Chapter 8
Development of Speech Perception

Judit Gervain

Abstract Infants start their journey into language as universal listeners but by the end of the first year of life they become native language experts, as their perceptual systems and brains attune to the sound patterns of their native language(s). This chapter describes this attunement process and its neural correlates. Speech is the auditory medium that allows us to externalize language. Speech perception and language acquisition are thus tightly connected, especially during development. While focusing primarily on the development of speech perception, this chapter, therefore, necessarily touches upon the growth of language more generally. It discusses the major milestones of this developmental trajectory in chronological order, starting out with prenatal experience and newborns' speech perception abilities, and following the attunement process in phoneme and tone perception during the first year of life, early word learning and the prosodic bootstrapping of grammar during the toddler years.

Keywords Newborn · Infant · Prenatal experience · Postnatal experience · Perceptual attunement · Perceptual reorganization · Native language · Critical period · Neural plasticity · Language input

8.1 Introduction

Speech perception undergoes dramatic changes during the first years of human development. Infants are born with speech perception abilities that allow them to acquire any of the world's languages. After months of experience, these initially broadly based, universal abilities get tuned to the native language(s). This attunement process implies a reorganization and/or narrowing of perceptual categories, with the maintenance or refinement of native sound categories, and a loss or decrease

---

J. Gervain (✉)
DPSS, Università degli Studi di Padova, Padova, Italy

INCC, CNRS and Université de Paris, Paris, France
e-mail: judit.gervain@unipd.it

L. L. Holt et al. (eds.), *Speech Perception*, Springer Handbook of Auditory Research 74, https://doi.org/10.1007/978-3-030-81542-4_8

in sensitivity to non-native ones. At the neural level, it is accompanied by an increasingly focal brain specialization for native language processing.

This chapter describes this language attunement process and its neural correlates. Importantly, speech is the auditory medium that allows us to externalize language. Speech perception cannot thus be described without reference to language, the representation and rule system humans possess. This is particularly true in development. Hearing infants only have access to speech to learn language; they receive no formal or explicit instructions about the words or rules of their native languages. Yet, they successfully acquire the lexicon and the grammar of their mother tongue in the span of just a few years with amazing ease and efficiency. This fact has led to the assumption that the sound patterns of language are intimately intertwined with the other levels of language such as grammar and lexicon. Correlations between the sound patterns and abstract linguistic regularities allow infants to use speech to learn about or "bootstrap" the grammar and the lexicon (Morgan and Demuth 1996). These abstract acquisitions in turn help infants further fine-tune their perception of the speech signal.

Speech perception and language acquisition are thus tightly connected and interact synergistically. Empirical evidence for these connections is steadily increasing (Werker 2018; Swingley 2021). Given this interactive view of speech and language, this chapter, while focusing primarily on the development of speech perception, will necessarily touch upon the growth of language more generally.

The ultimate mechanisms of language development have long been debated, with some theories arguing for genetically endowed factors (Lenneberg 1967; Chomsky et al. 2002), and others for experiential and learning-based ones (Elman et al. 1997; Tomasello 2000). With the advent of brain imaging and especially genetic and epigenetic studies (Werker and Hensch 2015), it is becoming increasingly clear that biologically endowed and experiential factors are likely to act synergistically and rely on each other to bring about language development. The strict binary dichotomy of the traditional nature-nurture debate is thus replaced by a more integrative view of the factors that contribute to the developmental changes in speech perception and language acquisition.

Related to this new perspective, the notion of critical periods in speech perception and language acquisition has been revisited. The original proposal (e.g., Lenneberg 1967) was based on observations about language development failing to reach native-like competence when acquisition starts late, typically after puberty. One example comes from cases of feral children. These children are raised in social deprivation and thus not spoken to. They only recover language if they are discovered and introduced to language before puberty (Curtiss et al. 1974). Immigrants to the USA constitute another example. They have been observed to achieve native competence in English if they arrived before age 8–10 years (Johnson and Newport 1989). But since language learning remains possible throughout the life span, with large individual variations in outcome, the notion of critical periods was sometimes debated. With a better understanding of the experiential, molecular, and neural mechanisms controlling critical period phenomena, both in humans (Weikum et al. 2012; Gervain et al. 2013) and in animals (Weaver et al. 2004; Hensch 2005), where invasive studies can be carried out to close or re-open critical periods, the notion of critical periods has taken on a new, biologically better-defined meaning. How brain plasticity changes during speech perception and language development as a result of the closure of the critical period has thus recently received considerable attention (Werker and Hensch 2015).

This chapter follows the development of speech perception chronologically. It starts by reviewing newborn infants' universal abilities and then following how these abilities narrow down and attune to the native language. Such attunement processes operate at different levels of phonological organization, from global ones such as rhythm to smaller units such as phonemes, syllables, tones, or words.

8.2 Newborns' Speech Perception Abilities

In the light of the theoretical debates on the role of innate and learned factors in language acquisition, newborn infants' abilities have received considerable attention. These abilities were viewed as the best approximation we can methodologically get of the "initial state," that is, the state of the perceptual and language learning system before experience begins. Since then, evidence has gathered that fetuses learn from the speech input they receive in utero, as hearing becomes operational between the 24th and 28th week of gestation (Eggermont and Moore 2012). Newborns thus show universal, broadly based speech perception abilities not specific to any language yet, as well as abilities that are already shaped by prenatal experience.

8.2.1 Newborns' Universal Speech Perception Abilities

The auditory system is immature at birth and continues developing into late childhood/early adolescence (Moore 2002; Eggermont and Moore 2012). Yet, newborns show a variety of speech perception abilities, many of which are universal and broadly based, allowing newborns to discriminate most sound patterns found in the world's languages and thus enabling them to start learning any language.

Newborns' first task is to identify speech among the sounds present in their environment. Newborns and 2-month-old[1] infants can indeed recognize speech, and show a strong preference for it over equally complex sine wave analogs (Vouloumanos and Werker 2004). However, the category "speech" may be relatively broad at birth, roughly corresponding to primate vocalizations, as newborns show equal preference for human speech and rhesus monkey vocalizations (Vouloumanos et al. 2010). Yet,

---

[1] Throughout this chapter, the ages specified indicate the ages of infants tested in the cited studies. While an individual infant of a given age may not show a specific ability, on average, infants as a group do so at the age indicated.

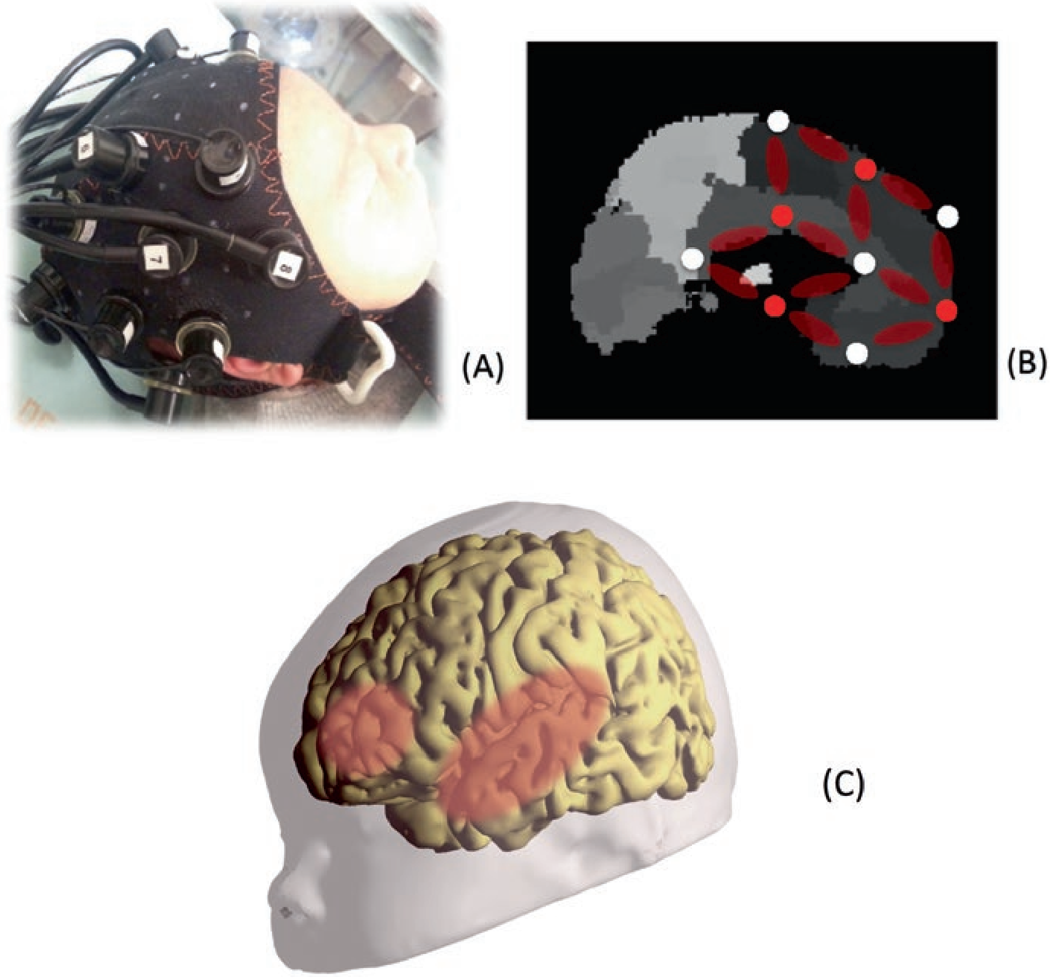

Fig. 8.1 Language in the newborn brain. (**a**) A near-infrared spectroscopy (NIRS) brain imaging cap on a newborn infant head, and (**b**) the corresponding sensor space overlaid on a newborn structural scan. NIRS is a commonly used imaging technique to localize the language network in infants' and young children's brains. (Images adapted from Abboub et al. (2016)). (**c**) A schematic illustration of the brain areas that have been found to be involved in a variety of speech and language processing tasks in young infants using brain imaging

by 3 months, infants show a unique preference for speech over both sine wave analogs and monkey calls (Vouloumanos et al. 2010).

