Cascading and Multisensory Influences on Speech Perception Development

Dawoon Choi, Alexis K. Black, and Janet F. Werker

University of British Columbia

Author Note

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

Abstract

Over the first weeks and months following birth, infants' initial, broad-based perceptual sensitivities become honed to the characteristics of their native language. In this paper, we review this process of emerging specialization within the context of a cascading "critical period" (CP) framework, in which periods of maximal openness to experience of different aspects of language occur at sequential, overlapping points in development. Importantly, as infants' experience of speech is not limited to auditory signals, but is informed by – for example – their experience of talking faces and their own oral motor movements, we review the trajectory of perceptual specialization in multisensory language processing. Throughout, we highlight the impact of increasing perceptual specialization on later language outcomes (e.g. word learning, foundations of syntax, literacy), and consider how the outcomes can be compromised if/when the timing of perceptual specialization has been perturbed.

Keywords: infant, neurodevelopment, language acquisition, speech perception, multisensory, critical periods

I.   Introduction

Infants emerge from the womb with a broad array of perceptual biases that set them on the path for language acquisition. Over the first months of life, infants attune these initial biases to become experts at perceiving and using only the properties of the native language or languages they are acquiring in the home. From a broad-based beginning wherein they are prepared to acquire any language, by their first birthday infants show diminished sensitivity to phonetic and acoustic characteristics they have not experienced, and sharpened sensitivities to the properties used in the native language(s). The initial changes in sensitivity do not mean that all plasticity to novel acoustic input has closed by the end of the first year of life; indeed, it is well known that young children adopted internationally even after several years of exposure to an initial language will go on to become fully fluent speakers of their new language (Pallier et al., 2003). Yet even in these cases, the imprint of the initial language environment remains (Hyltenstam, Bylund, Abrahamsson, & Park, 2009; Oh, Au, & Jun, 2010; Choi, Cutler, & Broersma, 2017).

In this paper, we explore this process of attunement within the context of a cascading 'critical period' framework (see Figure 1). This framework views the developing brain as optimally open to different perceptual properties of language at different points in development, beginning with more global, prosodic qualities of speech (e.g. rhythm and intonation), and progressing, for example, to specific segments (e.g. vowels and consonants). While considerable research has focused on how infants process auditory speech, language frequently occurs in face-to-face interactions wherein the infant not only hears speech, but also sees it being spoken, feels the rhythmical body movements that accompany it, and – from early on – produces oral-motor movements that may lay the groundwork for mapping speech sounds to their articulatory

configurations. In this paper, we examine the cascading consequences of early auditory, visual, and sensorimotor signals, and their interaction, on the developing infant's language competence.

Why talk about critical periods?

The term critical period (CP) has classically been used to describe windows in development during which particular systems are most open to experiential influences. However, it is known that windows of openness do not always begin and end abruptly, that their timing can be altered, and that plasticity can extend across the lifespan. Thus, some people have argued for the use of the term 'sensitive period' instead. However, we choose to use the term *critical* period because there are now mechanistic explanations for CP timing and its variations. The elucidation of the circuit and molecular level mechanisms that open and close plasticity allows not only an understanding of how CPs work, but also supports targeted interventions that can prematurely open CPs, keep them open indefinitely, close them prematurely, or reopen them at any time across the lifespan (see Werker & Hensch, 2015, for a broader discussion). To use a metaphor, just as having the 'key' to open a 'lock' doesn't mean we should reject the notion of a lock, knowing the keys for opening and closing critical periods – and hence changing their timing or even enabling lifelong plasticity – doesn't mean we should throw out the notion of CPs. Indeed, we argue it is all the more reason to use the term: the biological mechanism confirms that there are indeed CPs and acknowledging that provides a better path for targeted clinical intervention.

While a full elaboration of the circuit level mechanisms underlying CPs is beyond the scope of this chapter (see Takesian & Hensch, 2013, for detailed mechanistic explanations), generally speaking, there is an underlying push for the brain to be plastic, which is kept in check by the laying down of molecular brakes. Specifically, the onset, offset, and closing of CPs are controlled by a critical balance of inhibition over excitation in a particular class of GABA-

modulated interneurons. When this balance is met, the molecular brakes that are initially in place are removed and the circuit becomes plastic and is open to rewiring from experience, a complex process that requires an interaction of both maturation and experience. The onset of plasticity itself sets in motion another cascade of events, including myelination, that over time leads to CP closure via the laying down of new types of molecular brakes. These brakes have been shown to close the period of plasticity, hence leaving in place a stable and relatively permanent signature of the experiential shaping that occurred between the opening and closing of plasticity.

In the following sections, we outline what is known about early speech perception capacities and linguistic representations, and the emerging science on the relationship of these two phenomena to CPs and neural plasticity. We begin with a description of auditory CPs in non-human species and the hypothesized links to human auditory development, followed by a discussion of the evidence in support of CPs in human auditory speech perception development. We then describe multimodal experiences that contribute to phonetic perception across the first year of life, suggest that the speech percept itself may be multimodal, and consider the evidence in support of CPs on multisensory speech perception development. We conclude by outlining the cascading influence these early experiences potentially have on higher-order levels of linguistic competence, such as literacy.

II. Transforming early auditory experience to higher level functions

Acquiring sound-based representations: A CP model

From early studies on the role of sensory deprivation on development of the visual cortex (Hubel & Wiesel, 1977) and other sensory systems (e.g. Feng & Rogowski, 1980), to work elucidating the difference between experience-expectant and experience-dependent processes in perceptual development (Greenough, Black & Wallace, 1987), animal research has both revealed

that CPs are an evolutionarily preserved mechanism that serves an adaptive function across species, and has helped guide research with humans. Of direct relevance to language processing in humans, work with non-human animals has revealed CP effects in both the perception and production of sound. Perhaps the clearest example has been documented in passerine songbirds, whereby male songbirds learn their song from tutors during a constrained period that is determined by both maturation and experience (e.g. Jarvis, 2004; Brainard & Doupe, 2002). This process involves an initial period of sensory plasticity during which young birds attune to con-specific song, which is followed by (and depending on the species, overlaps with) a plastic period of sensorimotor development. CP effects on the auditory and sensorimotor systems are not limited to birds, however; in rodents, the onset of hearing initiates a period of rapid cortical reorganization that has long-term impacts on adult auditory perception. For example, rats begin to hear around day 12 of postnatal development and remain in a state of plasticity until around day 30. Typical organization of the rat auditory cortex (A1) requires appropriate and adequate environmental input during the described CP. When rat pups are exposed to pure tones during this period of plasticity, those tones become over-represented into adulthood in comparison to control rat populations with no tone exposure, or populations with tone exposure that begins past the period of plasticity (Zhang, Bao, & Merzenich, 2001; Chang & Merzenich, 2003). Interestingly, if the conditions of the sensory input are maladaptive – for instance, pervaded by a continuous level of moderate white noise – rat pups show a significant delay in A1 organization. When adult noise-reared rats are subsequently exposed to structured auditory input, A1 rapidly reorganizes to selectively respond to the novel acoustic environment. Thus, in rodent models the timing of an auditory CP can be shifted by alterations in the degree and quality of early auditory input (Chang & Merzenich, 2003).

