Language and Linguistics Compass 2/6 (2008): 1149–1170, 10.1111/j.1749-818x.2008.00089.x

How Infant Speech Perception Contributes to Language Acquisition

Judit Gervain* and Janet F. Werker
University of British Columbia

Abstract

Perceiving the acoustic signal as a sequence of meaningful linguistic representations is a challenging task, which infants seem to accomplish effortlessly, despite the fact that they do not have a fully developed knowledge of language. The present article takes an integrative approach to infant speech perception, emphasizing how young learners' perception of speech helps them acquire abstract structural properties of language. We introduce what is known about infants' perception of language at birth. Then, we will discuss how perception develops during the first 2 years of life and describe some general perceptual mechanisms whose importance for speech perception and language acquisition has recently been established. To conclude, we discuss the implications of these empirical findings for language acquisition.

1. Introduction

As part of our everyday life, we routinely interact with children and adults, women and men, as well as speakers using a dialect different from our own. We might talk to them face-to-face or on the phone, in a quiet room or on a busy street. Although the speech signal we receive in these situations can be physically very different (e.g., men have a lower-pitched voice than women or children), we usually have little difficulty understanding what our interlocutors say. Yet, this is no easy task, because the mapping from the acoustic signal to the sounds or words of a language is not straightforward. For instance, the physical signals corresponding to the same phoneme differ depending on the speaker (age, gender, dialect, individual variation, etc.), the context of the sound (position within word, surrounding phonemes, etc.), and a number of other factors (medium, noise level, etc.). The phoneme /d/, for instance, is pronounced differently when it is followed by different vowels, as Figure 1 illustrates. In addition, speech is continuous, while we represent language as a sequence of discrete phonemes, organized into discrete words. In our everyday lives, we often encounter manifestations of these problems when we interact with automated speech recognition systems, such as computers or automatic

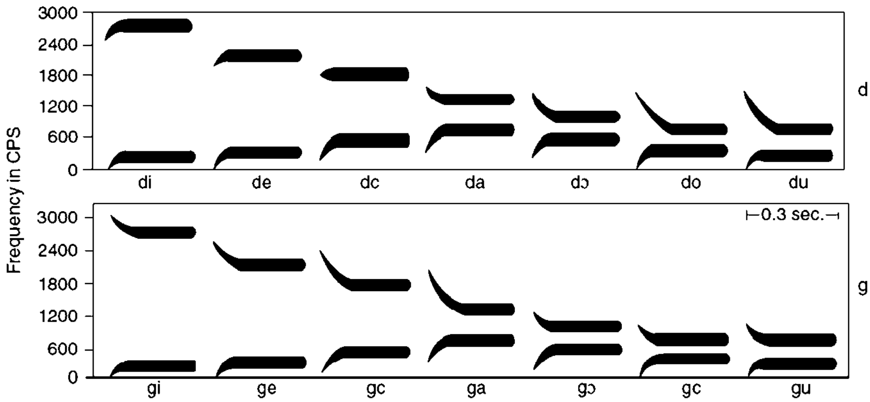

Fig. 1. The different acoustic realizations of the phonemes /d/ and /g/ when followed by different vowels. Reprinted from Tomasello and Bates (2001), p. 15, with permission; figure originally published in Liberman et al. (1967).

call centers. Machines usually fail to recognize a wide variety of inputs from a wide variety of speakers; either the number of speakers or the number of possible inputs has to be limited.

Perceiving the acoustic signal as a sequence of meaningful linguistic representations is thus a challenging task, which adults master with great efficiency, using mechanisms that we only partially understand. However, from what we do know, it is clear that their fully developed knowledge of language plays a considerable role in the process (Samuel 1996). It is, therefore, an even greater challenge to understand how infants perceive speech in the absence of a mature language faculty. In the past, speech perception and language acquisition were believed to be largely independent processes. Indeed, the systematic investigation of infant speech perception began only in the 1970s, much later than research into language acquisition. One of the main reasons for this separation was that the structure of language was believed to be an autonomous module (Chomsky 1957). Indeed, the focus of language acquisition research was mostly on understanding the innate language acquisition device (Chomsky 1959). The nature of the input and the role it plays in the acquisition process were largely ignored.

In the last two decades, however, it has been increasingly recognized that the input places a logical boundary condition on the hypothesized contents of the language acquisition device (Morgan et al. 1987). Information that is actually there in the input (and that infants are able to learn) need not be posited as innate. Even more importantly, the acoustic and phonological properties of the input have been found to correlate with some of the structural, that is, morphosyntactic, properties of language.

This has led to the formulation of the phonological (or prosodic) bootstrapping theory of language acquisition (Morgan and Demuth 1996). For instance, English nouns (N) have a strong-weak, that is, trochaic, stress pattern, while verbs (V) are weak-strong or iambic, for example, *'**re**cord* N vs. *re'**cord*** V (Cutler and Carter 1987; Davis and Kelly 1997). If a learner 'knows' (not necessarily explicitly) this correlation, then she can easily categorize a new word as N or V solely on the basis of its phonological form, without understanding its meaning or the sentential contexts in which it appears. The categorization, in turn, allows the learner to productively use the new word in morphological forms and syntactic contexts typically associated with other members of the same category (form the plural, etc.) Thus, the main strength of phonological bootstrapping theories is that they attempt to bridge the gap between sound, the perceptually available facet of language, and structure, the underlying and abstract organizational principles.

In the light of the above, the present article takes an integrative approach to infant speech perception, emphasizing how young learners' perception of speech helps them acquire abstract structural properties of language. We will first introduce what is known about infants' perception of language at the initial state, that is, before and immediately after birth. Then, we will discuss how perception develops during the first 2 years of life, reviewing perceptual mechanisms and their contribution to the acquisition of structure at different levels, from phonemes through word forms to larger prosodic units. We will then describe some general perceptual mechanisms whose importance for speech perception and language acquisition has recently been established. To conclude, we discuss the implications of the previously described empirical findings for language acquisition, returning to the phonological bootstrapping theory, and we will outline some open questions that we believe will shape the agenda of infant speech perception research in the years to come.

2. An Initial Preference for Speech: Newborn Speech Perception

Newborn babies have a rich repertoire of speech perception abilities. They have been shown to prefer human speech to equally complex speech analogues (see Figure 2; Vouloumanos and Werker 2007; the same was shown for 2-month-olds by Vouloumanos and Werker 2004), and were found to have larger left hemispheric brain activity when listening to speech played forward as opposed to speech played backward (see Figure 3; Peña et al. 2003; the same was shown for 3-month-olds by Dehaene-Lambertz et al. 2002).