Analogously with this behavioral preference for speech, the brains of young infants are specialized for speech processing. Three-month-old infants, full-term neonates, and even premature newborns activate similar brain networks as adults (the superior and middle temporal gyri, the inferior parietal cortex, and the inferior frontal gyrus, including Broca's area; Fig. 8.1) in response to language, but not to non-linguistic controls such as backward speech (Dehaene-Lambertz et al. 2002; Peña et al. 2003; Mahmoudzadeh et al. 2013). As discussed in Sect. 8.2.2, prenatal experience may already shape this specialization.

In addition to identifying speech in their environment, newborns are able to discriminate languages from one another, even if they never heard them before, on the basis of the languages' different rhythms (Mehler et al. 1988; Nazzi et al. 1998). Language rhythm was first quantified along three acoustic dimensions (Ramus et al. 1999): %V, which is the relative proportion of vowels in the speech signal as well as ΔC and ΔV, which are the variability in the length of consonant and vowel clusters, respectively. In the space defined by these variables (Fig. 8.2), languages cluster together into what was traditionally called the rhythm classes of languages. While language rhythmic is best understood as a continuum (Nespor 1990), the classes are still often used. They are named after the time unit that was once believed to be the basic isochronous element in the languages belonging to a given class (Abercrombie 1967). Japanese is thus a mora-timed language the mora is a unit larger than the phoneme, but smaller than the syllable). (Mora-timed language have the highest %V and the lowest ΔC values. Syllable-timed languages, such as French or Italian, still have relatively high %V, but medium ΔC values. Stress-timed languages such as English and Polish, in which the unit was believed to be the interval between stressed syllables, have lower %V and higher ΔC. Subsequently, other metrics have also been proposed to quantify rhythm (Grabe and Low 2002; Dellwo 2006). They are better at accounting for speech rate differences across speakers.

Rhythmic discrimination does not require familiarity with the languages. Newborns prenatally exposed to French are able to discriminate between English and Japanese, for instance, as can tamarin monkeys (Ramus et al. 2000). This

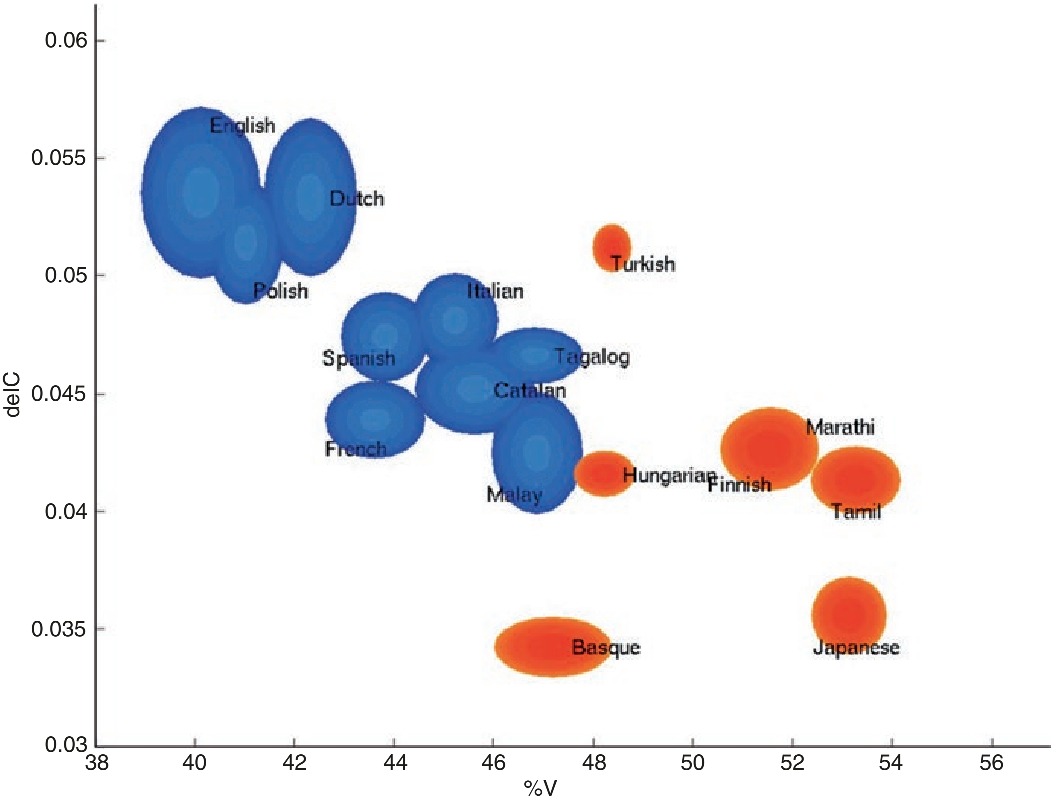

Fig. 8.2 Different languages in the space defined by %V and ΔC, two measures of speech rhythm. (Adapted from Mehler et al. (2004))

finding suggests that rhythmic discrimination might be a general property of the primate or mammalian auditory system, independent of experience with language or the ability to acquire it.

One important implication of newborns' ability to discriminate languages on the basis of rhythm is that infants born into a multilingual environment can immediately detect that they are being exposed to different languages, at least if the languages are rhythmically different. Bilingual newborns have indeed been shown to be able to discriminate their two languages from a third, rhythmically different language (Byers-Heinlein et al. 2010).

In addition to their abilities to identify speech in different languages in their environment, newborns are also able to process smaller units within the speech signal. Behavioral results show, for instance, that infants readily detect the acoustic cues correlated with the beginnings and ends of words (Christophe et al. 1994). They have also been found to be sensitive to syllables within words (Sansavini et al. 1997), readily discriminating words in which the stress is on the first syllable (e.g., ***doc****tor*) vs. those in which it is on the final one (e.g., *gui****tar***). Interestingly, however, infants cannot tell apart words with different numbers of phonemes if the number of syllables is the same.

During the first months of life, infants can also discriminate many of the phonemes appearing in the world's languages, as has been shown both behaviorally (Eimas et al. 1971) and electrophysiologically (Dehaene-Lambertz and Baillet 1998). This universal discrimination repertoire is one of the hallmarks of young infants' broad-based abilities, allowing them to learn any language to which they are exposed. Interestingly, chinchillas and songbirds can also discriminate phonemes at similar acoustic boundaries (Kuhl 1981, 1986), suggesting that phoneme perception builds on evolutionarily available perceptual abilities. It is, therefore, available to young infants prior to experience. How this ability is then shaped by language experience will be discussed in Sect. 8.3.

What features of the acoustic signal of speech newborns rely on to discriminate phonemes is only now starting to be investigated. When processing speech presented in silence (Chap. 4, Tune and Obleser; Chap. 7, Ullas, Bonte, Formisano, and Vroomen), adults can discriminate phonemes even on the basis of a strongly impoverished speech signal retaining only the slowest modulations (<16 Hz) of the amplitude envelope (Drullman 1995; Shannon et al. 1995), mimicking the signal available to cochlear implant users. A brain imaging study (Cabrera and Gervain 2020) investigated how newborns process consonant contrasts in three acoustic conditions. One condition consisted of the intact speech signal. In the second condition, the full envelope was preserved, but the temporal fine structure was suppressed. In the third condition, only the slowest modulations of the envelope were preserved. This study showed that newborns were able to discriminate consonants in all three conditions, suggesting that, like for adults, the slowest modulation cues of the speech envelope are sufficient for young infants to process the finest details of the speech signal. Interestingly, however, the three conditions activated different brain areas, suggesting early neural specialization for different aspects of the speech signal. Specifically, the condition containing the full envelope evoked a more left-lateralized activation than the slow envelope condition, suggesting adult-like brain specialization for the different aspects of the speech signal early in life.

Newborns are sensitive not only to sound patterns, but also to structural regularities in the speech input. Thus, they can detect repetition-based patterns such as ABB (e.g., “mu ba ba,” “pe na na,” etc.) or AAB (e.g., “ba ba mu,” “na na pe,” etc.), and discriminate them from otherwise similar random sequences such as ABC (e.g., “mu ba ge,” “pe na ku,” etc.; Gervain et al. 2008), or from one another (e.g., ABB vs. AAB; Gervain et al. 2012). Furthermore, this ability involves the bilateral temporal and left frontal areas, including Broca’s area, implying that the infant language network is already similar to the adult one.

In sum, newborns already possess a repertoire of basic auditory, speech perception, and learning mechanisms, many of them shared with chinchillas or songbirds, that allow them to crack the linguistic code in any language they encounter in their environment, independently of prenatal speech experience.