It is hypothesized that this model of plasticity may be mirrored in human speech perception development (Doupe & Kuhl, 1999; Werker & Hensch, 2015). First and foremost, it provides unequivocal evidence for auditory CPs. Second, the progression of human speech perception parallels the hallmarks of auditory tuning found in animal models, particularly songbirds' acquisition of species-specific songs. For example, human infants, at birth, are broadly sensitive to perceive many of the phonetic distinctions used across the world's languages (Streeter, 1976; Trehub, 1976; Werker & Tees, 1984; Best & McRoberts, 2003). By the end of the first year of life, infants' perception of linguistic sounds has undergone attunement: that is, there is increased sensitivity to those speech sound differences relevant in the native language environment and a decline in sensitivity to non-native speech sound differences (see Maurer & Werker, 2013, for review). The animal work is foundational to the CP model because it demonstrates that the timing of CPs can be altered depending on environmental input, which has direct implications to the human model. We return to these issues in greater detail below.

Fetal and neonatal perceptual experiences

The stereotypical view of an infant's initial experience is that of an abrupt transition from relative calm to a bright, cacophonous world. There is undoubtedly some truth to this image; however, there are significant sensory opportunities in the infant's prenatal experience that may serve to facilitate the transition to postnatal experience. The fetal ear can conduct sound as early as the 25th week of gestation, thus exposing infants to the sounds of the mother's body, loud external noises, and the speech of nearby talkers through the third trimester. This auditory input has measurable consequences on the infant's perceptual development. For example, fetuses can discriminate their mother's voice from that of a female stranger (Kisilevsky et al., 2003), and prefer to listen to the mother's voice at birth (e.g. DeCasper & Fifer, 1980). Newborn infants

recognize specific phrases that the mother has frequently repeated during pregnancy – even when those phrases are produced by an unfamiliar female (DeCasper & Spence, 1986). While the mother's voice is the loudest, and therefore most salient, fetuses can also hear and learn from speech produced by other speakers (e.g., their father; Lee & Kisilevsky, 2013) and speech repeated via audio recordings (Partanen et al., 2013). Such findings demonstrate remarkable prenatal perceptual and memory capacities.

Nonetheless, the fetal and postnatal auditory environments fundamentally differ from each other, as the tissue and liquid of the womb attenuate higher frequencies. While speech sounds generally extend up to around 10,000 Hz, the low-pass filter of the womb dampens frequencies above approximately 500 Hz (Gerhardt & Abrams, 1996). Thus, before birth, the sounds with which infants have the most experience include pitch, the melody of speech, and some frequency information relevant to vowel identity; they have far less experience with higher frequencies that are relevant for distinguishing many consonant sounds. These facts are reflected in behavioural and neurological experimentation. For example, neonates can perceive the difference between languages of different rhythmic classes – a feature that is determined through the timing of low frequency information (Nazzi, Bertoncini, & Mehler, 1998). Research has also demonstrated fetal discrimination of native vowel sounds (Lecanuet et al., 1987) and improvement in discrimination as the fetus develops (Shahidullah & Hepper, 1994). By birth, infants have acquired sufficient linguistic knowledge to be able to prefer the native language over an unfamiliar language from a different rhythmical class (Moon, Cooper, & Fifer, 1993), and even native over non-native vowel sounds (Moon, Lagercrantz, & Kuhl, 2013).

Accompanying – and perhaps driving – this impressive degree of prenatal learning are universal perceptual biases which position the infant to attend to and learn more about the native language. First and foremost is the preference from birth to attend to speech over similarly

complex non-speech sounds (Vouloumanos & Werker, 2007a; Vouloumanos, Hauser, Werker, & Martin, 2010). This does not appear to be driven simply by experience: when presented with speech and similarly complex non-speech signals that have both been low-pass filtered to simulate the prenatal learning environment, neonates fail to discriminate them (Vouloumanos & Werker, 2007b). Were the preference for speech seen at birth driven entirely by prenatal listening experience, infants should also prefer filtered speech over filtered complex non-speech, which – though similar in complexity, differ acoustically. The predisposition to listen to human language means that the infant directs more attention to and is hence more prepared to learn about language, as opposed to all the other sounds in their world, from the buzz of the computer monitor to the whir of traffic.

Further evidence that infants are born prepared to learn from speech can be found in neuroimaging studies that reveal cortical networks prepared for linguistic signals from birth. For example, neonatal to 3-month-old infants recruit the same cortical areas for processing speech stimuli as do adults (Dehaene-Lambertz et al., 2006), and these areas respond selectively to speech and not non-speech (May, Gervain, Carreiras, & Werker, 2017; Shultz, Vouloumanos, Bennett, & Pelphrey, 2014; Peña et al., 2003). This suggests that the foundation of the speech-specific network is present early. The dual pathway organization of speech processing in adults – comprising ventral and dorsal streams that respectively subserve auditory-to-meaning and auditory-to-motor mappings (Hickok & Poeppel, 2007) – is already present at 2-3 days (Perani et al., 2011), and a major white matter tract relevant for the speech pathways, called the arcuate fasciculus, shows accelerated maturation within the first 2-3 months of life (Leroy et al., 2011).

Speech perception across the first few months of life: Perceptual attunement

In the first few months after birth, infants readily perceive consonant and vowel sounds that are used by languages across the world. By the second half of the first year of life, however, infants' sensitivity to native contrasts is undergoing fine-tuning and enhancement (Narayan, Werker, & Beddor, 2009; Kuhl et al., 2006), while simultaneously their discrimination of non-native contrasts declines. This attunement is seen in the perception of minimal pair vowel, consonant, and lexical tone distinctions (see Werker & Curtin, 2005, for review). While the phonetic inventory that emerges from perceptual attunement is determined by listening experience with the native language, the timing of the attunement process itself appears to be maturationally constrained. Specifically, in a study comparing the ERP discriminative response of full-term infants to those born approximately 12 weeks premature, Peña, Werker, and Dehaene-Lambertz (2012) found that the timing of attunement was determined by gestational rather than chronological age. That is, despite having had an additional 12 weeks of postnatal linguistic experience, premature infants did not stop discriminating the non-native distinctions 12 weeks earlier, but instead continued to discriminate the non-native speech sounds until they reached the same gestational age as the full-term infants. Hence, the timing of openness to experience is gated by neurobiological maturation.