Newborns have also been found to discriminate between languages. They distinguish and prefer the language spoken by their mothers during pregnancy over other languages (Mehler et al. 1988; Moon et al. 1993). This suggests that learning about spoken language begins during the fetal

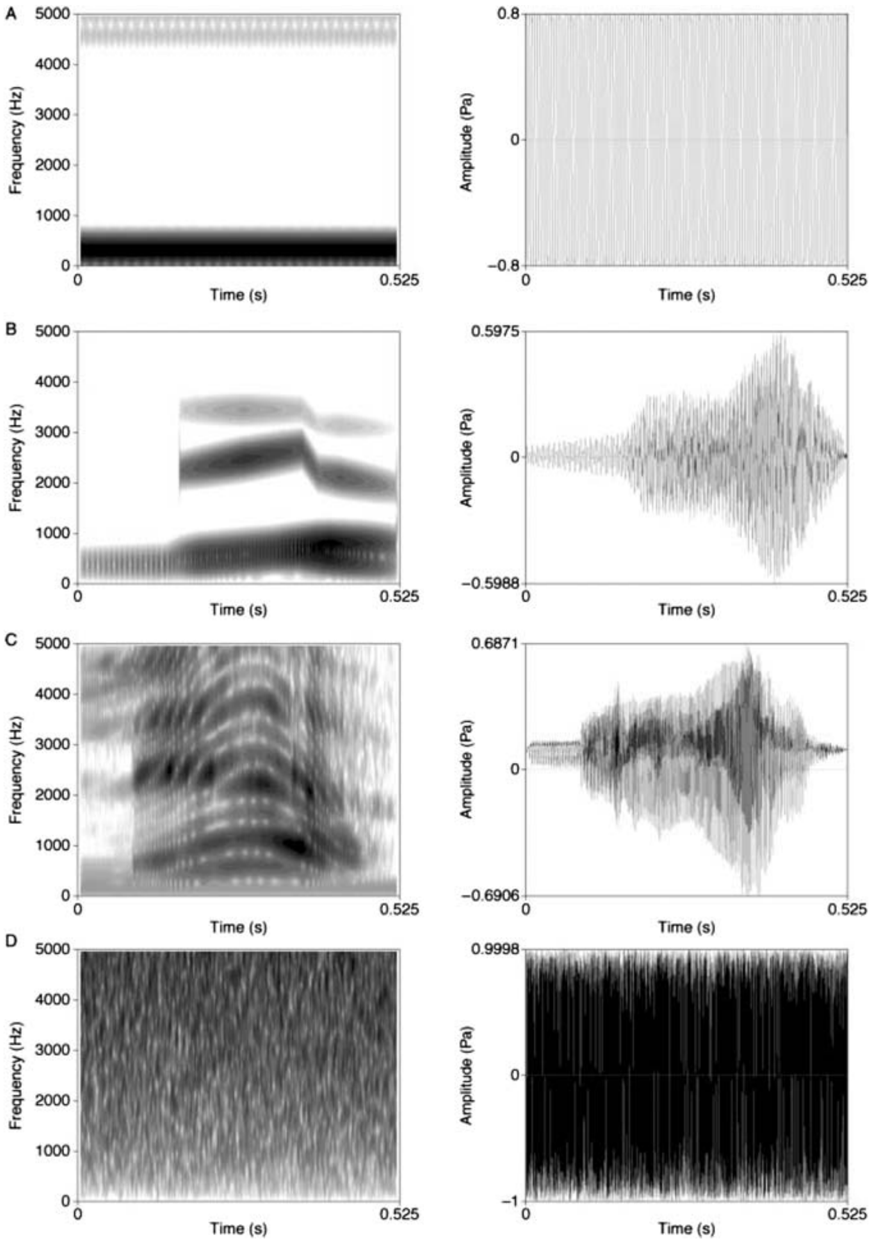

Fig. 2. The speech and speech-analog stimuli used in Vouloumanos and Werker (2007). Plots in the left column are spectrograms, those in the right are waveforms of (A) pure tones, (B) non-speech analogs, (C) speech, and (D) white noise. Reprinted from Vouloumanos and Werker (2007) with permission.

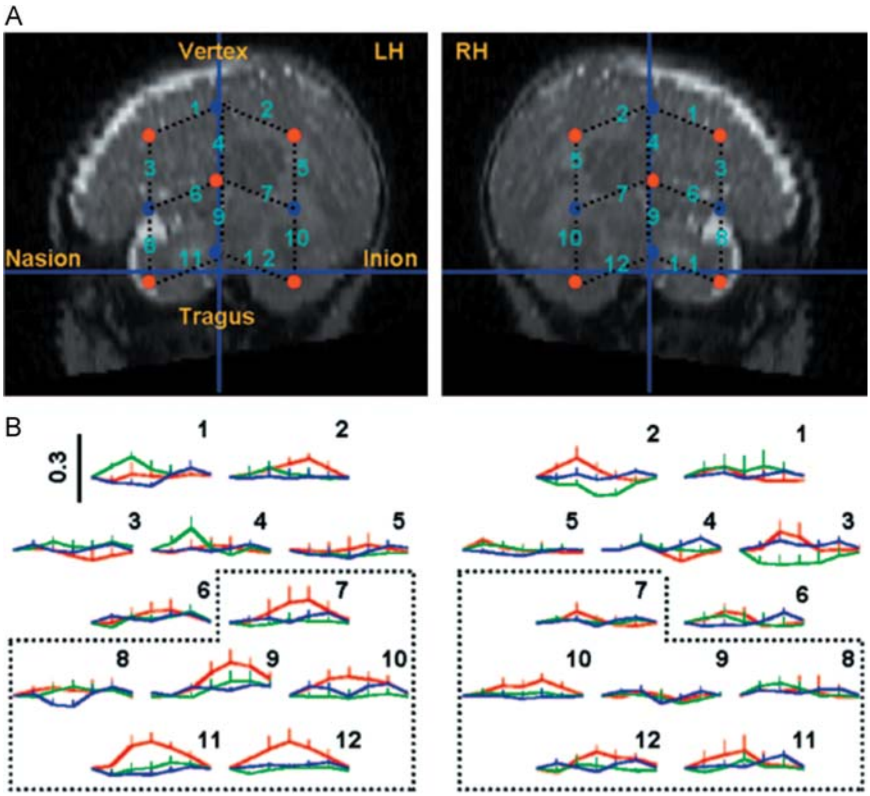

Fig. 3. Brain responses showing newborns' preference for forward speech (red curves) over backward speech (green curves) and silence (blue curves) in the left hemisphere. Panel (A) indicates the placement of the optical probes overlaid on a model newborn brain. Panel (B) shows the measured hemodynamic responses at each measurement point. Reprinted from Peña et al. (2003) with permission.

period. This is not implausible, because the auditory system is functional by about the 24th week of gestation (Mehler and Dupoux 1994; Thorburn and Harding 1994; Moore 2002), and although the womb filters out most of the fine details of speech, some of the more general properties, such as intonational contours or rhythmicity, are preserved. Interestingly, newborns can also discriminate two languages that they never heard before, if those are rhythmically different from each other, such as English and Japanese (Mehler et al. 1988; Nazzi et al. 1998). Infants' perception of linguistic rhythm and its role in language acquisition will be discussed in greater detail in Section 3 below.

Newborn infants also have surprising abilities to process acoustic information pertaining to word forms. They can detect the acoustic cues that signal word boundaries (Christophe et al. 1994), discriminate words with different patterns of lexical stress (Sansavini et al. 1997), and distinguish between function words (articles, pronouns, prepositions, determiners, etc.) and content words (nouns, verbs, adjectives, adverbs, etc.) on the basis of their different acoustic characteristics (Shi et al. 1999).

These findings suggest that newborns have an initial perceptual bias for speech, underlying all subsequent language acquisition processes. The discrimination abilities facilitate acquisition, allowing infants to tune in to the relevant environmental input. Moreover, they enable infants growing up in multilingual environments to separate and keep track of their different languages.

3. Rhythmicity

How infants growing up with several languages discriminate between them has been an exciting puzzle in language acquisition research for a long time. Linguists have long established (James 1940; Pike 1945; Abercrombie 1967; Ladefoged 1993) that the languages of the world may be grouped into three categories on the basis of their rhythmicity. In what linguists traditionally categorized as syllable-timed languages, such as Spanish or Italian, the organizing time unit was believed to be the syllable. In the group called stress-timed languages, like English, Dutch, or Arabic, the unit of isochrony was assumed to be the interstress interval. And in mora-timed languages, like Japanese and Tamil, the unit was believed to be the mora. These differences are easy to perceive. We rapidly realize if two rhythmically different languages are spoken in our environment, even if we never heard them before. English and Dutch sound like Morse code, with consonants tending to cluster together (e.g., ***spr****i****ngs***), while Spanish or French have what is sometimes referred to as a machine-gun rhythm, with a more equally paced alternation of consonants and vowels (e.g., Spanish ***b****u****f****a****nd****a* 'scarf'). If infants also perceive linguistic rhythm in a similar way, then this might help explain how multilingual infants differentiate their languages (Grosjean 1982).