8.2.2 *Newborns’ Speech Perception Abilities Shaped by Prenatal Experience*

A growing number of studies suggests that newborns also have abilities shaped by experience with speech heard in the womb in addition to their universal perceptual sensitivities. Auditory experience with speech starts in the womb. But the intrauterine speech signal is different from the signal heard outside of the womb. Maternal tissues and the amniotic fluid act as low-pass filters at about ~400–800 Hz, although the exact values can only be estimated from computational simulations and recordings in pregnant sheep models (Gerhardt et al. 1990; Lecanuet and Granier-Deferre 1993; DeCasper et al. 1994). As a result of this low-pass filtering, the melody and rhythm of speech, which jointly define the prosody of a language, are preserved. At the same time, the fine details necessary to identify individual sounds, especially consonants, are suppressed. As a result, words are mostly unintelligible. Infants’ first experience with speech thus consists mainly of prosodic information (Gervain 2015, 2018).

This prenatal experience already shapes fetuses’ speech perception abilities. Newborns recognize and prefer their mother’s voice over other female voices (DeCasper and Fifer 1980). They also show a preference for their native language over other languages (Mehler et al. 1988; Moon et al. 1993) and a story heard frequently in the womb over other stories (DeCasper et al. 1994; Kisilevsky et al. 2009).

Relevant to language acquisition, fetuses learn even more specific details about their native language. Since vowels have the highest energy in the speech signal and are the main carriers of prosody, some vowels seem to be already learned in part prenatally. Indeed, newborns show a preference for a vowel they did not hear prenatally over a native one (Moon et al. 2013). Fetuses can also learn word-level prosodic information (Partanen et al. 2013), readily detecting a change in lexical pitch trained prenatally, which untrained newborns do not recognize.

Infants also show evidence of learning prenatally about the prosody of larger linguistic units, such as utterances. Languages vary as a function of what acoustic cues mark prosodic prominence in their phonological phrases. In some languages, such as French or English, prominence is carried by a durational contrast, meaning that the prominent element is lengthened as compared to the non-prominent one (e.g., in the phrase *to Rome*, the vowel of the prominent content word *Rome* is longer than the vowel of the non-prominent preposition). In these languages, the prominent element typically occupies a phrase-final position, so phrases have an iambic prosodic pattern.

In other languages, like Japanese or Turkish, prominence is indicated by a pitch/intensity contrast. In these languages, prominence is phrase-initial (i.e., trochaic), so the higher or louder element is at the phrase onset (e.g., in the Japanese phrase, *Tokyo kara*, which literally translates as “Tokyo to” and means “to Tokyo,” the first vowel of *Tokyo* is higher than the vowel of the word *kara*). This alternation of prominent and non-prominent elements creates a rhythmic prosodic pattern readily perceivable even by listeners who are unfamiliar with a given language (Langus et al. 2016).

Newborn infants also seem to pick up on this pattern from their prenatal exposure (Abboub et al. 2016). Newborns were presented with pairs of pure tones contrasting in duration, pitch, or intensity. In one condition, the pairs were consistent with the patterns found in natural languages. Specifically, they were iambic pairs (e.g., short-long) for the durational contrast, like in the English example *to Rome*, and trochaic pairs for the pitch (e.g., high-low) and intensity contrasts (e.g., loud-soft), like in the Japanese example *Tokyo kara*. In the other condition, the pairs were inconsistent with these patterns, so trochees (e.g., long-short) for duration, iambs for pitch/intensity (e.g., low-high/soft-loud). The newborn brain showed a greater response to the inconsistent patterns, but only for the acoustic cue that marks prosodic prominence in the language the infants were exposed to prenatally.

Newborns’ knowledge of the native prosody might even go beyond perception. It has been suggested that newborns’ communicative cries reproduce the prosodic patterns of their native language (Mampe et al. 2009). Indeed, German newborns’ cry patterns were found to have initial prominence, just like typical declarative utterances in German do. By contrast, French babies’ cries were prominence-final mimicking the prosodic contour characteristic of French utterances. Recently, these findings received some criticism on the basis of the statistical analyses used (Gustafson et al. 2017). But in a subsequent study, automated classification algorithms could separate cries from French-, Arabic-, and Italian-exposed newborns according to native language (Manfredi et al. 2019). If confirmed to be true, these findings would indicate that prenatal experience is sufficiently strong to shape even production.

Prenatal experience also shapes the brain specialization for language processing. Newborns’ brain responses to speech in the native language are different from responses to non-native languages. Some studies find stronger left-lateralization for the native language played forward than backward. The response involves the same regions as in adults, mainly the middle and superior temporal areas and the inferior frontal regions, including Broca's area (Peña et al. 2003). When directly comparing neonates' responses to their native language vs. a non-native tongue, some studies reproduced the left hemisphere advantage for forward vs. backward speech in the native language, but no hemispheric difference in a non-native language (Sato et al. 2012; May et al. 2017). However, other studies found no hemispheric differences for either language, but an overall advantage for the native language over the non-native one (May et al. 2011). The lateralization issue notwithstanding, all studies found a difference between the responses to the native language and unfamiliar languages, strongly suggesting that the brain network for speech processing is sculpted by prenatal experience. Furthermore, this network is already specialized for processing speech, as a whistled language does not activate it despite being a human communication system (May et al. 2017).

In sum, despite their immature auditory system, newborns show sophisticated speech perception abilities. Some of these abilities are universal, allowing infants to start acquiring any language. Others, by contrast, are already tuned to the prenatal experience with speech, especially prosody, infants received in the womb.

8.3 Perceptual Attunement to the Native Language

After birth, experience with the full-band speech signal begins and infants start to learn about the sound patterns specific to their native language(s). The experience induces a perceptual reorganization or attunement to the native language, whereby the ability to discriminate linguistic contrasts found in the language(s) heard is maintained or even improved, while the ability to distinguish most contrasts that do not appear in the input decreases (Werker and Tees 1984; Kuhl 2004). This reorganization may show different developmental trajectories in different areas of language. In some, a simple decrease in non-native discrimination (with a concomitant improvement in native discrimination) is observed (Werker and Tees 1984). Other areas are characterized by a U-shaped trajectory, where after the initial ability to discriminate certain contrasts and a subsequent decline, the ability re-emerges (Weikum et al. 2007, 2013). This newly emerging ability is sometimes underpinned by mechanisms that are different in nature than those underlying the initial ability. The initial ability is acoustic, closely linked to acoustic discriminability, whereas the emerging one is shaped by native language experience.

Attunement to the native language comes about as an intricate interplay between experience and perceptual/cognitive mechanisms. Attunement is accompanied by reorganization at the neural level, with increasingly focal, lateralized brain specialization for native language processing. This, in turn, is tied to developmental changes in brain plasticity, brought about by the changing balance of inhibitory and excitatory connections, ultimately linked to synaptogenesis, myelination, and pruning (Casey et al. 2000; Tierney and Nelson 2009; Haartsen et al.

2016)—neurophysiological mechanisms that are particularly active between the prenatal period and adolescence (although they remain operational throughout the lifespan).

It is not surprising, therefore, that attunement to experience is not unique to speech and language. Similar phenomena have been observed in other perceptual domains, for instance, in face perception (Pascalis et al. 2002; Maurer and Werker 2014).

The general principle of attunement to the native language(s) notwithstanding, different areas of speech perception undergo different narrowing trajectories, and some non-native contrasts, such as click consonants or some tonal contrasts, remain discriminable throughout life. The following sections discuss each of these developmental trajectories in turn.

8.3.1 Linguistic Rhythm

The rhythmic discrimination ability observed in newborns provides a good explanation of how bilinguals exposed to rhythmically different languages may distinguish their languages from birth. However, some bilinguals are exposed to rhythmically similar languages, and newborns cannot discriminate these from one another at birth. From what age and on what basis do bilinguals of rhythmically similar languages start distinguishing between their mother tongues? Bilingual infants growing up with Spanish and Catalan, two rhythmically similar languages, were found to succeed on this discrimination task at 4 months of age (Bosch and Sebastian-Galles 1997). Monolingual Spanish and monolingual Catalan infants also performed similarly. Basque-Spanish bilinguals were also shown to distinguish the two languages between 3.5 and 4 months (Molnar et al. 2013). While monolingual Basque infants behaved similarly, interestingly, the monolingual Spanish infants in this study only discriminated the two languages when habituated to Basque, but not when habituated to Spanish. This asymmetry may be related to the geopolitical dominance of Spanish in the Spanish Basque Country, the location of the study.

Taken together, the above results suggest that familiarity and experience with at least one of the languages allow discrimination even within the rhythmic group after 3–4 months of experience. Specifically, this discrimination ability may rely on familiarity with the phoneme repertoire, syllable structure, and/or phonotactic regularities of at least one of the languages.

8.3.2 Audio-Visual Speech Perception

Speech is not only heard, but also seen. A considerable amount of visual information is available in the speaker's face/head when producing speech. This information includes the position and movement of the lips, and the tongue, as well as of the eyes, eyebrows, and head, about the global features of different languages, their prosody, as well as individual phonemes (Guellaï et al. 2014; Wagner et al. 2014; de la Cruz-Pavía et al. 2020a). Adults have been shown to readily integrate such visual information with the auditory signal while processing speech (McGurk and MacDonald 1976). They can also use it to discriminate languages presented only visually (Soto-Faraco et al. 2007; Weikum et al. 2013). Furthermore, visual information also supports and augments speech perception in a non-native language or when the signal is degraded (Birulés et al. 2020). It also helps listeners segment out words from continuous speech (Mitchel and Weiss 2014) or parse the speech input to larger prosodic units (de la Cruz-Pavía et al. 2020b).