Maturation may set constraints on when in development experience can shape native speech sound categories, but there are factors that impact the underlying neurobiology that can either accelerate or delay the CP timing. Weikum and colleagues (2012) examined one set of factors in a study that compared the discrimination abilities of infants of depressed mothers treated with serotonin reuptake inhibitor (SRI), a class of antidepressant medication, and two control groups: untreated depressed mothers, and mothers not suffering from depression. In rat pups, perinatal SSRI exposure had been previously shown to alter sensory processing and myelin sheath formation, as well as decrease cortical density in auditory and somatosensory cortex

(Simpson et al., 2011). Weikum and colleagues thus hypothesized that SRIs may similarly alter the timing of the CP for phonetic discrimination in infants. This hypothesis was confirmed: fetuses exposed to SRIs showed advanced phonetic discrimination abilities in utero, by perceiving consonant contrasts that were not perceived by unexposed fetuses. Furthermore, while the infants of non-depressed mothers followed the predicted time-course of postnatal phonological development (i.e. at 6 months, they perceived a non-native contrast, and at 10 months, they failed to do so), SRI-exposed infants had already ceased to discriminate the non-native sounds by 6 months. The infants of depressed but untreated mothers, on the other hand, continued to discriminate the non-native sounds at 10 months. Thus, this study supports the hypothesis that SRIs can open a period of plasticity in a developing infant. It furthermore revealed a factor that may delay the onset of the CP: maternal depression. The authors hypothesize that the mechanism behind this delay may relate to reduced input (either through a global reduction in speech heard by the infants, or of speech that is sufficiently engaging) or to reduced serotonin levels. Taken together, the studies described above suggest that – at least for the acquisition of consonants – phonological acquisition depends on a maturationally-constrained CP.

Two recent papers suggest that lasting effects of early exposure on speech sound perception is seen even if the exposure occurs in very early infancy, prior to the onset of what we have suggested is a critical period. Both papers showed that there was a lasting effect of early experience on later perception and production of the birth language, even if that experience ended by 6-months of age, prior to the window in which effects are typically seen, and long before speech production begins. Both papers involved testing of adults who been adopted from Korea into Dutch speaking homes either before 6-months of age or early in childhood but after 17 months of age. Only adults who had been adopted after age 17 months would have had

experience producing Korean prior to adoption, and would have had experience listening to it throughout the period of 6-12 months when perceptual attunement typically occurs, yet both groups of adults performed equally well on Korean perception and production, showing an advantage over Dutch speaking adults (Choi, Broersma, & Cutler, 2017; Choi, Cutler, & Broersma 2017). Taken at face value, these findings suggest that the age of exposure required for a lasting experiential advantage (e.g. evident even in adulthood) may predate the age at which evidence of a decline occurs. That is, the 'opening' of the critical period could be earlier than the timing of the beginning of attunement. This is consistent with other work showing that while there may be clearly demarcated times at which discrimination of many non-native distinctions declines, learning – as revealed by preference – begins with first exposure (Moon, Lagercrantz, & Kuhl, 2013), and strengthens gradually over the ensuing weeks and months (Yeung, Chen, & Werker, 2013). Thus, the accrual of the requisite experience – as a necessary precondition along with biological maturation for CP timing – could begin before the expression of attunement. Indeed, such a two-step CP process is seen in birdsong development (e.g. see Brainerd & Doupe, 2002).

A second possible explanation for the equal performance of the Dutch-speaking adults who had been adopted from Korea at 3-5 months to those adopted after 17-months is that the tasks on which they were tested as adults were not sufficiently sensitive to reveal underlying differences that might be present (e.g. see Werker & Curtin, PRIMIR, 2005). As such, it would be of interest in future research to compare these two groups of adults on their performance in phonological processing tasks that are more challenging, such as a gating task or one that tests speech perception or lexical processing in background noise.

III. Beyond audition: Transforming early multisensory experience to rich linguistic knowledge

Much of our perception and cognition is shaped and informed by the congruency of information from multiple modalities. This is especially apparent in speech, where we hear acoustic signals and see the shape and movement of the articulators (e.g. lips, tongue) and facial features (e.g. head, eyebrows) (Munhall & Vatikiotis-Bateson, 2004). Adult speakers integrate these dynamic bimodal cues. For example, conflicting audio-visual (AV) signals result in a fused phonetic percept (McGurk & MacDonald, 1976). Specifically, when observing a person visually producing one type of sound (e.g. "ba"), but hearing a recording of a different sound (e.g. "ga"), the resulting auditory percept can reflect visual dominance or a blend of the two. Interestingly, the blend reflects the intermediary articulation between the heard and seen syllables (in this case, "da"). This phenomenon is automatic, as it occurs even when subjects are aware of the content of the stimuli (Rosenblum & Saldaña, 1996). Visual speech moreover increases the intelligibility of the auditory signal by around 15dB (Sumby & Pollack, 1954), and improves detection of target sounds (Grant & Seitz, 2000). Neurally, synchronous visual information facilitates early cortical processing of the auditory signal, suggesting that the bimodal signals activate the same underlying abstract representation (Van Wassenhove, Grant, & Poeppel, 2005).

Dynamic faces and voices pervade infants' perceptual experience (Sugden, Mohamed-Ali, & Moulson, 2014; Fausey, Jayaraman, & Smith, 2016), and growing evidence suggests that speech is AV even in very early infancy. When presented with side-by-side images of a face silently articulating either an /a/ or an /i/ vowel, infants 2- and 4-months-old look longer to the face that matches the vowel they hear (Patterson & Werker, 1999; Patterson & Werker, 2003; Kuhl & Meltzoff, 1982). This ability extends to the matching of heard and seen consonants (Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009; see also Danielson et al., 2017). An

event-related potential study that examined the neural responses to auditorily-presented phonemes when primed either with auditory speech (unimodal) or silent visual articulation of the same sound (crossmodal) showed that, indeed, 3-month-old infants showed a similar response pattern to both unimodal and crossmodal priming conditions. Thus, habituating to visually-articulated vowels influences the evoked responses to vowels in a phoneme-specific manner (Bristow et al., 2009). These early perceptual biases indicate that it is not solely lifelong experience of AV congruency that results in bimodal perception of speech, but rather, that speech is multimodal from the outset.