Rhythm-based language discrimination was first explored by Mehler et al. (1988), who have shown that newborns are able to discriminate their mother tongue from another language, even if the stimuli are low-pass filtered, suppressing segmental, that is, phonemic, information. This was the first indication that language discrimination might depend on suprasegmental, prosodic cues. Nazzi et al. (1998) have extended these results by showing that suprasegmental cues are sufficient for discrimination; familiarity with the languages is not necessary. The authors have found that French newborns are able to discriminate between English and Japanese (low-pass filtered) stimuli. Moreover, they have also shown that newborns depend on linguistic rhythm when discriminating between the languages, as rhythmically similar languages are not differentiated.

While Mehler et al. (1988) and Nazzi et al. (1998) established that newborns are able to discriminate unknown languages on the basis of

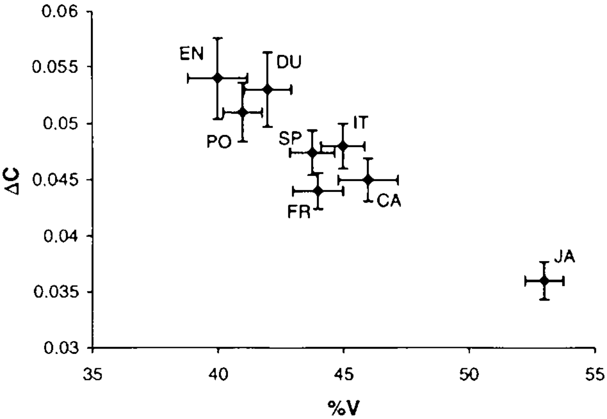

Fig. 4. The rhythmical categorization of languages in the %V and ΔC space. %V is the proportion of vocalic space in the speech stream, while ΔC is the variability of consonantal space. Reprinted from Ramus et al. (1999) with permission.

rhythm, it was not clear what acoustic features of the signal corresponded to the percept of rhythm. Low-pass filtering suggested that the relevant properties are suprasegmental, but did not allow their precise identification (Ramus and Mehler 1999; Ramus et al. 1999). The isochrony principle, proposed earlier by linguists, was not supported by empirical evidence (no isochrony was found; Dauer 1983), and subsequent findings by Nespor (1990) suggested that rhythmicity is better understood as a continuum, not as discrete categories, because languages were found with both stress-timed and syllable-timed properties. The solution came from Ramus et al. (1999), who operationalized rhythmical classes as clusters in three different two-dimensional spaces. The first was defined by %V, the relative length of vocalic space in the speech signal, and ΔC, the variability in the length of consonant clusters (see Figure 4). The second was defined by %V and ΔV, the variability of vocalic spaces. The third was defined by ΔC and ΔV. The authors measured these three properties in eight languages, English, Dutch, Polish, Catalan, Italian, Spanish, French, and Japanese, and plotted them in the two-dimensional spaces. In all three plots, the languages clustered into three classes, reproducing the original rhythmic categorization: English, Dutch, and Polish in one group (cf. stress-timed languages), the Romance languages in the second (cf. syllable-timed), and Japanese in the third group (cf. mora-timed). Yet, under this operational definition of rhythm, categorization was achieved using continuous dimensions of the acoustic signal, allowing for the accommodation of previous 'counterexamples' or mixed languages (Nespor 1990).

The above definition of rhythm, relying on the temporal distribution, but not on the precise identity of consonants and vowels, predicted that languages might be discriminated as long as the relative timings of vocalic and consonantal spaces are preserved, even if the identities of the individual phonemes are suppressed. Ramus and Mehler (1999) confirmed these predictions by showing that adults were able to discriminate rhythmically different languages, even if the signal was resynthesized replacing all vowels by /a/ and all consonants by /s/. However, discrimination failed if all segments were replaced by /a/, suggesting that suprasegmental properties other than rhythm were not sufficient to allow successful discrimination.

Ramus et al. (2000) later extended the above results showing successful discrimination in newborns for the resynthesized stimuli when they were played forward, but not when they were played backward, suggesting that rhythmic discrimination is specific to human speech. Additionally, Ramus et al. (2000) also showed that tamarin monkeys perform similarly to newborns, discriminating forward, but not backward resynthesized speech. This indicates that while rhythmic properties are preferentially computed over human speech as input stimulus, the ability to perform such computations is not restricted to human listeners, and might be a more general property of the primate or mammalian auditory system.

The rhythmic discrimination ability might thus provide a good explanation of how bilinguals exposed to rhythmically different languages distinguish their languages from birth. However, some bilinguals are exposed to rhythmically similar languages. To understand how they discriminate between their languages, Bosch and Sebastián-Gallés (2001) investigated 4-month-old bilingual infants growing up with Spanish and Catalan, and found that infants succeeded on the discrimination task. Monolingual Spanish and monolingual Catalan infants performed similarly. These results suggest that familiarity and experience with at least one of the languages allow discrimination even within the rhythmic group.

Shukla and colleagues, in yet unpublished work following upon Ramus et al.'s (1999) operational definition of rhythm, added further languages (Hungarian, Turkish, Tamil, Basque, Finnish, etc.). The classifications obtained coincided with previous rhythmic analyses of these languages, confirming the validity of the %V, ΔV, and ΔC cues. With the addition of a dozen geographically and typologically different languages, the classification still remained reliable, and interestingly, reflected not only rhythmical, but also morphosyntactic similarities among the languages. Languages with a high %V value were mostly agglutinating and had an Object–Verb (OV) word order (e.g., Japanese, Turkish), while lower %V typically correlated with poorer morphology and a VO order (e.g., English, Dutch). Similar correlations between syllable structure and morphosyntactic type have also been observed by language typologists (Fenk-Oczlon and Fenk 2005). Based on these observations, Mehler et al. (2004) proposed that rhythmical cues, such as %V, might help infants bootstrap into the morphosyntax of their native languages. While to date there is no direct empirical evidence that infants associate rhythmical cues to the relevant structural properties, the hypothesis is very appealing, because it links a robust acoustic cue detected even by neonates to more abstract structural properties.

4. Phoneme Perception

The phoneme, which corresponds roughly to an individual consonant or vowel, is the smallest unit of sound used to distinguish meaning in a language. Although they can discriminate the difference between two instances of the same phoneme (McMurray et al. 2002), adults show enhanced discrimination at phoneme category boundaries (Liberman et al. 1957). When presented with computer-generated stimuli that span two phonemes (e.g., a voice onset time difference from /b/ to /p/), adults label the first several steps along the continuum as one (e.g., /b/), and the next several steps as the other (e.g., /p/), with a sharp boundary in between. The term 'categorical perception' refers to the fact that labeling performance predicts discrimination: adults better discriminate pairs of stimuli of equal sized differences if the two stimuli cross their labeling boundary than they do if the two stimuli are from within a single phoneme category (Repp 1984). Like adults, 8-month-old infants can discriminate some within category phonetic differences (McMurray and Aslin 2005), but importantly, like adults, infants show enhanced discrimination of acoustic/phonetic differences that cross phoneme category boundaries (Eimas et al. 1971; Werker and Lalonde 1988; Dehaene-Lambertz and Dehaene 1994).