How infants use the visual correlates of speech has received increasing attention. Both monolingual and bilingual infants can readily discriminate two languages on the basis of visual speech alone at 4 and 6 months if at least one is their native language. By 8 months, only bilingual infants continue to do so (Weikum et al. 2007). This suggests that maintaining visual sensitivity helps infants in their daily task of discriminating between their two languages, a challenge that monolinguals do not face. Interestingly, this maintained perceptual sensitivity is general, as familiarity with the languages is not necessary to show successful discrimination. Indeed, both English-French and Spanish-Catalan bilinguals discriminate visual French and visual English at 8 months (Sebastián-Gallés et al. 2012). During the first 6 months of life, while their audio-visual sensitivity to speech is still broadly based, infants can also match talking faces to speech in languages that are unfamiliar to them. This ability weakens by 12 months of age when speech is adult-directed, showing perceptual narrowing (Kubicek et al. 2014a). Interestingly, 12-month-olds still succeed if the auditory stimuli used are infant-directed (Kubicek et al. 2014b).

The prosody of speech also has its visual correlates: speakers of Japanese and English produce eyebrow movements and head nods to mark phrase boundaries (de la Cruz-Pavía et al. 2020a), which adult listeners can use, in conjunction with auditory information, to parse speech into phrasal units (de la Cruz-Pavía et al. 2019). Eight-month-old, but not yet 4-month-old infants, also start to show sensitivity to these visual cues, and can integrate them with auditory prosodic information and word frequency. However, this integration process is not yet adult-like, in particular in the temporal asynchrony that infants expect and tolerate between the different cues (de la Cruz-Pavía et al. 2019).

Like in adults, infants' perception of speech in noise improves when they are provided with additional visual information (Hollich et al. 2005). A large body of work indicates that this facilitatory effect is based on infants' ability to match the auditory and visual signals at the syllable/phoneme level. Infants, for instance, can choose which of two silently talking faces articulates a syllable heard auditorily (Kuhl and Meltzoff 1982; MacKain et al. 1983; Patterson and Werker 1999, 2002). Audio-visual matching also undergoes perceptual narrowing, similarly to auditory phoneme perception (see Sect. 8.3.3). By 11–12 months of age, infants no longer match the auditory and visual signals of phonemes if those are not found in their native language (Pons et al. 2009; Danielson et al. 2017).

Interestingly, what cues infants rely on in a talking face also changes during development, reflecting the underlying perceptual reorganization. While infants and adults mainly look at the eyes of a (talking) face (Hunnius and Geuze 2004; Viola Macchi et al. 2004), around 8–12 months of age, infants shift their attention to the mouth region, and this shift is more pronounced if infants hear non-native speech (Lewkowicz and Hansen-Tift 2012; Kubicek et al. 2013) or if they are bilingual (Pons et al. 2015). This shift corresponds to the developmental timeline of perceptual narrowing to the native language, and might thus reflect infants' strategy to seek out visual information that supports the attunement process. Indeed, by 12 months of age, infants only look to the mouth region when hearing non-native speech, but not when hearing their native language (Lewkowicz and Hansen-Tift 2012). This audio-visual reorganization may be a crucial milestone in speech perception, as children with language impairment show reduced attention to the mouth (Pons et al. 2018).

8.3.3 *Phoneme Perception*

Very young infants, up to about 4–6-months of age, can discriminate almost all phonemes appearing in the world's languages, even those that do not appear in their native language and that adult speakers of a different language are unable to discriminate, as has been shown both behaviorally (Eimas et al. 1971; Werker and Curtin 2005) and electrophysiologically (Dehaene-Lambertz and Baillet 1998; Kujala et al. 2004). Infants' phoneme perception, like that of adults, is categorical, especially for consonants, possibly less so for vowels (Swingley 2021). Perception is categorical when a given acoustic difference between two sounds is discriminated and treated as contrastive if it spans a phoneme boundary, but not discriminated if it falls within the boundaries of a phoneme category (even though infants are able to perceive the acoustic difference; McMurray and Aslin 2005). This universal discrimination repertoire is one of the hallmarks of young infants' broad-based abilities, allowing infants to learn any language they are exposed to.

After several months of experience with the native language, non-native sound discrimination declines (Werker and Tees 1984), while the discrimination of contrasts found in the native language is maintained or even improves (Kuhl et al. 2006; Narayan et al. 2010). This perceptual attunement toward the native sound repertoire takes place around 4–6 months for vowels (Kuhl et al. 1992) and 10–12 months for consonants (Werker and Tees 1984). The system nevertheless remains plastic for several years after attunement. It is thus possible to learn the phoneme inventory of another language until age 6–8 years (or the onset of puberty the latest), as studies with immigrants (Johnson and Newport 1989) and international adoptees suggest (Ventureyra et al. 2004; Pierce et al. 2014). Infants growing up multilingually go through the same perceptual narrowing, although for some sounds, they also show different developmental patterns (Byers-Heinlein and Fennell 2014). For instance, when a phoneme pair is distinguished in one of their languages, but not in the other, bilingual infants may go through a phase when they do not discriminate between the two sounds.

Interestingly, the ability to discriminate non-native contrasts does not always get lost. Some features of click sounds, found for example in the African language Zulu, remain discriminable to non-native adults (Best et al. 1988). This has been explained by the unusual, almost non-linguistic nature of these sounds.

Phoneme discrimination may be facilitated by systematic associations between sounds and objects, implying that the relationship between phoneme perception and word learning is mutual (Werker and Yeung 2005). Thus, 9-month-old infants can successfully discriminate a non-native sound contrast if each phoneme occurs in a nonword that is associated with an object (Yeung and Werker 2009), whereas at this age, they would already fail without the association with objects, due to perceptual attunement.

This relationship between word learning and perceptual attunement notwithstanding, the lexicon is still relatively small between 4 and 12 months of life, when perceptual attunement takes place. To explain how phonetic perception changes without a sizeable lexicon, different mechanisms based on similarity-matching and distributional learning have been proposed (Kuhl 2004; Werker and Curtin 2005). These models assume that the distributional characteristics of different phonemes in the speech signal reflect those perceptible differences that are contrastive in the language and de-emphasize differences that are not.

Existing results also point toward another factor in the development of early phoneme perception, the contribution of the motor system. In adult speech perception, a long tradition has argued for the key role of the motor system in phoneme perception (e.g., Liberman and Mattingly 1985). According to the motor theory of speech perception, the motor schemes necessary to produce a speech sound play an important role in its identification, when the sound is perceived. Whereas the original conception of a necessary role for the motor system in speech perception is not supported empirically, there is a strong case to be made for the interplay of speech perception and production in adults (Hickok et al. 2003).

This theory received relatively little attention in developmental research, since infants' motor and production skills so clearly lag behind their perceptual skills, although correlational evidence between infants' babbling/production and phoneme perception abilities has been reported (Guellaï et al. 2014; Majorano et al. 2014; Vilain et al. 2019). However, a study by Bruderer et al. (2015) has provided direct experimental evidence that the position of infants' tongue and lips may impact how they perceive speech sounds. When 6-month-old English-learning infants were tested on a non-native speech contrast produced with movement of the tongue tip, they showed successful discrimination, replicating earlier results about infants' quasi-universal ability to discriminate consonant contrasts before about 10 months (Werker and Tees 1984). However, when the same infants had to accomplish the same task with a teething toy in their mouth that specifically inhibited tongue tip movements, infants failed. This effect was specific as teething toys with other shapes not impacting the position of the tongue tip, but that of the lips (lip spreading), did not prevent infants from making the discrimination. These results are remarkable in that the infants tested were preverbal, barely starting to babble, and had no experience with the phoneme contrast tested, yet showed the influence of the position of the articulators on their discrimination abilities, providing experimental evidence for the auditory-motor link at the earliest age in development (Bruderer et al. 2015).

8.3.4 *Tone Perception*

Perceptual attunement to the native language has been observed not only for phonemes, but also for lexical tones, the linguistic function of which is similar to that of phonemes in that they are minimal units of distinguishing meaning in tonal languages like Thai or Mandarin Chinese. The perception of lexical tone follows a similar attunement pattern to phonemes, with infants exposed to tone languages maintaining discrimination, and unexposed infants losing it over the second half of the first year of life, although some studies paint a more complex picture. The acoustic distance between the tested tone pairs seem to play a role, and some studies have also shown U-shaped developmental patterns whereby the ability to discriminate non-native tones returns after a drop even in non-exposed infants (Mattock and Burnham 2006; Liu and Kager 2012).

8.3.5 *Increasing Brain Specialization as a Correlate of Perceptual Attunement*

Perceptual attunement observed behaviorally is paralleled by increasing brain specialization at the neural level. Brain activation in response to language features found in the native language becomes more focal and more lateralized, with phoneme discrimination lateralizing to the left hemisphere and prosody-related discrimination lateralizing to the right (Minagawa-Kawai et al. 2011). As an example, 3-month-old Parisian infants respond bilaterally to Parisian French, their native dialect, and Quebecois French, a non-native regional dialect. Their brain responses to the two dialects are similar. By 5 months of age, however, Parisian infants show a differential response to the native dialect, which is left lateralized and more focal than 3-month-olds' responses (Cristia et al. 2014).