So far, we have only discussed the development of speech perception through the auditory and visual modalities. In adults, however, auditory speech perception is also influenced by sensorimotor signals. Articulatory movements influence the perception of acoustic speech (Sams, Möttönen, & Sihvonen, 2005; Scott, Yeung, Gick, & Werker, 2013), and auditory feedback is thought to help establish a common articulatory representation that enables us to match AV-presented vowels (Skipper, van Wassenhove, Nusbaum, & Small, 2007). The auditory perception of speech furthermore activates cortical regions responsive to motor responses (Skipper et al., 2007; Okada & Hickok, 2009). While these findings are often explained as resulting from experience, similar findings are now being reported for infants even before they have gained sufficient experience to have learned to link auditory signals with specific self-generated motor movements. For example, Bruderer, Danielson, Kandhadai, and Werker (2015) found that infants' own oral-motor movements can influence their speech perception. Six-month-old English-learning infants completed a speech discrimination task (the Hindi retroflex-dental contrast) while having simultaneously given a teething toy that would control the movement and position of the infants' own tongue tip. Both the control group of infants who were not given the teething toy and the second control group of infants who were given a teething toy that

maintained a neutral tongue position discriminated the non-native contrast as expected in infants this age. However, the group of infants who were given a teething toy that restricts tongue-tip movement - the relevant oral-motor movements for production of the speech contrast - no longer showed evidence of discrimination.

The first experimental evidence that oral-motor perturbations can affect young infants' integration of AV speech was provided by Yeung and Werker (2013). In this study, two faces silently articulating /i/ and /u/ vowels were presented side by side, while an auditory recording of either /i/ or /u/ was played. Infants' lips were simultaneously manipulated to either match or mismatch the configuration needed to produce these phonemes – that is, the infant's lips were spread wide (as is needed for /i/) by placing a wooden teething ring in the mouth, or were caused to purse (as in /u/) by putting the tip of the caregiver's finger in the mouth. When the oral-motor configuration matched the auditory signal, infants preferred to look at the opposing visual configuration (e.g. the face producing /u/ if the infant is mimicking and hearing /i/). When the infants' lip movements contrasted with the heard speech, they preferred to look at the face with a matching visual configuration. This demonstration that specific sensorimotor configurations influence the matching of heard and seen speech suggests an early integration of the productive and perceptual speech systems.

To the extent that parallels can be drawn between bird song and human speech, one can speculate that just as vocalization experience during the sensorimotor phase is a crucial step in consolidating auditory sensory experience in swamp sparrows (Brainard & Doupe, 2002; see also Bolhuis, Okanoya & Scharff, 2010, it might also be a crucial step for human infants to have the opportunity to vocalize, or engage in oral-motor movements, to properly acquire and attune to the phonology of their native language. In the field of molecular genetics, the identification of a FOXP2 mutation – which causes speech apraxia, as well as impairments in expressive and

receptive language in humans – has allowed for an opportunity to study the neural bases and the

evolutionary origins of speech and language, as FOXP2 has been found to exist in non-animals

(Fisher & Scharff, 2009). In songbirds, FOXP2 is most ubiquitously expressed in young song

birds in the striatal region known to be important for song learning, and its knockout crucially

impairs the songbirds' ability to correctly and consistently learn the conspecifics' songs (Haesler

et al., 2007). While only speculative at this time, the work from bird song is a further impetus for

us to continue studying the influence of oral-motor movements.

Critical periods and multisensory development

While the phenomenon of perceptual narrowing and CP timing has been robustly

documented in the auditory domain, it is only recently that researchers have begun to study

perceptual attunement of visual or auditory-visual speech perception in typically developing

infants. These studies have revealed that perceptual attunement unfolds in the same time period

in visual-only and auditory-visual speech perception as it does in auditory-only speech

perception, and that, again, there may be CP-like effects. In Weikum et al. (2007), infants were

habituated to silent videos of a bilingual actor speaking either in French or English, whereby

only the visual information in the talking face was available to the infant. Following habituation,

infants were shown another silent video of the same actor speaking either the same or the other

language (e.g. a change from English to French). English-learning infants aged 4-6 months

looked longer to the talking faces when the language changed, but 8-month-old infants did not.

However, English-French exposed infants, who have had input from both languages, continued

to discriminate the languages visually even at 8 months of age. Thus, similar to the trajectory

observed in auditory speech discrimination, continued exposure (and/or bilingualism in general,

see Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012) is necessary for infants to

maintain perceptual sensitivity. Moreover, just as exposure to SRIs seems to accelerate the timing (i.e. it is apparent at a younger age) of perceptual attunement to non-native auditory speech contrasts, so did exposure to SRIs appear to accelerate the timing of perceptual attunement to visual language discrimination. And, just as exposure to maternal depression delayed the timing of attunement to auditory speech perception, it also did so for visual language discrimination (Weikum et al., 2012).

The process of perceptual attunement has also been documented for infants' sensitivities to AV speech. When 6-month-old Spanish-learning infants were habituated to either auditory /va/ or /ba/, a phonetic contrast not present in Spanish, they showed a preference to look longer towards the visual speech that matched what they heard; however, 11-month-old infants did not. Thus, by the time infants reach one year of age, they are no longer sensitive to AV matching in non-native speech (Pons et al., 2009; for related findings, see Danielson et al., 2017). In sum, multisensory speech perception also undergoes perceptual reorganization with experience. The similarity in the patterns of attunement to auditory, visual, and AV speech suggests that the developing phonological system may have a multisensory basis from the start. We expect that there may be a similar process of perceptual attunement to the sensorimotor information in speech, but of a somewhat different nature. In particular, motor influences on speech perception might be most important before and during the period of perceptual attunement, but no longer as influential once a native phonological representation is established. Work consistent with this possibility comes from an MEG study that demonstrates activation of motor areas (Broca's and cerebellum) in infants aged 7-months when listening to both native and non-native speech sound differences, but at 12-months showed activation in motor areas only when listening to non-native, and hence unfamiliar, speech (Kuhl, Ramírez, Bosseler, Lin, & Imada, 2014).