The ability to treat acoustic phonetic differences from within a single phoneme category as equivalent is essential for language processing. It is what allows us to rapidly recognize different pronunciations of the same word as equivalent, and hence to instantaneously map sound on to meaning when engaged in discourse. At the same time, the continuing sensitivity to within category phonetic differences is also important for language processing. It provides cues, for example, to word boundaries (Christophe et al. 1994), and is thus essential for tasks such as segmenting the ongoing speech stream into individual words (Werker and Curtin 2005; Curtin and Werker 2007).

Languages differ in their phoneme inventories. English, for instance, contains a contrast between /r/ and /l/ that is lacking in Japanese, but English lacks the retroflex /D/ vs. dental /d/ distinction that is used in Hindi. Very young infants discriminate these phonemes, even though adult non-native speakers fail. In illustration, at 6–8 months English infants discriminate the Hindi retroflex /D/ vs. dental /d/ that English adults find very difficult (Werker et al. 1981). Yet, by the end of the first year of life, English infants perform like English adults and no longer discriminate this (non-native) distinction (Werker and Tees 1984). This oft replicated pattern of broad-based sensitivity in young infants and language-specific phonetic discrimination in older infants and adults

(Saffran et al. 2006) indicates that phonetic perceptual development involves maintenance of initial sensitivities rather than induction of new categories (Aslin et al. 1981; 'universal theory'). In the few documented cases of improvement in phonetic discrimination across infancy, it appears that initial categories have been sharpened (Kuhl et al. 2006; Maye et al. 2008) or realigned (Burns et al. 2007), or that an initial broad distinction has been further refined (see Narayan et al. forthcoming about nasal discrimination). Critically, there is no evidence that language listening experience can induce the formation of a phonemic distinction that was not present in some form in early infancy. Thus, from birth human being has perceptual biases that impose broad based categorizations on phonetic forms.

Considerable research is currently focused on trying to elucidate the mechanisms by which initial phonetic sensitivities become reorganized to match those used in the native language, and on how changing phonetic sensitivities contribute to language acquisition. Because the phoneme is defined as the smallest unit to contrast meaning, initial models assumed that in building a lexicon, infants would learn which phonetic differences to continue to pay attention to and which to ignore (e.g., *ball* and *doll* refer to different objects, hence the /b/–/d/ distinction would be maintained, but the phonetic differences in the pronunciation of the retroflexed [d] in *our doll* and the dental [d] in *this doll* do not contrast meaning and hence would be ignored). This lexical contrast based explanation was eventually dismissed because phonetic perception becomes language specific in the second half of the first year of life, long before a sizeable receptive vocabulary is in place. Moreover, in associative word learning tasks, infants confuse similar sounding words such as 'bin' and 'din' even though they can discriminate the phonemes distinguishing them (Stager and Werker 1997; Werker et al. 2002; Pater et al. 2004; Thiessen 2007; but see Swingley and Aslin 2002).

To explain how phonetic perception changes without a sizeable lexicon, researchers have proposed similarity-matching mechanisms (Best and McRoberts 2003; Kuhl 2004). These models assume that the distributional characteristics of the speech heard highlight those perceptible differences that are phonemic in the language and de-emphasize differences that are not. Thus, even without access to meaning, an infant growing up in Hindi might frequently hear two extreme /d/ sounds whereas an infant growing up in English is more likely to hear variation in pronunciations around a single, intermediate /d/. And indeed, infants can track frequency information and can use that to change their phonetic category boundaries. Infants aged 6–8 months were presented with stimuli from along an 8-step voicing continuum, but the number of times each stimulus was presented yielded a bimodal frequency distribution in one condition and a monomodal frequency distribution in the other (see Figure 5). Following 2–3 min of familiarization, the infants were tested on their ability to discriminate the endpoints of the continuum. Those infants who were

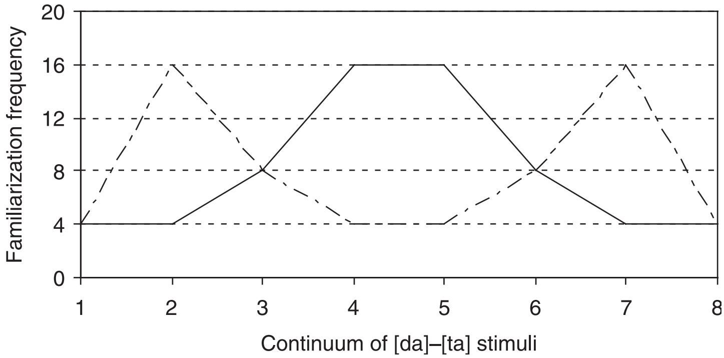

Fig. 5. Bimodal vs. unimodal distributions of [da]–[ta] stimuli during familiarization. The continuum of speech sounds is shown on the abscissa, with Token 1 corresponding to the endpoint [da] stimulus, and Token 8 the endpoint [ta] stimulus. The ordinate axis plots the number of times each stimulus occurred during the familiarization phase. Presentation frequency for infants in the Bimodal group is shown by the dotted line, and for the Unimodal group by the solid line. Reprinted from Maye et al. (2002) with permission.

first familiarized in the bimodal condition showed better discrimination than did infants in the monomodal condition (Maye et al. 2002, 2008). Subsequent studies have shown that adult speech directed to infants contains sufficient distributional phonetic information (Werker et al. 2007) to support a distributional learning mechanism (Vallabha et al. 2007; McMurray et al. forthcoming).

5. Phonotactics

Languages differ not only in the sets of phonemes they use, but also in how they combine them. Phonotactics describes what are the allowed and disallowed phoneme combinations in a language. For instance, Japanese is a strict C(onsonant)–V(owel) language. Its syllables can only consist of Vs or CVs; no other syllable types are possible (with the exception of nasals and the first members of geminate consonants in syllable final positions, for example, Ho**n**da). No consonant clusters are allowed, and words cannot end in consonants. Such restrictions are not only part of speakers' abstract knowledge, but they also directly influence speech perception. Japanese speakers tend to perceive English words as containing additional vowels breaking up the English consonant clusters, for example, *Christmas* is perceived as *K**u**ris**u**mas**u***.

Are infants also sensitive to such phonotactic restrictions? Since these constraints are language-specific, a certain amount of experience with the language is required. But by 9 months of age, as Jusczyk et al. (1993a) demonstrated, infants are able to discriminate words in their native language from words in another language on the basis of the different phoneme sequences that the two languages allow. Later studies (Jusczyk et al. 1993b) suggested that frequency modulates the effects of phonotactics, as 9-month-old American infants preferred phonotactically legal nonsense words with frequent phoneme combinations to those with infrequent sequences.

While the previous studies investigated how much infants know about the phonotactics of their native language through experience, another line of research focused on infants' abilities to learn new phonotactic rules. Onishi, Chambers and colleagues (Onishi et al. 2002; Chambers et al. 2003) taught 16-month-old infants first-order phonotactic constraints restricting the positions of consonants in CVC words (/b/ allowed only in the onset) and second-order phonotactic regularities conditioning the distribution of consonants on their vowel neighbors (/b/ after /ae/, but not after /i/) and have obtained successful learning. Interestingly, conditioning consonantal distributions on the speaker's voice/identity in a third study did not induce learning. Thus, these experiments not only demonstrate infants' rich abilities to learn novel phonotactic constraints, but they also indicate that such learning is restricted to plausible regularities, that is, the ones typically occurring in language.