The processing of smaller linguistic units also gets lateralized. Lexical pitch accent contrasts, such as high-low vs. low-high, are readily discriminated both by 4-month-old and 10-month-old Japanese infants behaviorally, but brain imaging reveals important underlying differences in processing (Sato et al. 2010). The younger infants process the contrast bilaterally, with the activation patterns closely resembling their brain responses to pure tones, suggesting that processing is mostly based on the acoustic properties of the stimuli. The older infants, by contrast, show a left-lateralized discrimination response to the pitch accent contrast, the intensity of which is greater than that of the response to pure tones, indicating more specialized and more linguistically based processing. Similarly, Japanese infants have been found to discriminate the vowel duration contrast such as the short and long /a/ in Japanese (Minagawa-Kawai et al. 2007) at 6–7 months, not at 10–11 months, and then again from 13 to 14 months onward until adulthood. The initial discrimination response at 6–7 months is bilateral, whereas it becomes left lateralized from 13 to 14 months on, after reorganization.

These results clearly illustrate the development of the brain specialization for the native language. Processing and discrimination are initially acoustically based, and hence more bilateral. During reorganization, response patterns may vary or even weaken, and then re-emerge as more linguistic in nature, indexed by their more focal and lateralized location.

8.4 Learning Word Forms

As infants attune to their native sound repertoire, they also start acquiring their first words (Jusczyk and Aslin 1995; Tincoff and Jusczyk 1999; Bergelson and Swingley 2012). Speech is a continuous signal in which words are not systematically separated by pauses or other acoustic cues in a fully reliable manner. Thus, one challenge infants face when learning words is to segment out the possible word form candidates from the speech stream so that they can associate them with appropriate meanings. Here, we will only be addressing the word segmentation problem. How infants associate the extracted word forms with meaning goes beyond speech perception; the reader is, therefore, referred to existing overviews on this issue (Markman 1994; Golinkoff et al. 2000).

What cues do infants rely on to identify possible word forms? Several types of cues have been identified and statistical cues have received considerable attention. It has long been recognized that the statistical regularities of phoneme co-occurrences are also reliable indicators of word boundaries (Harris 1955; Brent and Cartwright 1996). Thus, the syllable /ti/ follows the syllable /pri/ with a greater probability than /bei/ follows /ti/, for example, as in the sequence *pretty baby*, because /pri/ and /ti/ frequently co-occur in the same word, while the adjective *pretty* might be followed by a large number of other words; thus, /ti/ and /bei/ do not necessarily co-occur. Saffran et al. (1996) and much subsequent work have shown that infants are able to pick up such regularities and use them to segment speech. Thus, infants expect a boundary between words when the probabilities between syllables are low.

Other segmentation cues have also been proposed in the literature. First, infants might rely on typical stress patterns, such as the strong-weak (trochaic) pattern commonly found in English content words (e.g., ***‘doc**tor*). This is plausible, because infants have been shown to develop sensitivity to the stress patterns typical of their native language between 6 and 9 months (Jusczyk and Aslin 1995; Morgan and Saffran 1995; Morgan 1996). Such a stress-based segmentation mechanism, called the Metrical Segmentation Strategy (Cutler and Carter 1987; Cutler 1994), has been shown to underlie 7.5-month-old English-learning infants' recognition of familiar words. In a series of studies, Jusczyk et al. (1999b) have shown that when familiarized with trochaic English words (e.g., ***'doc****tor*, ***'can****dle*), 7.5-month-olds prefer passages containing these words over passages that do not contain them. This preference is specific to the trochaic word form, because passages containing only the first strong syllables of the words (e.g., *dock*, *can*) did not give rise to a similar preference.

Moreover, by this age, English infants use language-specific stress cues to segment words from the ongoing speech stream. When presented with a continuous stream of syllables consisting of a consonant and a vowel, where every third syllable was stressed, 7- and 9-month-olds treated as familiar only those trisyllabic sequences that had initial stress (soft-weak-weak). Infants showed no recognition of trisyllabic sequences that were not trochaic (weak-soft-weak or weak-soft-soft; Curtin et al. 2001). The Metrical Segmentation Strategy also predicts that weak-strong, that is, iambic, words (e.g., *gui****'tar***) might initially be missegmented, which turns out to be the case (Jusczyk et al. 1999b).

A second possible language-specific cue to segmentation is phonotactics, that is, the regularities of how phonemes can be combined in a language. Knowing that the sequence /br/ is frequent in the initial positions of English words, while /nt/ typically appears at the end can help the learner posit word boundaries. Indeed, Saffran and Thiessen (2003) tested the acquisition of phonotactic constraints using segmentation as the experimental task. Using a different approach, Mattys et al. (1999) explored how 9-month-old English-learning infants' knowledge of English phonotactics helps them posit word boundaries. They familiarized infants with non-sense words consisting of a sequence of consonants (C) and vowels (V) in the following order CVCCVC. The CC cluster in the middle was either frequent word-internally in English, but infrequent across word boundaries (e.g., /nk/) or vice versa (e.g., /nt/). Infants segmented the non-sense words into two monosyllables when the CC cluster was infrequent word-internally and frequent across word boundaries. No segmentation was observed for the other type of CC clusters, indicating that 9-month-old infants can use their phonotactic knowledge to assist them in word segmentation (Mattys and Jusczyk 2001). Phonotactic biases, that is frequent, typical phonotactic patterns that appear in a language, can also aid segmentation. Thus, infants have been found to be perceptually sensitive to the Labial-Coronal bias by 10 months of age in languages, like French, in which this bias is present in the lexicon. The Labial-Coronal bias means that in the vocabulary of many languages, words with two consonants in them are such that the initial consonant is labial and the subsequent one is coronal, rather than the other way round. Studies suggest that infants show a preference for words that are Labial-Coronal over words with the opposite pattern (Nazzi et al. 2009). Similarly, infants learning languages with vowel harmony, but not those exposed to a non-harmonic language, are sensitive to this property of their native language by about 7–13 months of life (Altan et al. 2016; Gonzalez-Gomez et al. 2019). Vowel harmony is the tendency found in some languages for vowels within a word to be similar to one another in some feature. For instance, in Hungarian, vowels harmonize in frontness/backness (e.g., the word *ajtó* "door" only has back vowels, while the word *edény* "dish" only has front vowels). Sensitivity to such biases can help infants identify possible word forms in the input, and thus contribute to segmentation.

A third segmentation cue comes from the distributions of allophones, different realizations of the same phoneme in different positions within words. In English, aspirated stop consonants, consonants produced with a small puff of air, appear in the initial positions of stressed syllables (Church 1987), their unaspirated allophones appear elsewhere. Consequently, aspirated stops are good cues to word onsets. Because infants as young as 2 months are able to discriminate the different allophones of a phoneme (Hohne and Jusczyk 1994), it is not implausible to assume that they might use them as cues for segmentation. Indeed, Jusczyk et al. (1999a) have shown that at 9 months, infants are able to posit word boundaries (e.g., *night rates* vs. *nitrates*) based on allophonic and distributional cues, and at 10.5 months, allophonic cues alone are sufficient for successful segmentation.

In the speech input that infants receive, the above cues never occur in isolation. Therefore, it is important to understand how these cues interact during the actual process of language acquisition. Work by Mattys, Jusczyk, and colleagues (Mattys et al. 1999; Mattys and Jusczyk 2001) has shown that when stress and phonotactic cues are pitted against each other, that is, provide conflicting information about word boundaries, 9-month-old infants prefer to rely on stress cues. When stress and statistical information are contrasted, 6-month-olds follow the statistical information (Saffran and Thiessen 2003), while 8-month-olds rely more on stress (Johnson and Jusczyk 2001). This developmental trajectory might indicate a shift from universal to more language-specific strategies, reflecting infants' growing knowledge of the specifics of their native phonology.

By the end of the first year of life, infants thus develop powerful strategies to segment the continuous speech stream into words and start building a small vocabulary of candidate word forms. This development happens in parallel with the attunement to the native phoneme repertoire, and the two processes mutually influence each other. As a consequence, the native phoneme categories only become stable enough to support word learning in highly demanding contexts by about 18 months of age, but not yet at 14 months (although they are sufficiently reliable to allow word learning when context and task demands are low). Indeed, while 14-month-old infants can reliably learn to associate one non-sense word with a novel object and another non-sense word with another novel object when the non-sense words are phonologically distinct, such as "lif" and "neem," they have difficulty with minimal pairs. Minimal pairs are word that differ in a single phoneme, such as "bih" and "dih," and succeed in the latter task only by 18 months (Stager and Werker 1997). By about this age, they seem to encode even subsegmental detail in word forms (White and Morgan 2008).

Infants thus first show evidence of recognizing some word forms and reliably associate them with possible meanings between 6 and 9 months. Between this age and about 18 months, as their native phoneme repertoire stabilizes and they develop language-specific strategies for segmenting words, they start to build a sizeable lexicon as they become expert word learners during the second year of life.

8.5 Prosodic Perception

Infants' first linguistic experience largely consists of the rhythm and melody of the language(s) spoken by their mothers before birth (Gervain 2018). Throughout early language acquisition, prosody continues to play an important role in scaffolding language learning—this is known as prosodic bootstrapping.