Of interest, similar patterns are seen in the perception of a manual language, sign language. At birth, infants show a preference for possible signs over non-sign gestures, but stop doing so by 9-10 months unless they are growing up with a sign language (Krentz & Corina, 2008). Moreover, young hearing infants discriminate linguistic signs in a categorical manner (Baker, Golinkoff, & Petitto, 2006) just as do signing adults (Baker, Idsardi, Golinkoff, & Petitto, 2005). However, hearing infants stop discriminating signs categorically by the end of the first year of life (Baker et al., 2006), with maintenance only if they have been exposed to a sign language as a full language (in this sample, one of the parents was a fluent signer who used sign with the infant) across the first year of life (Palmer, Fais, Golinkoff, & Werker, 2012).

IV. Long-term consequences of early linguistic representations

In the preceding sections we have described the developmental trajectory of the perceptual foundations of language acquisition within the first year of life from the perspective of CP processes, including both early sensory experiences and multimodal linguistic representations. From the initial linguistic establishment set in place during these early periods, continued acquisition and development of linguistic skills take place. Conversely, impoverished or compromised input both in timing and quality may have lasting consequences in later years. In the following paragraphs, we will briefly explore how changing phonological sensitivities – to phonetic category distinctions as well as to prosody – contribute to building a more complex linguistic system in the typically-developing individual.

Building up from the early signal

Languages are often described as having a kind of musical quality to them – we referred to this previously as the "melody of speech". This quality (termed prosody) is achieved through

a constellation of cues that include rhythm, duration, loudness, silence, and intonation. As described above, these cues are among the first to which infants are exposed, and among the first to which they demonstrate sensitivity. But the ability to perceive prosodic contours is not useful merely for discriminating one language from another; rather, prosodic chunks of speech correspond (albeit imperfectly) to higher-order linguistic structures like word category identity or constituent boundaries (Truckenbrodt, 1999; see Selkirk, 2011, for review). It has thus been proposed that the ability to discriminate prosodic cues is important for the acquisition of underlying syntactic structure (Morgan, 1986; Christophe, Guasti, & Nespor, 1997). Indeed, this hypothesis has been supported by demonstrations that infants' perception of prosodic cues leads to better word segmentation (Cristia & Seidl, 2011), phonotactic awareness (which in turn helps word segmentation) (Brent & Cartwright, 1996), and recognition of syntactic categories (Shi, Werker, & Morgan, 1999; Shi & Werker, 2001) – all within the first year of life. Two recent studies, moreover, revealed that infants can also rely on prosody to detect language-specific word order (Bernard & Gervain, 2012; Gervain & Werker, 2013). In the Gervain and Werker (2013) study, bilingual 7-month-olds who were learning languages with two different standard word orders (i.e. [object-verb] versus [verb-object]) were exposed to identical artificial languages and tested on their ability to extract the underlying word order. Languages with these two different word-order structures have been shown to use different prosodic strategies (i.e. differences in pitch or duration) for demarcating the constituent units in syntactic phrases. Bilingual infants exposed to a stream that used the prosodic strategy of languages with [object-verb] order segmented that language accordingly. Infants exposed to an identical stream, but one with the alternative prosodic strategy, segmented that language with a [verb-object] order.

While prosody facilitates segmentation and lays the groundwork for acquiring the native word order, acquiring the phonological categories of the native language works hand in hand

with pulling out, representing, and learning the meaning of specific words. Infants who show better phonetic discrimination at birth have larger vocabularies after 2 years of age (Molfese, 2000). Moreover, the timing of phonetic attunement predicts later vocabulary (Tsao, Liu, & Kuhl, 2004): infants with better native and worse non-native discrimination in the second half of the first year of life have larger vocabularies by 2 years. Again, while the establishment of native speech sound categories is fundamentally under maturational control (i.e. prematurity does not lead to earlier perceptual attunement; Peña et al., 2012), exogenous factors have been shown to influence both the production and perception of native speech sounds. Infants whose caregivers produce more infant directed speech are better able to discriminate consonant contrasts (Liu, Kuhl, & Tsao, 2003) and those whose caregivers engage in a high proportion of contingent communicative interactions show an accelerated timing of attunement (Elsabbagh et al., 2013). Contingent communicative interactions also impact the nature of young infants' babbles (Goldstein, King, & West, 2003; Goldstein & Schwade, 2008). Similarly, responsive maternal-infant speech interaction style facilitates infants' ability to pull out and recognize words in continuous speech (Thiessen, Hill, & Saffran, 2005; Bosseler, Teinonen, Tervaniemi, & Huotilainen, 2016; Floccia et al., 2016), skills that predict later vocabulary size (Junge, Kooijman, Hagoort, & Cutler, 2012; Marchman & Fernald, 2008; Newman, Rowe, & Ratner, 2016; Singh, Reznick, & Xuehua, 2012). Thus, speech perception development in the first year of life has a direct influence on later vocabulary (but see Cristia et al., 2014).

One means by which the relationship between phonetic development and later vocabulary size might be mediated is that phonological categories enable infants to recognize when various pronunciations constitute variations on a single word versus two different words (Dietrich, Swingley, & Werker, 2007). And indeed, toddlers (Bernhardt, Kemp, & Werker, 2007; Werker, Fennell, Corcoran, & Stager, 2002; also see Kemp et al., 2017, for more mixed results) –

including bilingual-learning toddlers (Fennell, Byers-Heinlein, & Werker, 2007) – who are better able to use native phonological categories to drive word learning are more advanced on many measures of later language development. Increasingly, researchers are beginning to also consider the extent to which a budding vocabulary drives speech perception development (e.g. see Swingley, 2009; Werker, in press); thus, the process may be bidirectional.

The relevance of phonological development to long-term language outcomes beyond the stage of word-learning has long been noted. As just one example: children who are better at isolating, manipulating, or rhyming speech sounds (i.e. "phonological awareness") experience a boost when learning to read (Anthony & Francis, 2005). This is especially true in languages with alphabetical writing systems such as English, where each symbol reflects (more or less) one phoneme; it has also been demonstrated, however, for children learning to read and write in syllabic-based systems of writing (Song, Georgiou, Su, & Hua, 2016; Shu, Peng, & McBride-Chang, 2007). While literacy is a relatively recent advance in human history, the readiness and capacity to read is of crucial importance in literate societies. To wit, children who are poor readers suffer in other areas of academic achievement and have lower rates of completing secondary schooling and enrolling in post-secondary education (Lesnick, Goerge, Smithgall, & Gwynne, 2010). It is important to note, however, that phonological awareness is but one predictor of eventual reading competence, which is a skill that also develops in response to children's lexical and syntactic development as well as their exposure to the many functions and types of literacy (Hoff, 2013).