Saffran and Thiessen (2003) took a further step by exposing 9-month-old infants to phonotactic regularities, and measuring learning through infants' subsequent performance in segmenting out words from a continuous speech stream. During the familiarization phase, they taught infants a rule concerning permissible syllable forms (word consisted of only CV or only CVC syllables, for example, *boda* and *bikrub*, respectively) and another relating to permissible consonantal distributions (voiced stop consonants in syllable onsets, unvoiced stops in codas, for example, *dakdot*, or vice versa). They found successful segmentation in both cases.

6. Segmentation and Learning Word Forms

Segmentation is not only a laboratory task. Young language learners need to identify word forms in speech in order to attach meanings to them and build a lexicon. However, speech is continuous and words are only rarely separated by pauses. Moreover, while the names of a few common objects and actions might occasionally occur in isolation, most words do not. (In fact, grammatical words usually cannot, for example, **is*, **might*, **of*, and **a*.). Indeed, it has been shown that when explicitly asked to teach words to their infants, parents typically do not use isolated words, but embed them in simple carrier phrases, for example, *This is the . . .* , *Look at the . . .* (Aslin et al. 1996).

Segmenting the continuous speech input is thus one of the earliest tasks that language-learning infants encounter. Indeed, some time between 6 months and 7.5 months, they are already able to detect familiar word forms in continuous speech, as work by Jusczyk and Aslin (1995) suggests. The authors familiarized infants with two English words, for example, *bike* and *feet*, for a few minutes, then played them short passages that either contained these words or not. They found that 7.5-month-olds, but not 6-month-olds, preferred to listen to the passages that contained the familiar words.

What information do infants use to identify possible word forms? At least four different types of cues have been identified. First, infants might rely on typical stress patterns, such as the strong–weak (trochaic) pattern commonly found in English content words (e.g., ˈ***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 et al. 1993a; 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. (1999a) have shown that when familiarized with trochaic English words (e.g., ˈ***doc****tor*, ˈ***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 CV syllables where every third syllable was stressed, 7- and 9-month-olds treated as familiar only those trisyllabic sequences that had initial stress (SWW). Infants showed no recognition of trisyllabic sequences that were not trochaic (WSW or WSS; Curtin et al. 2001). The Metric Segmentation Strategy also predicts that weak–strong, that is, iambic, words (e.g., *gui*ˈ***tar***) might initially be missegmented, which turns out to be the case (Jusczyk et al. 1999a).

A second possible cue to segmentation is phonotactics. Knowing that the sequence /br/ is frequent in the initial positions of English words/syllables, while /nt/ typically appears at the end can help the learner posit word boundaries. Indeed, as discussed earlier, 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 CVCCVC words, in which the CC cluster is either frequent word-internally in English, but infrequent across word boundaries (e.g., /ŋk/) or vice versa (e.g., /ŋt/). Infants segmented the nonsense 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).

A third segmentation cue comes from the distributions of allophones in different positions within words. In English, aspirated stop consonants 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. (1999b) 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.

Fourthly, it has long been recognized that the statistical regularities of phoneme co-occurrences are also reliable indicators of morpheme and word boundaries (Harris 1955; Brent and Cartwright 1996). Thus, the syllable /ti/ follows the syllable /prɪ/ with a greater probability than /beɪ/ follows /ti/, for example, as in the sequence *pretty baby*, because /prɪ/ 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 /beɪ/ 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. In their seminal work, Saffran et al. (1996) familiarized infants with a monotonous, continuous artificial speech stream, in which the only cue to word boundaries was the transitional probability[1] (TP) between adjacent syllables. Within-word syllable pairs had TP = 1, while the TP between syllable pairs spanning a word boundary was 0.33. In a subsequent test phase, infants were able to discriminate statistically coherent 'words' from missegmented 'non-words', which never appeared in the stream or contained a dip in TPs word internally.

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.

Identifying potential word forms is indispensable for building a lexicon. But recent work suggests that it might also help bootstrap syntax. Morgan et al. (1995) observed that function words, as opposed to content words, are acoustically and phonologically 'minimal' (short, carry no stress, often reduced, etc.), and subsequent work by Shi and colleagues (Shi et al.

1999, 2006; Shi and Werker 2001, 2003) demonstrated that newborns are able to categorize content and function words based on their forms. In addition, by 6 months, they develop a preference for content words. Moreover, 13-month-old infants are able to use frequent function words as cues to identifying the associated content words (e.g., *the* + N). This acoustic discrimination, together with the high frequency of function words, provides information about word order, an important structural difference among languages, which young learners have been shown to acquire very early on (Brown 1973; Gervain et al. 2008). Indeed, Gervain et al. (2008) has shown that 7-month-old, that is, pre-lexical Japanese and Italian infants prefer their native word order over the word order of the other language.

7. *Perceiving Larger Phonological Units*

Speech is organized into hierarchically embedded prosodic units (Nespor and Vogel 1986), such as utterances, intonational phrases, phonological phrases, and phonological words. Jusczyk et al. (1992) have shown that 6-month-olds are able to detect whether pauses are inserted into passages at real intonational phrase boundaries or at non-boundary positions, and by 9 months, they exhibit a similar sensitivity to phonological phrase boundaries. Work by Christophe et al. (1994) and Jusczyk et al. (1999b) also suggests that newborns and 9-month-olds can detect the boundaries of even smaller units, that is, phonological words.

Although prosodic units are not identical to the corresponding syntactic units, they are aligned with them (Nespor and Vogel 1986), that is, they share at least one boundary. Therefore, infants' ability to perceptually chunk speech into prosodic units might offer cues about the structural units of language. For instance, perceiving prepositional phrases [e.g., *in the box*, *on the table*) as units might, for instance, provide the infant with well-formed syntactic domains from which the relative order of function words and content words, that is, some general word order properties of the target language, can easily be derived. Without such prosodic chunking, irrelevant syntactic chunks might be identified (e.g., **toy is in*, **is in the* from the utterance *Your toy is in the box*).

8. *Early Perceptual Biases*

Research in infant speech perception has mostly focused on how infants perceive or learn about different linguistic units, as reviewed above. Recently, a new line of research emerged, investigating how the inherent representational and processing properties of the auditory system influence speech perception. Similarly to visual Gestalts (e.g., Kanizsa 1955), certain auditory input configurations have been identified that the perceptual system detects in an efficient, automatic manner. In a series of adult studies, Endress et al. (2005, 2007) demonstrated that identical repetitions and sequence edges constitute such perceptual biases, allowing adults to learn regularities that they otherwise fail to acquire (i.e., based on non-repeating structures or not located at sequence edges). Recently, Gervain et al. (forthcoming) have shown using optical imaging that newborn babies are also sensitive to adjacent repetitions and they are able to discriminate trisyllabic sequences containing identical second and third syllables (ABB: 'mubaba', 'penana') from random controls (ABC: 'mubage', 'penaku').

While these auditory biases are general, applying to speech and non-speech sounds alike, they are exploited by the linguistic system and they most probably contribute to the efficiency of language acquisition. Infant-directed speech, for instance, abounds in identical and immediate repetitions of words and phrases (Ferguson 1983). Similarly, typical 'child words' in different languages often contain full or partial reduplications (e.g., *baby*, *daddy* in English; *bébé* 'baby', *dodo* 'sleep' in French; *baba* 'baby', *tata* 'grandpa' in Hungarian; and *papà* 'daddy' in Italian). These are also very often the child's first words. Thus, the perceptual saliency of such words might help infants discover and learn the first entries in their lexicons. Sequence edges, in contrast, appear to be more important for the acquisition of morphological and syntactic structure. Inflectional and derivational morphology universally prefers pre-fixation (beginnings of words: ***re****start*, ***re****play*, etc.) and suffixation (ends of words: *walk****ing***, *talk****ing***, *walk****ed***, *talk****ed***, etc.) to infixation [in Seri, spoken in Mexico, the plural verb stem is formed by inserting *too* into the singular stem after the initial vowel, for example, *ic* 'plant' (singular verb), *itooc* 'plant' (plural verb)]. Also, function words typically occur in phrase initial (prepositions as in English, French, Spanish, etc.) or phrase final (postpositions as in Turkish, Basque, Hungarian, etc.) positions. Therefore, the enhanced processing of these positions might help infants detect and encode morphosyntactic regularities.

