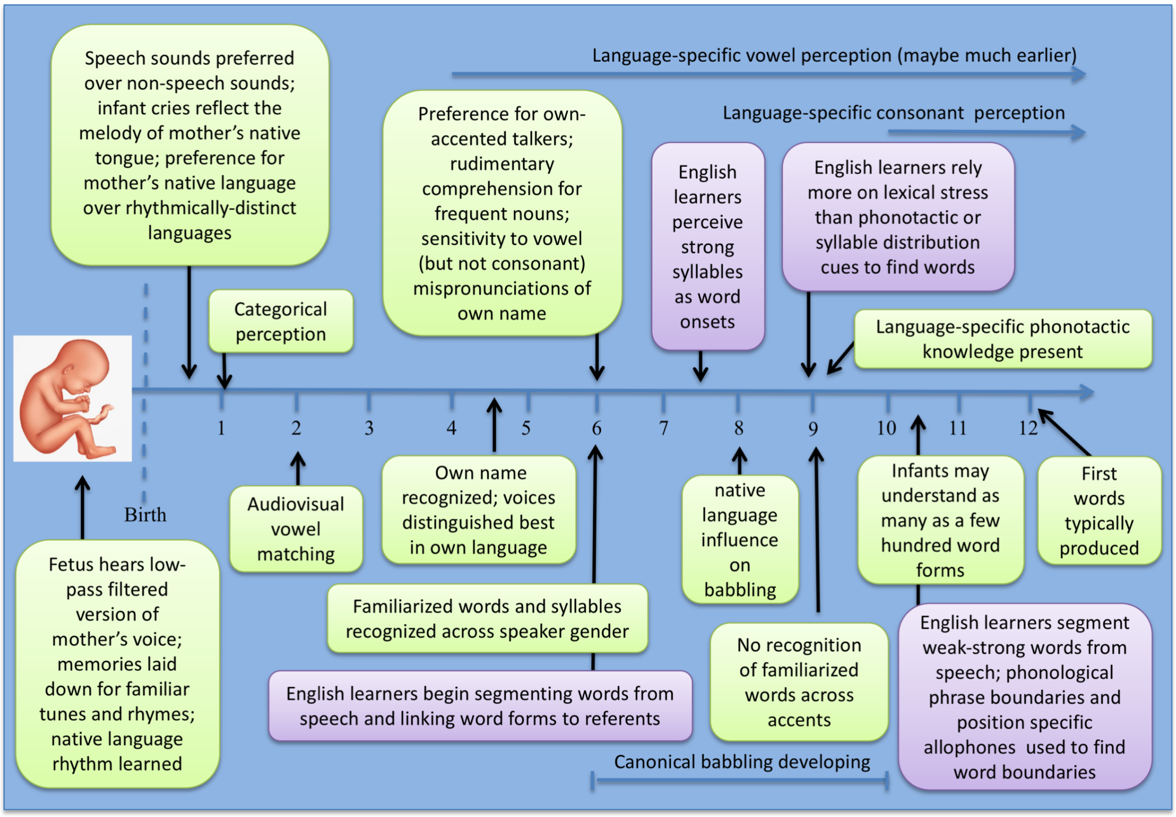

Figure 1. Timeline of infant speech development in the first 12 months of life. Lavender boxes indicate developmental milestones specific to English, whereas green boxes indicate milestones thought to be more language general.