Literacy is a challenging and complex skill to acquire, yet the majority of those who receive sufficient training become proficient readers and writers. A subset of the population, however, has immense difficulty in the acquisition of this skill, despite having normal intelligence, educational opportunities, and typical audition and vision. Dyslexia is a

developmental language disorder that involves impaired word recognition or fluency in reading, inaccurate decoding, and poor spelling. While a number of theories regarding the core processing deficits leading to dyslexia have been proposed, a deficit in phonological awareness has had the most consistent empirical support (Frost, 1998; Snowling, 2000; Ziegler & Goswami, 2005; Ramus & Szenkovits, 2008). Interestingly, there is also some evidence suggesting that children with developmental apraxia of speech suffer as well from an impairment in phonological awareness and reading skill (McNeill, Gillon, & Dodd, 2009).

Infants with severe hearing loss miss out enormously on the input that would otherwise be impacting their speech and language development (see Levine, Strother-Garcia, Golinkoff, & Hirsh-Pasek, 2016, for an elucidation), and without early intervention (either through auditory means such as hearing aids or cochlear implants, or through early introduction of a linguistic system such as sign language) are at high risk for life-long difficulties in speech and language processing (Ching et al., 2017). Less is known about the long-term impacts of deficits in multimodal speech perception. If it is established that the typical sensorimotor system is indeed important for speech perception, which in turn guides the establishment of the native phonological system, then congenital oral-motor deficits such as cleft palate and ankyloglossia (tongue tie) may also have implications for early acquisition of the native phonological units and later development of linguistic structures. Further research is needed to establish the degree to which the perceptual and productive speech systems guide one another for the acquisition of the native phonological system within the early developmental period.

V. Conclusion

Babies begin life with perceptual biases that prepare them for learning the native language. There are cascading critical periods in development during which the developing brain

is maximally open to specific kinds of linguistic sensory input. As such, under typical conditions, input during these periods has a particularly strong and lasting influence on linguistic perceptual organization and processing. This in turn has consequences for later aspects of language processing and use. In this review, we highlighted but a few of those known and potential lasting consequences, including effects on vocabulary acquisition, the acquisition of syntax, and later literacy. Because speech perception is not only auditory, but also involves other sensory modalities including visual and sensorimotor influences, it is essential in future work to characterize more deeply whether the critical periods in these other sensory domains emerge in parallel with those for auditory speech perception, whether they build on them, or whether they have somewhat distinct trajectories. One particularly important area of research is cross-modal plasticity, i.e. how a lack of input in one sensory modality can result in the neural areas that typically subserve that modality being used for processing input in a different modality (for reviews, see Bavelier & Neville, 2002; Merabet & Pascual-Leone, 2012).

An unambiguous implication of the cascading critical period(s) framework is that it is essential to ensure infants – from the first moments after birth – have the opportunity to experience rich linguistic input. The communicative exchanges that characterize parent-infant interactions provide not only the acoustic/phonetic information that enables infants to begin learning about their native language, but these interactions also provide rich visual and somatosensory support. With inadequate experience at the right points in development – whether through a noisy home environment, a sensory deficit in the infant, a perturbation in timing of infant brain development, or a paucity of interpersonal communicative exchanges – critical periods may close prior to the establishment of sufficiently precise representations, which can in turn place the child at risk for difficulties. With adequate experience, the infant is better

positioned to successfully establish the requisite foundations for successful language and literacy

acquisition.

References

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14,* 255-259.

Baker, S.A., Golinkoff, R. M., & Petitto, L. A. (2006). New insights into old puzzles from infants' categorical discrimination of soundless phonetic units. *Language Learning and Development, 2*(3), 147-162.

Baker, S. A., Idsardi, W. J., Golinkoff, R. M., & Petitto, L. A. (2005). The perception of handshapes in American Sign Language. Memory & Cognition, 33(5), 887-904.

Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: where and how? *Nature Reviews Neuroscience, 3*(6), 443-452.

Bernard, C., & Gervain, J. (2012). Prosodic cues to word order: what level of representation? *Frontiers in Psychology, 3,* 451.

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

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

Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience, 11*(11), 747-759.

Bosseler, A. N., Teinonen, T., Tervaniemi, M., & Huotilainen, M. (2016). Infant Directed Speech Enhances Statistical Learning in Newborn Infants: An ERP Study. *PloS One, 11,* e0162177.

Brainard, M. S., & Doupe, A. J. (2002). What songbirds teach us about learning. *Nature, 417,* 351-358.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61,* 93-125.

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., Mangin, J. F., (2009). Hearing faces: how the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience, 21,* 905-921.

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

Chang, E. F., & Merzenich, M. M. (2003). Environmental noise retards auditory cortical development. *Science, 300,* 498-502.

Ching, T. Y., Dillon, H., Button, L., Seeto, M., Van Buynder, P., Marnane, V., ... & Leigh, G. (2017). Age at intervention for permanent hearing loss and 5-year language outcomes. *Pediatrics, 140*(3), e20164274.

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,* e160660.

Choi, J., Broersma, M., & Cutler, A. (2017). Early phonology revealed by international adoptees' birth language retention. *Proceedings of the National Academy of Sciences, 114,* 7307-7312.

Christophe, T., Guasti, M., & Nespor, A. (1997). Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. *Language and Cognitive Processes, 12,* 585-612.

Cristia, A., & Seidl, A. (2011). Sensitivity to prosody at 6 months predicts vocabulary at 24 months. In *BUCLD 35: Proceedings of the 35th annual Boston University Conference on Language Development,* 145-156.

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting Individual Variation in Language From Infant Speech Perception Measures. *Child Development, 85,* 1330-1345.

Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., & Werker, J. F. (2017). The organization and reorganization of audiovisual speech perception in the first year of life. *Cognitive Development, 42,* 37-48.

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

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development, 9,* 133-150.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., & Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences, 103,* 14240-14245.

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

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience, 22,* 567-631.

Elsabbagh, M., Hohenberger, A., Campos, R., Van Herwegen, J., Serres, J., de Schonen, S., Aschersleben, G., & Karmiloff-Smith, A. (2013). Narrowing perceptual sensitivity to the native language in infancy: exogenous influences on developmental timing. *Behavioral Sciences, 3,* 120-132.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition, 152,* 101-107.

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

Feng, A. S., & Rogowski, B. A. (1980). Effects of monaural and binaural occlusion on the

    morphology of neurons in the medial superior olivary nucleus of the rat. *Brain*

    Research, 189(2), 530-534.