9. Conclusion and Future Directions

The research on infant speech perception described above suggests that infants are born with auditory sensitivities that are tuned to human speech. These early abilities allow infants to start learning about their native language. During this learning process, some of their initial capabilities get reorganized: sensitivity to information relevant for the native language(s) is strengthened, while the discrimination of unused features is lost or reduced.

Our understanding of this learning process has increased significantly in the last decades. However, several questions remain and will, in our opinion, shape the research agenda for the years to come. To conclude, we will briefly outline two of these. First, with the advent of efficient and completely non-invasive brain imaging techniques, it has become possible to investigate the neural substrates of infant speech perception. In fact, there is growing consensus that the infant brain shows at least some of the functional and anatomical specializations typical of the mature auditory processing system (Dehaene-Lambertz et al. 2006). For instance, studies by Dehaene-Lambertz et al. (2002) and Peña et al. (2003) have found left lateralization for speech processing in 3-month-olds and newborns (see Figure 3), respectively. Future research into the neural mechanisms of speech perception will explore the developing auditory system in greater detail, answering questions about how and why the early 'window of opportunity' that makes infants such efficient language learners closes during development (e.g., Fagiolini and Hensch 2000; Hensch 2003).

A second issue that has received increasing attention concerns the evolutionary origins of speech perception and language (Hauser et al. 2002). Researchers have found that primates and other animal species perform similarly to infants on several speech perception tasks. Tamarin monkeys, for instance, are able to discriminate languages on the basis of rhythm (Ramus et al. 2000), and both rats and tamarins can segment continuous speech on the basis of statistical information (Hauser et al. 2001; Toro and Trobalon 2005). However, neither tamarins nor rats have language. Therefore, future research will need to clarify what are the subcomponents of language that we share with other species, and what are the evolutionary innovations of humans enabling them to have language.

Short Biographies

Judit Gervain has an MA in theoretical linguistics from the University of Szeged, Hungary, and a PhD in cognitive neuroscience from the International School for Advances Studies, Trieste, Italy. Currently, she is a postdoctoral research fellow at the Department of Psychology at the University of British Columbia. Her interest ranges from formal theories of syntax through computational linguistics to early language acquisition. Her current research projects include the investigation of the perceptual precursors of language acquisition in newborn infants using brain imaging methods, as well as the first acquisitions of language structure during the first year of life. She has recently published a number of papers in these areas in *Cognitive Psychology*, *Proceedings of the National Academy of Sciences of the United States of America*, and *Lingua*.

Janet F. Werker is a Professor and Canada Research Chair in the Department of Psychology at the University of British Columbia. She holds a BA from Harvard University, and an MSc and PhD from the University of British Columbia (UBC). Werker's first academic position was at Dalhousie University before she returned to the UBC. Werker has received recognition for her work with the Killam Research Prize, the UBC Alumni Prize in the Social Sciences, and Fellowships in the Royal Society of Canada, the Canadian Psychological Association, the Canadian

Institutes for Advanced Research, and the American Association for the Advancement of Science. In 2008, she was awarded the Jacob Bieley Research Prize, UBC's most prestigious award. Werker is on the editorial board of six journals spanning cognitive and developmental psychology as well as linguistics, and an editor in the book series, 'Essays in Developmental Psychology'. Werker has contributed empirical and theoretical papers in the broad areas of human experimental, methodological, developmental, cross-cultural, and cognitive neuroscience, with her primary focus on the effects of differential language experience on infant speech perception development and early language acquisition. Classic contributions include early works (Werker et al. 1981; Werker and Tees 1984) showing that listening experience during the first year of life changes infant speech perception but that the link between native speech sound categories and word learning is complex (Stager and Werker 1997; Werker and Yeung 2005; Fennell et al. 2007; Vallabha et al. 2007). More recent work from her laboratory suggests that distributional learning might be a mechanism that leads to changes in speech perception in the first year of life (Maye et al. 2002), that the characteristics of maternal speech support such distributional learning (Werker et al. 2007), and that this learning can be simulated via computational modelling algorithms (Dietrich et al. 2007). More recently, her laboratory has begun a series of studies examining perception and word learning in infants growing up bilingual (Weikum et al. 2007; Fennell et al. 2007; Werker and Byers-Heinlein 2008).

Notes

* Correspondence address: Judit Gervain, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, British Columbia, Canada V6T 1Z4. Email: gervain@psych.ubc.ca.

[1] $TP(X \rightarrow Y) = F(X)/F(XY)$, where $TP(X \rightarrow Y)$ is the transition probability from a syllable X to a subsequent syllable Y, F(X) is the frequency of item X.

Works Cited

Abercrombie, D. 1967. Elements of general phonetics. Edinburgh, UK: Edinburgh University Press.

Aslin, R. N., D. B. Pisoni, B. L. Hennessy, and A. J. Perey. 1981. Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience. Child Development 52.1135–45.

Aslin, R. N., J. Z. Woodward, N. P. LaMendola, and T. G. Bever. 1996. Models of word segmentation in fluent maternal speech to infants. Signal to syntax, ed. by J. L. Morgan and K. Demuth, 263–83. Hillsdale, NJ: Lawrence Erlbaum Associates.

Best, C. T., and G. W. McRoberts. 2003. Infant perception of non-native consonant contrasts that adults assimilate in different ways. Language and Speech 46.183–216.

Bosch, L., and N. Sebastian-Galles. 2001. Early language differentiation in bilingual infants. Trends in bilingual acquisition, ed. by J. Cenoz and F. Genesee, 71–93. Amsterdam, The Netherlands: John Benjamins.

Brent, M. R., and T. A. Cartwright. 1996. Distributional regularity and phonotactic constraints are useful for segmentation. Cognition 61.93–125.