Many lexical and grammatical properties of language are accompanied by characteristic prosodic patterns. The theory of prosodic bootstrapping (Morgan and Demuth 1996) holds that young learners can exploit the prosodic cues that are directly available in their input to learn about the perceptually unavailable, abstract lexical, and grammatical properties with which those cues are correlated. In English, for instance, bisyllabic nouns (N) and verbs (V) with the same segmental make-up are distinguished by lexical stress: nouns tend to have initial stress, verbs final stress, such as the noun *record* /ˈrekə(r)d/ vs. the verb *record* /riˈko(r)d / (Cutler and Carter 1987). Knowing this regularity, a learner is able to categorize novel words as nouns or verbs even if she does not know their meanings.

Experimental findings over the past two decades suggest that infants are indeed able to exploit such correlations to break into the lexicon and grammar of their native language(s), thus alleviating the learning problem they face when confronted with the acquisition of abstract linguistic properties (Gervain et al. 2021).

As reviewed in Sect. 8.2, many of newborns' speech perception abilities rely on prosody. These sensitivities constitute the basis of the subsequent bootstrapping role of prosody. One area in which this has been extensively documented is word learning (Sect. 8.4). Once infants learn the lexical stress pattern typical of their native language on the basis of the first few words they encounter, they can then use this knowledge to constrain and support further learning.

Another important mechanism of prosodic bootstrapping is prosodic grouping, also known as the Iambic-Trochaic Law (ITL) (Hayes 1995), which states that sound sequences contrasting in duration are naturally perceived iambically (e.g., as forming pairs in which the first sound is short, the second one is long), whereas sound sequences that contrast in pitch or intensity are perceived trochaically (e.g., as forming pairs in which the first sound is high/loud, the second one is low/soft). The position as well as the acoustic realization of phrase-level prosodic prominence co-varies with word order (Nespor et al. 2008; Gervain and Werker 2013). In languages in which phrases start with grammatical words called functors, (e.g., *in Rome*), such as English or Italian, prosodic prominence in phonological phrases, which falls on the content word, is phrase-final (i.e., iambic) and is realized as a durational contrast—that is, as the lengthening of the stressed vowel of the content word (e.g., *in* ***Ro****me*). By contrast, languages, such as Japanese, Turkish, or Basque, where grammatical words appear at the end of phrases, the prominence is initial (i.e., trochaic) and is realized as increased pitch or intensity (e.g., Japanese: ***To****kyo ni* "to Tokyo"). While other cues may accompany prominence in any language, pitch or intensity serves as the contrastive cue in languages with final grammatical functors, whereas duration plays this role in functor-initial languages. Infants as young as 8–9 months of age can align phrasal prosody with the underlying syntactic pattern within phrases, as they expect functors to be non-prominent and content words to be prominent (Bernard and Gervain 2012). Even more importantly, 7-month-old bilinguals exposed to a functor-initial and a functor-final language use the different prosodic realizations to select the relevant word order (Gervain and Werker 2013). Upon hearing a durational contrast (short-long), they select sequences with a functor-initial order, while, when presented with a pitch contrast (high-low), they prefer functor-final sequences. This is strong evidence that infants start using prosody to bootstrap syntax even before they have a sizeable lexicon, suggesting that they set abstract syntactic parameters rather than memorize or rote-learn lexical patterns or item-based expressions. In this regard, the role of the ITL is particularly relevant. As mentioned before, newborns already show familiarity with the predominant iambic or trochaic prosodic patterns of their native language from prenatal experience (Abboub et al. 2016). This knowledge may guide young infants from very early on in how they segment and parse the language input, and allow them to determine basic properties of their native grammar such as its word order. For instance, an infant expecting a functor-content word order on the basis of prosody will be able to directly assign the correct lexical category to the novel words she encounters in an input sentence. This is further aided by young infants' ability to distinguish functors and content words on the basis of their phonological differences (Shi et al. 1999). Thus, on the basis of auditory cues alone, infants may be able to already build a rudimentary representation of the basic word order of functors and content words, which then further correlates with other word order phenomena, such as the relative order of verbs and their objects, or main clauses and subordinate clauses, etc., (Dryer 1992), providing infants with a powerful strategy to break into the grammar of their native language.

Later, children can also use prosody to support the processing of syntactic structures (Christophe et al. 2008, 2016; Hawthorne and Gerken 2014). Infants perceive intonational phrase boundaries from 5 months of age (Hirsh-Pasek et al. 1987; Männel and Friederici 2009). To test the effect of phrasal prosody on syntactic analysis, sentences with syntactically ambiguous phrases were presented to toddlers such as *the baby flies*, which can be interpreted as a noun phrase as in *The baby flies hide in the shadows*, or as a noun and a verb as in *The baby flies her kite*. In these sentences, prosody disambiguates the two possibilities, as in one sentence there is a prosodic boundary before the ambiguous *word fly*, in the other case, the boundary follows *fly*. When listening to the critical phrase in such sentences (with the end of the sentence being masked by noise), toddlers as young as 20 months of age are able to exploit the prosodic information, and looked at the picture depicting the intended meaning (Carvalho et al. 2016).

Children thus use prosody from the very beginning of language development starting with their prenatal experience with speech to identify and break into the native language, relying on prosodic cues to extract words from the input, learn the basic word order of the native language, and subsequently to constrain syntactic parsing.

8.6 Chapter Summary

Infants start their journey into language as universal listeners, but by the end of the first year of life they become native language experts, as their perceptual systems and brains reorganize to better perceive those linguistic contrasts that they encounter in the native language, losing sensitivity to non-native sound patterns. Attunement to the native language starts prenatally, as infants first experience speech in the womb. Accordingly, newborns possess speech perception abilities, some of which already show the impact of prenatal experience, while many others are universal and broadly based, allowing infants to learn any of the world's languages. After several months of experience with their native languages, infants start to lose these universal abilities, becoming unable to discriminate most contrasts (phonemes, tones, etc.) that are not used in the native language(s), while improving and fine-tuning their native sound categories. This perceptual attunement is accompanied by an increasing hemispheric specialization for language at the neural level. In parallel with the perceptual reorganization, infants also start learning their first words and the basics of their native grammar. The acquisition of the different levels of language thus proceeds in parallel and interacts with one another in synergistic ways.

Acknowledgment This work was funded by the ERC Consolidator Grant "BabyRhythm" (no. 773202) to J.G.

Compliance with Ethics Requirements Judit Gervain declares that she has no conflict of interest.

References

Abboub N, Nazzi T, Gervain J (2016) Prosodic grouping at birth. Brain Lang 162:46–59

Abercrombie D (1967) Elements of general phonetics. Edinburgh University Press, Edinburgh

Altan A, Kaya U, Hohenberger A (2016) Sensitivity of Turkish infants to vowel harmony in stem-suffix sequences: preference shift from familiarity to novelty. BUCLD 40 Online Proceedings Supplements

Bergelson E, Swingley D (2012) At 6–9 months, human infants know the meanings of many common nouns. Proc Natl Acad Sci USA 109:3253–3258

Bernard, C., & Gervain, J. (2012). Prosodic cues to word order: What level of representation? *Frontiers in Language Sciences, 3*:, 451. https://doi.org/10.3389/fpsyg.2012.00451

Best CT, McRoberts GW, Sithole NM (1988) Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. J Exp Psychol Hum Percept Perform 14:345–360

Birulés J, Bosch L, Pons F, Lewkowicz DJ (2020) Highly proficient L2 speakers still need to attend to a talker's mouth when processing L2 speech. Lang Cogn Neurosci 35(10):1314–1325

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

Brent MR, Cartwright TA (1996) Distributional regularity and phonotactic constraints are useful for segmentation. Cognition 61:93–125

Bruderer AG, Danielson DK, Kandhadai P, Werker JF (2015) Sensorimotor influences on speech perception in infancy. Proc Natl Acad Sci USA 112:13531–13536

Byers-Heinlein K, Fennell CT (2014) Perceptual narrowing in the context of increased variation: insights from bilingual infants. Dev Psychobiol 56:274–291

Byers-Heinlein K, Burns TC, Werker JF (2010) The roots of bilingualism in newborns. Psychol Sci 21:343–348

Cabrera L, Gervain J (2020) Speech perception at birth: the brain encodes fast and slow temporal information. Sci Adv 6:eaba7830

Carvalho A, Dautriche I, Christophe A (2016) Preschoolers use phrasal prosody online to constrain syntactic analysis. Dev Sci 19:235–250

Casey BJ, Giedd JN, Thomas KM (2000) Structural and functional brain development and its relation to cognitive development. Biol Psychol 54:241–257

Chomsky N, Belletti A, Rizzi L (2002) On nature and language. Cambridge University Press, Cambridge

Christophe A, Dupoux E, Bertoncini J, Mehler J (1994) Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. J Acoust Soc Am 95:1570–1580

Christophe A, Millotte S, Bernal S, Lidz J (2008) Bootstrapping lexical and syntactic acquisition. Lang Speech 51:61–75

Christophe A, de Carvalho A, Dautriche I et al (2016) Synergies in early language acquisition. Int J Psychol 51:811

Church KW (1987) Phonological parsing and lexical retrieval. Cognition 25:53–69

Cristia A, Minagawa-Kawai Y, Egorova N et al (2014) Neural correlates of infant accent discrimination: an fNIRS study. Dev Sci 17:628–635