Fisher, S. E., & Scharff, C. (2009). FOXP2 as a molecular window into speech and

    language. *Trends in Genetics, 25*(4), 166-177.

Floccia, C., Keren-Portnoy, T., DePaolis, R. A., Duffy, H., Delle Luche, C., Durrant, S., White, L.,

    Goslin, J. & Vihman, M. (2016). British English infants segment words only with

    exaggerated infant-directed speech stimuli. *Cognition, 148,* 1-9.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: true issues and

    false trails. *Psychological Bulletin, 123,* 71-99.

Gerhardt, K. J., & Abrams, R. M. (1996). Fetal hearing: characterization of the stimulus and

    response. *Seminars in Perinatology, 20,* 11-20.

Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual

    infants. *Nature Communications, 4*.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing

    parallels between birdsong and speech. *Proceedings of the National Academy of Sciences,*

    100, 8030-8035.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid

    phonological learning. *Psychological Science, 19,* 515-522.

Grant, K.W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection

    of spoken sentences. *Journal of Acoustical Society of America, 108,* 1197-1208.

Greenough, W. T., Black, J. E., & Wallace, C. S. (1987). Experience and brain development. *Child*

    Development, 539-559.

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biology, 5*(12), e321.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8,* 393-402.

Hoff, E. (2013). Language development. Belmont, CA: Wadsworth Cengage Learning.

Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture: Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London B: Biological Sciences, 198*(1130), 1-59.

Hyltenstam, K., Bylund, E., Abrahamsson, N., & Park, H. S. (2009). Dominant-language replacement: The case of international adoptees. *Bilingualism: Language and Cognition, 12,* 121-140.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences, 1016,* 749-777.

Junge, C., Kooijman, J., Hagoort, P., & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science, 15,* 463-473.

Kemp, N., Scott, J., Bernhardt, B. M., Johnson, C. E., Siegel, L. S., & Werker, J. F. (2017). Minimal Pair Word Learning and Vocabulary Size: Links with Later Language Skills. *Applied Psycholinguistics, 38,* 289-314.

Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., Zhang, K., & Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science, 14,* 220-224.

Krentz, U., & Corina, D. (2008). Infant perception of American Sign Language and non-linguistic biological motion: The language instinct is not speech specific. *Developmental Science, 11,* 1-9.

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

Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J. F., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National Academy of Sciences, 111*(31), 11238-11245.

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

Lecanuet, J. P., Granier-Deferre, C., DeCasper, A. J., Maugeais, R., Andrieu, A. J., & Busnel, M. C. (1987). Fetal perception and discrimination of speech stimuli; demonstration by cardiac reactivity; preliminary results. *Comptes Rendus de l'academie des Sciences, Serie III, Sciences de la Vie, 305,* 161-164.

Lee, G. Y., & Kisilevsky, B. S. (2013). Fetuses respond to father's voice but prefer mother's voice after birth. *Developmental Psychobiology, 56,* 1-11.

Leroy, F., Glasel, H., Dubois, J., Hertz-Pannier, L., Thirion, B., Mangin, J. F., & Dehaene-Lambertz, G., (2011). Early Maturation of the Linguistic Dorsal Pathway in Human Infants. *Journal of Neuroscience, 31,* 1500-1506.

Lesnick, J., Goerge, R., Smithgall, C., & Gwynne, J. (2010). *Reading on grade level in third grade: How is it related to high school performance and college enrollment?* Chicago, IL: Chapin Hall at the University of Chicago.

Levine, D., Strother-Garcia, K., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Language

Development in the First Year of Life: What Deaf Children Might Be Missing Before

Cochlear Implantation. *Otology & Neurotology, 37*(2), e56-e62.

Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003) An association between mothers' speech clarity and

infants' speech discrimination skills. *Developmental Science, 6,* F1-F10.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in

infancy predict cognitive and language outcomes in later childhood. *Developmental Science,

11,* F9-F16.

Maurer, D., & Werker, J. F. (2013). Perceptual narrowing during infancy: A comparison of

language and faces. *Developmental Psychobiology, 56,* 154-178.

May, L., Gervain, J., Carreiras, M., & Werker, J. F. (2017). The specificity of the neural response to

speech at birth. *Developmental Science,* e12564.

McNeill, B. C., Gillon, G. T., & Dodd, B. (2009). Phonological awareness and early reading

development in childhood apraxia of speech (CAS). *International Journal of Language &

Communication Disorders, 44,* 175-192.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

Merabet, L.B., & Pascual-Leone, A. (2012). Neural reorganization following sensory loss: The

opportunity of change. *Nature Reviews Neuroscience, 11*(1), 44-52.

Molfese, D. L. (2000). Predicting dyslexia at 8 years of age using neonatal brain responses. *Brain

and Language, 72,* 238-245.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant

Behavior and Development, 16,* 495-500.

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel

perception after birth: A two-country study. *Acta Paediatrica, 102,* 156-160.

Morgan, J. L. (1986). *From simple input to complex gramma*r. Cambridge, MA: MIT Press.

Munhall, K. G., & Vatikiotis-Bateson, E. (2004). *Spatial and Temporal Constraints on Audiovisual Speech Perception*. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), The Handbook of Multisensory Processes (177-188). Cambridge, MA: MIT Press.

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2009). 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., & 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.

Newman, R. S., Rowe, M. L., & Ratner, N. B. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language, 43,* 1158-1173.

Oh, J. S., Au, T. K. F., & Jun, S. A. (2010). Early childhood language memory in the speech perception of international adoptees. *Journal of Child Language, 37,* 1123-1132.

Okada, K., & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience Letters, 452,* 219-223.

Pallier, C., Dehaene, S., Poline, J. B., LeBihan, D., Argenti, A. M., Dupoux, E., & Mehler, J. (2003). Brain imaging of language plasticity in adopted adults: Can a second language replace the first? *Cerebral Cortex, 13,* 155-161.

Palmer, S.B., Fais, L., Golinkoff, R.M., & Werker, J.F. (2012). Perceptual narrowing of linguistic sign occurs in the first year of life. *Child Development, 83*(2), 543-553.

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013).

Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences, 110,* 15145-15150.

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

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

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

Peña, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier speech exposure does not accelerate speech acquisition. *Journal of Neuroscience, 32,* 11159-11163.

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

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

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *The Quarterly Journal of Experimental Psychology, 61,* 129-141.

Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology-Human Perception and Performance, 22,* 318-330.

Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research, 23,* 429-435.

Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustical Society of America, 133,* EL286-EL292.

Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science, 23*(9), 994-999.

Selkirk, E. (2011). The syntax-phonology interface. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (2nd ed.) (pp .435-483).

Shahidullah, S., & Hepper, P. G. (1994). Frequency discrimination by the fetus. *Early Human Development, 36,* 13-26.

Shi, R., & Werker, J. F. (2001). Six-month-old infants' preference for lexical words. *Psychological Science, 12,* 70-75.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72,* B11-B21.

Shu, H., Peng, H., & McBride-Chang, C. (2007). Phonological awareness in young Chinese children. *Developmental Science, 11,* 171-181.

Shultz, S., Vouloumanos, A., Bennett, R. H., & Pelphrey, K. (2014). Neural specialization for speech in the first months of life. *Developmental Science, 17,* 766-774.

Simpson, K. L., Weaver, K. J., de Villers-Sidani, E., Lu, J. Y. F., Cai, Z., Pang, Y., Rodriguez-Porcel, F., Paul, I. A., Merzenich, M., & Lin, R. C. (2011). Perinatal antidepressant exposure alters cortical network function in rodents. *Proceedings of the National Academy of Sciences, 108,* 18465-18470.

Singh, L., Steven Reznick, J., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Developmental Science, 15*(4), 482-495.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing

    voices: how cortical areas supporting speech production mediate audiovisual speech

    perception. *Cerebral Cortex, 17,* 2387-2399.

Snowling, M. J. (2000). Dyslexia. Oxford, England: Blackwell Publishing.

Song, S., Georgiou, G. K., Su, M., & Hua, S. (2016). How well do phonological awareness and

    rapid automatized naming correlate with Chinese reading accuracy and fluency? A meta-

    analysis. *Scientific Studies of Reading, 20,* 99-123.

Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate

    mechanisms and experience. *Nature, 259,* 39-41.

Sugden, N. A., Mohamed-Ali, M. I., & Moulson, M. C. (2014). I spy with my little eye: Typical,

    daily exposure to faces documented from a first-person infant perspective. *Developmental*

    Psychobiology, 56, 249-261.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. The

    Journal of the Acoustical Society of America, 26, 212-215.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical*

    Transactions of the Royal Society B, 364, 3617-3622.

Takesian, A. E., & Hensch, T. K. (2013). Balancing plasticity/stability across brain development.

    Progress in Brain Research, 207, 3-34.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word

    segmentation. *Infancy, 7,* 53-71.

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

    Development, 47, 466-472.

Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological

    phrases. *Linguistic Inquiry, 30,* 219-255.

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development, 75,* 1067-1084.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, 102,* 1181-1186.

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

Vouloumanos, A., & Werker, J. F. (2007b). Why voice melody alone cannot explain neonates' preference for speech. *Developmental Science, 10,* 169-171.

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

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

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual Language Discrimination in Infancy. *Science, 316,* 1159.

Werker, J. F. (in press). Perceptual beginnings to language acquisition. *Applied Psycholinguistics.*

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

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

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

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.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect

    phonetic perception? The precocious case of lexical tone. *Journal of Memory and*

    Language, 68, 123-139.

Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech

    perception. *Psychological Science, 24,* 603-612.

Zhang, L. I., Bao, S., & Merzenich, M. M. (2001). Persistent and specific influences of early

    acoustic environments on primary auditory cortex. *Nature Neuroscience, 4,* 1123-1130.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled

    reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131,*
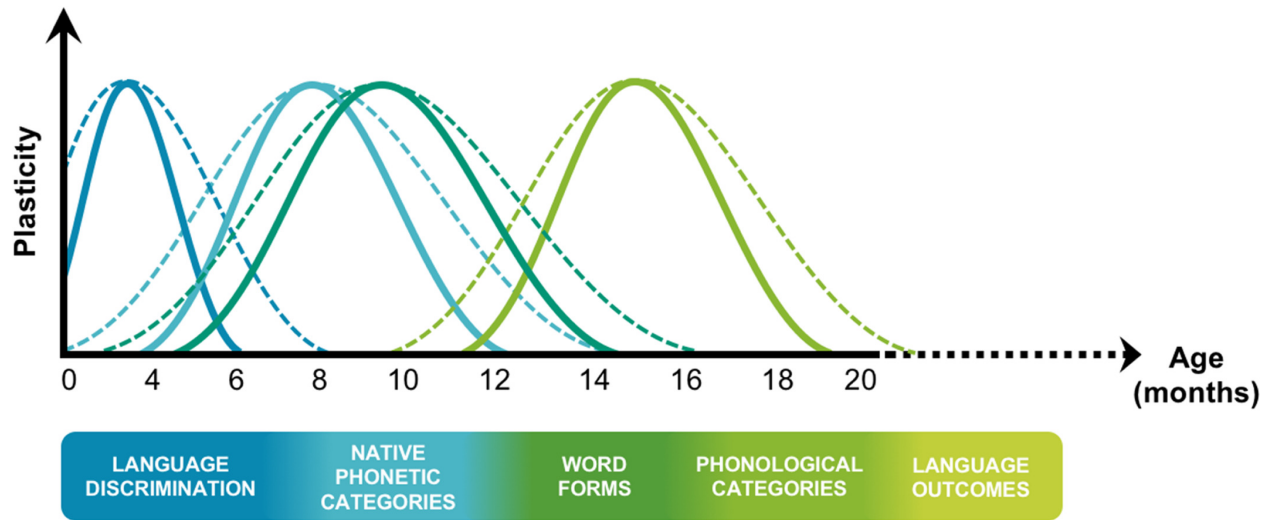
    3-29.

Figures

Figure 1.

Figure Caption

Figure 1. An illustration of cascading and mutually influential critical periods (CPs) in language development across the first months of life. CPs are represented with curves to illustrate periods of maximal openness (plasticity) to experiential input for particular linguistic sensitivities. The typical timing of a CP (solid curved lines) can be altered (dotted curved lines) by the properties of the environmental context. Note that the CP curves are overlapping to various degrees, illustrating how they cascade rather than occur in a step-wise fashion. The linguistic sensitivities themselves are labelled within boxes below the CP curves and are represented with merging colours to illustrate their mutually-influential nature: while experiential attunement at one stage influences later acquisition, it is also the case that attunement of later components can exert influence on previously-established representations. For example, just as an infant's attunement to native phonetic categories (light blue) can facilitate their identification of word forms (dark green) via speech segmentation, learning to segment word forms can also sharpen the boundaries of their native phonetic categories.