Brown, Roger William. 1973. A first language; the early stages. Cambridge, MA: Harvard University Press.
Burns, Tracey C., Katherine A. Yoshida, Karen Hill, and Janet F. Werker. 2007. The development of phonetic representation in bilingual and monolingual infants. Applied Psycholinguistics 28.455–74.
Chambers, K. E., K. H. Onishi, and C. Fisher. 2003. Infants learn phonotactic regularities from brief auditory experiences. Cognition 87.B69–77.
Chomsky, Noam. 1957. Syntactic structures. The Hague, The Netherlands: Mouton.
——. 1959. A review of B. F. skinner's verbal behavior. Language 35.26–58.
Christophe, A., E. Dupoux, J. Bertoncini, and J. Mehler. 1994. Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. Journal of the Acoustical Society of America 95.1570–80.
Church, K. W. 1987. Phonological parsing and lexical retrieval. Cognition 25.53–69.
Curtin, S., T. H. Mintz, and D. Byrd. 2001. Coarticulatory cues enhance infants' recognition of syllable sequences in speech. Proceedings of the 25th Annual Boston University Conference on Language Development, ed. by A. H. J. Do, L. Dominguez and A. Johansen, 190–201. Somerville, MA: Cascadilla Press.
Curtin, Susan, and Janet F. Werker. 2007. The perceptual foundation of phonological development. The Oxford handbook of psycholinguistics, ed. by Gareth Gaskell, 579–99. Oxford, UK: Oxford University Press.
Cutler, A. 1994. Segmentation problems, rhythmic solutions. Lingua 92.81–104.
Cutler, A., and D. M. Carter. 1987. The predominance of strong initial syllables in the English vocabulary. Computer Speech and Language 2.133–42.
Dauer, Rebecca M. 1983. Stress-timing and syllable-timing reanalyzed. Journal of Phonetics 11.51–62.
Davis, Sally M., and Michael H. Kelly. 1997. Knowledge of the English noun–verb stress difference by native and nonnative speakers. Journal of Memory and Language 36.445–60.
Dehaene-Lambertz, G., and S. Dehaene. 1994. Speed and cerebral correlates of syllable discrimination in infants [see comments]. Nature 370.292–5.
Dehaene-Lambertz, G., L. Hertz-Pannier, J. Dubois, S. Meriaux, A. Roche, M. Sigman, and S. Dehaene. 2006. Functional organization of perisylvian activation during presentation of sentences in preverbal infants. Proceedings of the National Academy of Sciences of the United States of America 103.14240–5.
Dehaene-Lambertz, G., S. Dehaene, and L. Hertz-Pannier. 2002. Functional neuroimaging of speech perception in infants. Science 298.2013–5.
Dietrich, C., D. Swingley, and J. F. Werker. 2007. Native language governs interpretation of salient speech sound differences at 18 months. Proceedings of the National Academy of Sciences of the USA 104(41).16027–31.
Eimas, P. D., E. R. Siqueland, P. W. Jusczyk, and J. Vigorito. 1971. Speech perception in infants. Science 171.303–6.
Endress, A. D., B. J. Scholl, and J. Mehler. 2005. The role of salience in the extraction of algebraic rules. Journal of Experimental Psychology. General 134.406–19.
Endress, A. D., G. Dehaene-Lambertz, and J. Mehler. 2007. Perceptual constraints and the learnability of simple grammars. Cognition 105.577–614.
Fagiolini, M., and T. K. Hensch. 2000. Inhibitory threshold for critical-period activation in primary visual cortex. Nature 404.183–6.
Fenk-Oczlon, G., and A. Fenk. 2005. Crosslinguistic correlations between size of syllables, number of cases, and adposition order. Sprache und natürlichkeit, gedenkband für Willi Mazerthaler, ed. by G. Fenk-Oczlon and C. Winkler. Tübingen, Germany: Narr.
Fennell, C. T., K. Byers-Heinlein, and J. F. Werker. 2007. Using speech sounds to guide word learning: the case of bilingual infants. Child Development 78(5).1510–25.
Ferguson, C. A. 1983. Reduplication in child phonology. Journal of Child Language 10.239–43.
Gervain, J., M. Nespor, R. Mazuka, R. Horie, and J. Mehler. 2008. Bootstrapping word order in prelexical infants: a Japanese-italian cross-linguistic study. Cognitive Psychology 57.56–74.
Gervain, Judit, Francesco Macagno, Silvia Cogoi, Marcela Peña, and Jacques Mehler. forthcoming. The neonate brain extracts speech structure. Proceedings of the National Academy of Sciences USA, <http://www.pnas.org/content/early/2008/09/02/0806530105.abstract>.

Grosjean, François. 1982. Life with two languages: an introduction to bilingualism. Cambridge, MA/London, UK: Harvard University Press.
Harris, Z. 1955. From phoneme to morpheme. Language 31.190–222.
Hauser, M. D., E. L. Newport, and R. N. Aslin. 2001. Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. Cognition 78.B53–64.
Hauser, M. D., N. A. Chomsky, and W. T. Fitch. 2002. The faculty of language: what is it, who has it, and how did it evolve? Science 298.1569–79.
Hensch, T. K. 2003. Controlling the critical period. Neuroscience Research 47.17–22.
Hohne, E. A., and P. W. Jusczyk. 1994. Two-month-old infants' sensitivity to allophonic differences. Perception and Psychophysics 56.613–23.
James, A. Lloyd. 1940. Speech signals in telephony. London, UK: Sir Isaac Pitman & Sons.
Johnson, E. K., and P. W. Jusczyk. 2001. Word segmentation by 8-month-olds: when speech cues count more than statistics. Journal of Memory and Language 44.548–67.
Jusczyk, P. W., and R. N. Aslin. 1995. Infants' detection of the sound patterns of words in fluent speech. Cognitive Psychology 29.1–23.
Jusczyk, P. W., A. Cutler, and N. J. Redanz. 1993a. Infants' preference for the predominant stress patterns of English words. Child Development 64.675–87.
Jusczyk, P. W., A. D. Friederici, J. M. Wessels, V. Y. Svenkerud, and A. M. Jusczyk. 1993b. Infants' sensitivity to the sound patterns of native language words. Journal of Memory and Language 32.402–20.
Jusczyk, P. W., D. M. Houston, and M. R. Newsome. 1999a. The beginnings of word segmentation in English-learning infants. Cognitive Psychology 39.159–207.
Jusczyk, P. W., E. A. Hohne, and A. Bauman. 1999b. Infants' sensitivity to allophonic cues to word segmentation. Perception & Psychophysics 61.1465–76.
Jusczyk, P. W., K. Hirsh-Pasek, D. G. K. Nelson, L. J. Kennedy, A. L. Woodward, and J. Piwoz. 1992. Perception of acoustic correlates of major phrasal units by young infants. Cognitive Psychology 24.252–93.
Kanizsa, G. 1955. Margini quasi-percettivi in campi con stimolazione omogenea. Rivista di Psicologia 49.7–30.
Kuhl, P. K. 2004. Early language acquisition: cracking the speech code. Nature Reviews Neuroscience 5.831–43.
Kuhl, Patricia K., Erica Stevens, Akiko Hayashi, Toshisada Deguchi, Shigeru Kiritani, and Paul Iverson. 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Developmental Science 9.F13–F21.
Ladefoged, P. 1993. A course in phonetics. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.
Liberman, A. M., F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. 1967. Perception of the speech code. Psychological Review 74.431–61.
Liberman, A. M., K. S. Harris, H. S. Hoffman, and B. C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology 54.358–268.
Mattys, S. L., and P. W. Jusczyk. 2001. Phonotactic cues for segmentation of fluent speech by infants. Cognition 78.91–121.
Mattys, S. L., P. W. Jusczyk, P. A. Luce, and J. L. Morgan. 1999. Phonotactic and prosodic effects on word segmentation in infants. Cognitive Psychology 38.465–94.
Maye, J., D. Weiss, and R. N. Aslin. 2008. Statistical phonetic learning in infants: facilitation and feature generalization. Developmental Science 11.122–34.
Maye, J., J. F. Werker, and L. Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82.B101–11.
McMurray, B., and R. N. Aslin. 2005. Infants are sensitive to within-category variation in speech perception. Cognition 95.B15–26.
McMurray, B., M. K. Tanenhaus, and R. N. Aslin. 2002. Gradient effects of within-category phonetic variation on lexical access. Cognition 86.B33–42.
McMurray, B., R. N. Aslin, and J. C. Toscano. forthcoming. Statistical learning of phonetic categories: insights from a computational approach. Developmental Science.
Mehler, J., P. W. Jusczyk, G. Lambertz, N. Halsted, J. Bertoncini, and C. Amiel-Tison. 1988.

A precursor of language acquisition in young infants. Cognition 29.143–78.

Mehler, Jacques, and Emmanuel Dupoux. 1994. What infants know: the new cognitive science of early development. Cambridge, MA: Blackwell.

Mehler, Jacques, Núria Sebástian-Gallés, and Marina Nespor. 2004. Biological foundations of language: language acquisition, cues for parameter setting and the bilingual infant. The new cognitive neuroscience, ed. by M. Gazzaniga, 825–36. Cambridge, MA: MIT Press.