Curtin S, Mintz TH, Byrd D (2001) Coarticulatory cues enhance infants' recognition of syllable sequences in speech. In: Proceedings of the 25th annual Boston University conference on language development. Cascadilla Press, Somerville, pp 190–201

Curtiss S, Fromkin VA, Krashen S et al (1974) The linguistic development of Genie. Language 50:528–554

Cutler A (1994) Segmentation problems, rhythmic solutions. Lingua 92:81–104

Cutler A, Carter DM (1987) The predominance of strong initial syllables in the English vocabulary. Comput Speech Lang 2:133–142

Danielson DK, Bruderer AG, Kandhadai P et al (2017) The organization and reorganization of audiovisual speech perception in the first year of life. Cogn Dev 42:37–48

de la Cruz-Pavía I, Gervain J, Vatikiotis-Bateson E, Werker JF (2019) Finding phrases: on the role of co-verbal facial information in learning word order in infancy. PLoS One 14:e0224786

de la Cruz-Pavía I, Gervain J, Vatikiotis-Bateson E, Werker JF (2020a) Coverbal speech gestures signal phrase boundaries: a production study of Japanese and English infant-and adult-directed speech. Lang Acquis 27:160–186

de la Cruz-Pavía I, Werker JF, Vatikiotis-Bateson E, Gervain J (2020b) Finding phrases: the interplay of word frequency, phrasal prosody and co-speech visual information in chunking speech by monolingual and bilingual adults. Lang Speech 63:264–291

DeCasper AJ, Fifer WP (1980) Of human bonding: newborns prefer their mothers' voices. Science 208:1174–1176

DeCasper AJ, Lecanuet J-P, Busnel MC, Granier-Deferre C (1994) Fetal reactions to recurrent maternal speech. Infant Behav Dev 17:159–164

Dehaene-Lambertz G, Baillet S (1998) A phonological representation in the infant brain. Neuroreport 9:1885–1888

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

Dellwo V (2006) Rhythm and speech rate: a variation coefficient for ΔC. In: Language and language-processing. Peter Lang, Frankfurt/Main, pp 231–241

Drullman R (1995) Temporal envelope and fine structure cues for speech intelligibility. J Acoust Soc Am 97:585–592

Dryer MS (1992) The Greenbergian word order correlations. Language 68:81–138

Eggermont JJ, Moore JK (2012) Morphological and functional development of the auditory nervous system. In: Human auditory development. Springer, New York, pp 61–105

Eimas PD, Siqueland ER, Jusczyk PW, Vigorito J (1971) Speech perception in infants. Science 171:303–306

Elman J, Bates E, Johnson M et al (1997) Rethinking innateness: a connectionist perspective on development. The MIT Press, Boston

Gerhardt KJ, Abrams RM, Oliver CC (1990) Sound environment of the fetal sheep. Am J Obstet Gynecol 162:282–287

Gervain J (2015) Plasticity in early language acquisition: the effects of prenatal and early childhood experience. Curr Opin Neurobiol 35:13–20

Gervain J (2018) The role of prenatal experience in language development. Curr Opin Behav Sci 21:62–67

Gervain J, Werker JF (2013) Prosody cues word order in 7-month-old bilingual infants. Nat Commun 4:1490

Gervain J, Macagno F, Cogoi S et al (2008) The neonate brain detects speech structure. Proc Natl Acad Sci USA 105:14222–14227

Gervain J, Berent I, Werker JF (2012) Binding at birth: the newborn brain detects identity relations and sequential position in speech. J Cogn Neurosci 24:564–574

Gervain J, Vines BW, Chen LM et al (2013) Valproate reopens critical-period learning of absolute pitch. Front Syst Neurosci 7:102

Gervain J, Christophe A, Mazuka R (2021) Prosodic bootstrapping. In: The handbook of prosody. Oxford University Press, Oxford

Golinkoff RM, Hirsh-Pasek K, Bloom L et al (2000) Becoming a word learner: a debate on lexical acquisition. Oxford University Press, New York

Gonzalez-Gomez N, Schmandt S, Fazekas J et al (2019) Infants' sensitivity to nonadjacent vowel dependencies: the case of vowel harmony in Hungarian. J Exp Child Psychol 178:170–183

Grabe E, Low EL (2002) Durational variability in speech and the rhythm class hypothesis. In: Papers in laboratory phonology. Mouton de Gruyter, Berlin, pp 515–546

Guellaï B, Langus A, Nespor M (2014) Prosody in the hands of the speaker. Front Psychol 5:700

Gustafson GE, Sanborn SM, Lin H-C, Green JA (2017) Newborns' cries are unique to individuals (but not to language environment). Infancy 22:736–747

Haartsen R, Jones EJ, Johnson MH (2016) Human brain development over the early years. Curr Opin Behav Sci 10:149–154

Harris Z (1955) From phoneme to morpheme. Language 31:190–222

Hawthorne K, Gerken L (2014) From pauses to clauses: prosody facilitates learning of syntactic constituency. Cognition 133:420–428

Hayes B (1995) Metrical stress theory: principles and case studies. University of Chicago Press, Chicago

Hensch TK (2005) Critical period plasticity in local cortical circuits. Nat Rev Neurosci 6:877–888

Hickok G, Buchsbaum B, Humphries C, Muftuler T (2003) Auditory–motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J Cogn Neurosci 15:673–682

Hirsh-Pasek K, KemlerNelson DG, Jusczyk PW et al (1987) Clauses are perceptual units for young infants. Cognition 26:269–286

Hohne EA, Jusczyk PW (1994) Two-month-old infants' sensitivity to allophonic differences. Percept Psychophys 56:613–623

Hollich G, Newman RS, Jusczyk PW (2005) Infants' use of synchronized visual information to separate streams of speech. Child Dev 76:598–613

Hunnius S, Geuze RH (2004) Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: a longitudinal study. Infancy 6:231–255

Johnson EK, Jusczyk PW (2001) Word segmentation by 8-month-olds: when speech cues count more than statistics. J Mem Lang 44:548–567

Johnson JS, Newport EL (1989) Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. Cogn Psychol 21:60–99

Jusczyk PW, Aslin RN (1995) Infants' detection of the sound patterns of words in fluent speech. Cogn Psychol 29:1–23

Jusczyk PW, Hohne EA, Bauman A (1999a) Infants' sensitivity to allophonic cues to word segmentation. Percept Psychophys 61:1465–1476

Jusczyk PW, Houston DM, Newsome MR (1999b) The beginnings of word segmentation in English-learning infants. Cogn Psychol 39:159–207

Kisilevsky BS, Hains SMJ, Brown CA et al (2009) Fetal sensitivity to properties of maternal speech and language. Infant Behav Dev 32:59–71

Kubicek C, De Boisferon AH, Dupierrix E et al (2013) Face-scanning behavior to silently-talking faces in 12-month-old infants: the impact of pre-exposed auditory speech. Int J Behav Dev 37:106–110

Kubicek C, De Boisferon AH, Dupierrix E et al (2014a) Cross-modal matching of audio-visual German and French fluent speech in infancy. PLoS One 9:e89275

Kubicek C, Gervain J, Hillairet de Boisferon A et al (2014b) The influence of infant-directed speech on 12-month-olds' intersensory perception of fluent speech. Infant Behav Dev 37:644–651

Kuhl PK (1981) Discrimination of speech by nonhuman animals: basic auditory sensitivities conducive to the perception of speech-sound categories. J Acoust Soc Am 70:340–349

Kuhl PK (1986) Theoretical contributions of tests on animals to the special-mechanisms debate in speech. Exp Biol 45:233–265

Kuhl PK (2004) Early language acquisition: cracking the speech code. Nat Rev Neurosci 5:831–843

Kuhl PK, Meltzoff AN (1982) The bimodal perception of speech in infancy. Science 218:1138–1141

Kuhl PK, Williams KA, Lacerda F et al (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255:606–608

Kuhl PK, Stevens E, Hayashi A et al (2006) Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Dev Sci 9:F13–f21

Kujala A, Huotilainen M, Hotakainen M et al (2004) Speech-sound discrimination in neonates as measured with MEG. Neuroreport 15:2089–2092

Langus A, Mehler J, Nespor M (2016) Rhythm in language acquisition. Neurosci Biobehav Rev 81:158–166

Lecanuet JP, Granier-Deferre C (1993) Speech stimuli in the fetal environment. In: Developmental neurocognition: speech and face processing in the first year of life. Kluwer Academic Press, Dordrecht, pp 237–248

Lenneberg EH (1967) The biological foundations of language. Wiley Press, New York

Lewkowicz DJ, Hansen-Tift AM (2012) Infants deploy selective attention to the mouth of a talking face when learning speech. Proc Natl Acad Sci USA 109:1431–1436

Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. Cognition 21:1–36

Liu L, Kager R (2012) Non-native tone perception from infant to adult: how consistent and flexible is it? In: Proceedings of speech prosody

MacKain KS, Studdert-Kennedy M, Spieker S, Stern DN (1983) Infant intermodal speech perception is a left-hemisphere function. Science 219:1347–1349

Mahmoudzadeh M, Dehaene-Lambertz G, Fournier M et al (2013) Syllabic discrimination in premature human infants prior to complete formation of cortical layers. Proc Natl Acad Sci USA 110:4846–4851

Majorano M, Vihman MM, DePaolis RA (2014) The relationship between infants' production experience and their processing of speech. Lang Learn Dev 10:179–204