Moon, C., R. P. Cooper, and W. P. Fifer. 1993. Two-day-olds prefer their native language. Infant Behavior and Development 16.495–500.

Moore, D. R. 2002. Auditory development and the role of experience. British Medical Bulletin 63.171–81.

Morgan, J. L. 1996. A rythmic bias in preverbal speech segmentation. Journal of Memory and Language 35.666–88.

Morgan, J. L., and J. R. Saffran. 1995. Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. Child Development 66.911–36.

Morgan, J. L., and K. Demuth. 1996. Signal to syntax: bootstrapping from speech to grammar in early acquisition. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Morgan, J. L., R. P. Meier, and E. L. Newport. 1987. Structural packaging in the input to language learning: contributions of intonational and morphological marking of phrases to the acquistion of language. Cognitive Psychology 19.498–550.

Morgan, J. L., R. Shi, and P. Allopenna. 1995. Perceptual Bases of Rudimentary grammatical categories: toward a broader conceptualization of bootstrapping. Signal to syntax, ed. by J. L. Morgan and K. Demuth, 263–83. Hillsdale, NJ: Lawrence Erlbaum Associates.

Narayan, C. R., J. F. Werker, and P. S. Beddor. forthcoming. Acoustic salience affects speech perception in infancy: evidence from nasal place discrimination.

Nazzi, T., J. Bertoncini, and J. Mehler. 1998. Language discrimination by newborns: toward an understanding of the role of rhythm. Journal of Experimental Psychology: Human Perception and Performance 24.756–66.

Nespor, Marina. 1990. On the rhythm parameter in phonology. Logical issues in language acquisition, ed. by I. Roca, 157–75. Dordrecht, The Netherlands: Foris.

Nespor, Marina, and Irene Vogel. 1986. Prosodic phonology. Dordrecht, The Netherlands/ Riverton, NJ: Foris.

Onishi, K. H., K. E. Chambers, and C. Fisher. 2002. Learning phonotactic constraints from brief auditory experience. Cognition 83.B13–23.

Pater, J., C. L. Stager, and J. F. Werker. 2004. The perceptual acquisition of phonological contrasts. Language 80.384–402.

Peña, M., A. Maki, D. Kovacic, G. Dehaene-Lambertz, H. Koizumi, F. Bouquet, and J. Mehler. 2003. Sounds and silence: an optical topography study of language recognition at birth. Proceedings of the National Academy of Sciences of the United States of America 100.11702–5.

Pike, Kenneth Lee. 1945. The intonation of American English. Ann Arbor, MI: University of Michigan Press.

Ramus, F., and J. Mehler. 1999. Language identification with suprasegmental cues: a study based on speech resynthesis. Journal of the Acoustical Society of America 105.512–21.

Ramus, F., M. D. Hauser, C. Miller, D. Morris, and J. Mehler. 2000. Language discrimination by human newborns and by cotton-top tamarin monkeys. Science 288.349–51.

Ramus, F., M. Nespor, and J. Mehler. 1999. Correlates of linguistic rhythm in the speech signal. Cognition 73.265–92.

Repp, B. H. 1984. Categorical perception: issues, methods and findings. Speech and language. Advances in basic research and practice, ed. by N. Lass, 244–335. Orlando, FL: Academic.

Saffran, J. R., and E. D. Thiessen. 2003. Pattern induction by infant language learners. Developmental Psychology 39.484–94.

Saffran, J. R., R. N. Aslin, and E. L. Newport. 1996. Statistical learning by 8-month-old infants. Science 274.1926–8.

Saffran, Jenny R., Janet F. Werker, and Lynne A. Werner. 2006. The infant's auditory world: hearing, speech, and the beginnings of language. Handbook of child psychology: Vol 2. Cognition, perception, and language (6th edn), ed. by Deanna Kuhn, Robert S. Siegler, William Damon and Richard M. Lerner, 58–108. Hoboken, NJ: John Wiley & Sons Inc.

Samuel, Arthur G. 1996. Does lexical information influence the perceptual restoration of phonemes? Journal of Experimental Psychology: General 125.28–51.
Sansavini, A., J. Bertoncini, and G. Giovanelli. 1997. Newborns discriminate the rhythm of multisyllabic stressed words. Developmental Psychology 33.3–11.
Shi, R., A. Cutler, J. Werker, and M. Cruickshank. 2006. Frequency and form as determinants of functor sensitivity in English-acquiring infants. The Journal of the Acoustical Society of America 119.EL61–7.
Shi, R., and J. F. Werker. 2001. Six-month-old infants' preference for lexical words. Psychological Science 12.71–6.
——. 2003. The basis of preference for lexical words in 6-month-old infants. Developmental Science 6.484–8.
Shi, R., J. F. Werker, and J. L. Morgan. 1999. Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. Cognition 72.B11–21.
Stager, C. L., and J. F. Werker. 1997. Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature 388.381–2.
Swingley, D., and R. N. Aslin. 2002. Lexical neighborhoods and the word-form representations of 14-month-olds. Psychological Science 13.480–4.
Thiessen, Erik D. 2007. The effect of distributional information on children's use of phonemic contrasts. Journal of Memory and Language 56.16–34.
Thorburn, Geoffrey D., and Richard Harding. 1994. Textbook of fetal physiology. Oxford, UK/New York, NY: Oxford University Press.
Tomasello, Michael, and Elizabeth Bates. 2001. Language development: the essential readings. Oxford, UK/Malden, MA: Blackwell Publishers.
Toro, J. M., and J. B. Trobalon. 2005. Statistical computations over a speech stream in a rodent. Perception & Psychophysics 67.867–75.
Vallabha, Gautam K., James L. McClelland, Ferran Pons, Janet F. Werker, and Shigeaki Amano. 2007. Unsupervised learning of vowel categories from infant-directed speech. PNAS Proceedings of the National Academy of Sciences of the United States of America 104.13273–8.
Vouloumanos, A., and J. F. Werker. 2004. Tuned to the signal: the privileged status of speech for young infants. Developmental Science 7.270.
——. 2007. Listening to language at birth: evidence for a bias for speech in neonates. Developmental Science 10.159–64.
Weikum, W., A. Vouloumanos, J. Navarro, S. Soto-Faraco, N. Sebastian-Galles, and J. F. Werker. 2007. Visual language discrimination in infancy. Science 316(5828).1159.
Werker, J. F., and Krista Byers-Heinlein. 2008. Bilingualism in infancy: first steps in perception and comprehension. Trends in Cognitive Sciences 12(4).144–51.
Werker, J. F., and Susan Curtin. 2005. PRIMIR: a developmental model of speech processing. Language Learning and Development 1.197–234.
Werker, J. F., C. T. Fennell, K. M. Corcoran, and C. L. Stager. 2002. Infants' ability to learn phonetically similar words: effects of age and vocabulary size. Infancy 3.1–30.
Werker, J. F., J. H. V. Gilbert, K. Humphrey, and R. C. Tees. 1981. Developmental aspects of cross-language speech perception. Child Development 52.349–55.
Werker, J. F., and H. H. Yeung. 2005. Infant speech perception bootstraps word learning. Trends in Cognitive Science 9(11).519–27.
Werker, J. F., and C. E. Lalonde. 1988. Cross-language speech perception: initial capabilities and developmental change. Developmental Psychology 24.672–83.
Werker, J. F., F. Pons, C. Dietrich, S. Kajikawa, L. Fais, and S. Amano. 2007. Infant-directed speech supports phonetic category learning in English and Japanese. Cognition 103(1).147–62.
Werker, J. F., and R. C. Tees. 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behavior and Development 7.49–63.