Mampe B, Friederici AD, Christophe A, Wermke K (2009) Newborns' cry melody is shaped by their native language. Curr Biol 19:1994–1997

Manfredi C, Viellevoye R, Orlandi S et al (2019) Automated analysis of newborn cry: relationships between melodic shapes and native language. Biomed Signal Process Control 53:101561

Männel C, Friederici AD (2009) Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults. J Cogn Neurosci 21:1988–2006

Markman EM (1994) Constraints on word meaning in early language acquisition. Lingua 92:199–227

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

Mattys SL, Jusczyk PW (2001) Phonotactic cues for segmentation of fluent speech by infants. Cognition 78:91–121

Mattys SL, Jusczyk PW, Luce PA, Morgan JL (1999) Phonotactic and prosodic effects on word segmentation in infants. Cogn Psychol 38:465–494

Maurer D, Werker JF (2014) Perceptual narrowing during infancy: a comparison of language and faces. Dev Psychobiol 56:154–178

May L, Byers-Heinlein K, Gervain J, Werker JF (2011) Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech? Front Lang Sci 2:222

May L, Gervain J, Carreiras M, Werker JF (2017) The specificity of the neural response to speech at birth. Dev Sci 21:e12564

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748

McMurray B, Aslin RN (2005) Infants are sensitive to within-category variation in speech perception. Cognition 95:B15–B26

Mehler J, Jusczyk PW, Lambertz G et al (1988) A precursor of language acquisition in young infants. Cognition 29:143–178

Mehler J, Sebastian-Galles N, Nespor M (2004) Biological foundations of language: language acquisition, cues for parameter setting and the bilingual infant. In: The new cognitive neuroscience. MIT Press, Cambridge, MA, pp 825–836

Minagawa-Kawai Y, Mori K, Naoi N, Kojima S (2007) Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. J Neurosci 27:315–321

Minagawa-Kawai Y, Cristia A, Dupoux E (2011) Cerebral lateralization and early speech acquisition: a developmental scenario. Dev Cogn Neurosci 1:217–232

Mitchel AD, Weiss DJ (2014) Visual speech segmentation: using facial cues to locate word boundaries in continuous speech. Lang Cogn Process 29:771–780

Molnar M, Gervain J, Carreiras M (2013) Within-rhythm class native language discrimination abilities of Basque-Spanish monolingual and bilingual infants at 3.5 months of age. Infancy 19:326–337

Moon C, Cooper RP, Fifer WP (1993) Two-day-olds prefer their native language. Infant Behav Dev 16:495–500

Moon C, Lagercrantz H, Kuhl PK (2013) Language experienced in utero affects vowel perception after birth: a two-country study. Acta Paediatr 102:156–160

Moore DR (2002) Auditory development and the role of experience. Br Med Bull 63:171–181

Morgan JL (1996) A rythmic bias in preverbal speech segmentation. J Mem Lang 35:666–688

Morgan JL, Demuth K (1996) Signal to syntax: bootstrapping from speech to grammar in early acquisition. Lawrence Erlbaum Associates, Inc, Hillsdale

Morgan JL, Saffran JR (1995) Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. Child Dev 66:911–936

Narayan CR, Werker JF, Beddor PS (2010) The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. Dev Sci 13:407–420

Nazzi T, Bertoncini J, Mehler J (1998) Language discrimination by newborns: toward an understanding of the role of rhythm. J Exp Psychol Hum Percept Perform 24:756–766

Nazzi T, Bertoncini J, Bijeljac-Babic R (2009) A perceptual equivalent of the labial-coronal effect in the first year of life. J Acoust Soc Am 126:1440

Nespor M (1990) On the rhythm parameter in phonology. In: Logical issues in language acquisition. Foris, Dordrecht, pp 157–175

Nespor M, Shukla M, van de Vijver R et al (2008) Different phrasal prominence realization in VO and OV languages. Lingue E Linguaggio 7:1–28

Partanen E, Kujala T, Näätänen R et al (2013) Learning-induced neural plasticity of speech processing before birth. Proc Natl Acad Sci USA 110:15145–15150

Pascalis O, deHaan M, Nelson CA (2002) Is face processing species-specific during the first year of life? Science 296:1321–1323

Patterson ML, Werker JF (1999) Matching phonetic information in lips and voice is robust in 4.5-month-old infants. Infant Behav Dev 22:237–247

Patterson ML, Werker JF (2002) Infants' ability to match dynamic phonetic and gender information in the face and voice. J Exp Child Psychol 81:93–115

Peña M, Maki A, Kovacic D et al (2003) Sounds and silence: an optical topography study of language recognition at birth. Proc Natl Acad Sci USA 100:11702–11705

Pierce LJ, Klein D, Chen J-K et al (2014) Mapping the unconscious maintenance of a lost first language. Proc Natl Acad Sci USA 111:17314–17319

Pons F, Lewkowicz DJ, Soto-Faraco S, Sebastián-Gallés N (2009) Narrowing of intersensory speech perception in infancy. Proc Natl Acad Sci USA 106:10598

Pons F, Bosch L, Lewkowicz DJ (2015) Bilingualism modulates infants' selective attention to the mouth of a talking face. Psychol Sci 26:490–498

Pons F, Sanz-Torrent M, Ferinu L et al (2018) Children with SLI can exhibit reduced attention to a talker's mouth. Lang Learn 68:180–192

Ramus F, Nespor M, Mehler J (1999) Correlates of linguistic rhythm in the speech signal. Cognition 73:265–292

Ramus F, Hauser MD, Miller C et al (2000) Language discrimination by human newborns and by cotton-top tamarin monkeys. Science 288:349–351

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.

Saffran JR, Thiessen ED (2003) Pattern induction by infant language learners. Dev Psychol 39:484–494

Sansavini A, Bertoncini J, Giovanelli G (1997) Newborns discriminate the rhythm of multisyllabic stressed words. Dev Psychol 33:3–11

Sato Y, Sogabe Y, Mazuka R (2010) Development of hemispheric specialization for lexical pitch-accent in Japanese infants. J Cogn Neurosci 22:2503–2513

Sato H, Hirabayashi Y, Tsubokura H et al (2012) Cerebral hemodynamics in newborn infants exposed to speech sounds: a whole-head optical topography study. Hum Brain Mapp 33:2092–2103

Sebastián-Gallés N, Albareda-Castellot B, Weikum WM, Werker JF (2012) A bilingual advantage in visual language discrimination in infancy. Psychol Sci 23:994–999

Shannon RV, Zeng F-G, Kamath V et al (1995) Speech recognition with primarily temporal cues. Science 270:303–304

Shi R, Werker JF, Morgan JL (1999) Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. Cognition 72:B11–B21

Soto-Faraco S, Navarra J, Weikum WM et al (2007) Discriminating languages by speech-reading. Percept Psychophys 69:218–231

Stager CL, Werker JF (1997) Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature 388:381–382

Swingley D (2021) Infants' learning of speech sounds and word forms. In: Oxford handbook of the mental lexicon. Oxford University Press, Oxford

Tierney AL, Nelson CA (2009) Brain development and the role of experience in the early years. Zero Three 30:9

Tincoff R, Jusczyk PW (1999) Some beginnings of word comprehension in 6-month-olds. Psychol Sci 10:172–175

Tomasello M (2000) Do young children have adult syntactic competence? Cognition 74:209–253

Ventureyra VA, Pallier C, Yoo H-Y (2004) The loss of first language phonetic perception in adopted Koreans. J Neurolinguistics 17:79–91

Vilain A, Dole M, Lœvenbruck H et al (2019) The role of production abilities in the perception of consonant category in infants. Dev Sci 22:e12830

Viola Macchi C, Turati C, Simion F (2004) Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? Psychol Sci 15:379–383

Vouloumanos A, Werker JF (2004) Tuned to the signal: the privileged status of speech for young infants. Dev Sci 7:270

Vouloumanos A, Hauser MD, Werker JF, Martin A (2010) The tuning of human neonates' preference for speech. Child Dev 81:517–527

Wagner P, Malisz Z, Kopp S (2014) Gesture and speech in interaction: an overview. Elsevier

Weaver IC, Cervoni N, Champagne FA et al (2004) Epigenetic programming by maternal behavior. Nat Neurosci 7:847–854

Weikum WM, Vouloumanos A, Navarra J et al (2007) Visual language discrimination in infancy. Science 316:1159

Weikum WM, Oberlander TF, Hensch TK, Werker JF (2012) Prenatal exposure to antidepressants and depressed maternal mood alter trajectory of infant speech perception. Proc Natl Acad Sci USA 109:17221–17227

Weikum WM, Vouloumanos A, Navarra J et al (2013) Age-related sensitive periods influence visual language discrimination in adults. Front Syst Neurosci 7:86

Werker JF (2018) Perceptual beginnings to language acquisition. Appl Psycholinguist 39:703–728

Werker JF, Curtin S (2005) PRIMIR: a developmental model of speech processing. Lang Learn Dev 1:197–234

Werker JF, Hensch TK (2015) Critical periods in speech perception: new directions. Annu Rev Psychol 66:173–196

Werker JF, Tees RC (1984) Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behav Dev 7:49–63

Werker JF, Yeung HH (2005) Infant speech perception bootstraps word learning. Trends Cogn Sci 9:519–527

White KS, Morgan JL (2008) Sub-segmental detail in early lexical representations. J Mem Lang 59:114–132

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