3

Infant Speech Perception

Derek Houston

Introduction

Speech perception can be described as a mode of hearing specialized for speech. When people engage in conversation, they do not hear simply the information conveyed in a waveform or a spectrogram. Instead, they perceive linguistic and indexical information that conveys words and sentences as well as identifying qualitative characteristics of the talkers. People are able to extract linguistic and indexical information from speech because of the specialized way the human hearing instrument is tuned. Accordingly, the study of infant speech perception is concerned with the tuning of that instrument during early development.

A complete understanding of how speech perception develops would require descriptions of the initial and end states of infants' speech perception and an explanation of how the change of state happens. The field of infant speech perception is not yet able to describe with certainty exactly what information infants perceive from speech at any stage of development or what drives speech perception to change with development and language experience. Nevertheless, developmental scientists have made a great deal of progress over the last 40 years toward these goals.

Early work in infant speech perception was strongly influenced by Noam Chomsky's theories, which were revolutionizing the field of linguistics at the time. Chomsky (1968, 1975) posited that language was not learnable from the input alone and required a specialized universal language acquisition device that was innately endowed to humans. Because of this prevailing view, most work focused on identifying speech perception skills that were thought to be innate and universal. Gradually, however, infant speech perception scientists have focused increasingly more on what infants are able to learn from the input, and mounting evidence suggests that general learning mechanisms may play a larger role in language acquisition than previously thought (Behme & Deacon, 2008).

This chapter reviews what is known about normal-hearing infants' initial preferences and sensitivities to both suprasegmental properties of speech (e.g., stress, intonation, rhythm) and fine-grained information contained in phonemes. We then discuss learning mechanisms that have been found to play a role in tuning infants' speech perception skills. Finally, we describe the speech perception skills that infants acquire as a result of these learning mechanisms, which put them in a position to build a vocabulary and acquire language.

Birth—Six Months: Initial Preferences and Sensitivities

The aim of this section is to describe infants' speech perception skills during the early postnatal period. This period may not reflect infants' initial state, however, because fetuses are able to hear and learn in utero. Fetuses demonstrate consistent responses to auditory stimulation by 25 to 29 weeks gestational age (Birnholz & Benacerraf, 1983), which means that full-term newborns have more than two months of auditory experience. In utero experience may shape the initial postnatal state, and we consider its role in the initial postnatal state of speech perception.

The fetal hearing experience may affect some aspects of speech perception more than others. Studies of the in utero acoustic environment suggest that frequencies above 1000 Hz are attenuated 20 to 30 dB in transmission to fetuses (e.g., Lecanuet et al., 1998), suggesting that they have, at best, limited access to acoustic information important for discriminating segmental information, especially consonants. However, the sound that does reach fetuses provides sufficient information for perception of suprasegmental aspects of speech.

Perception of Suprasegmental Information

Suprasegmental information, such as intonation and rhythm, transmits very well to fetuses. Several research teams have investigated how prenatal experience with suprasegmental information affects infants' early speech perception. One issue of interest is whether or not fetuses encode suprasegmental information into memory. If so, their in utero experience with speech may affect their initial postnatal speech processing and preferences.

Evidence of Early Encoding

There is evidence that fetuses not only have access to auditory information, they also encode speech information into memory—especially suprasegmental information. Both newborns and fetuses demonstrate the ability to discriminate their native language from a foreign language (Kisilevsky et al., 2009; Mehler et al., 1988; Moon, Cooper, & Fifer, 1993; Nazzi, Bertoncini, & Mehler, 1998) and their mother's voice from another woman's voice (DeCasper & Fifer, 1980; Kisilevsky et al., 2003). Because most fine-grained segmental information is filtered out before it reaches the fetus, these findings suggest that fetuses encode some suprasegmental characteristics of speech.

To explore the possibility that suprasegmental properties encoded during the last trimester persist in memory, DeCasper and colleagues investigated newborns' and fetuses' memory of their mothers' speech. They instructed women to read a nursery rhyme three times a day, starting six weeks before their due date. Newborns showed a preference for the familiar nursery rhyme over a novel nursery rhyme (DeCasper & Spence, 1986). A follow-up investigation found that fetuses who were tested two weeks before birth showed differential heart rate responses for the familiar compared to a novel rhyme (DeCasper, LeCanuet, Busnel, Granier-Deferre, & Maugeais, 1994). Thus, four weeks of exposure for a few minutes a day is *sufficient* for fetuses to encode some properties of a nursery rhyme. However, it is unknown at this time how much exposure is *necessary* for fetuses to encode suprasegmental properties of their language, their mother's voice, or a nursery rhyme. Moreover, we do not know how long suprasegmental information like this persists in memory.

Sensitivity to Rhythm and Intonation

The above findings suggest that infants are highly attuned to what they hear in utero. But what aspects of speech are they encoding? One possibility is that infants' ability to discriminate what they hear in utero from novel stimuli may reflect simple pattern-matching skills. However, their discrimination abilities may instead reflect a more general attunement to suprasegmental properties of speech. Nazzi, Bertoncini, and Mehler (1998) investigated this possibility by testing newborn infants' discrimination of languages they were not exposed to in utero. Specifically, they investigated the role of rhythm—the timing of syllables in a language—in infants' language discrimination. French newborns demonstrated discrimination of unfamiliar languages that were rhythmically dissimilar (English and Japanese) but not languages that were rhythmically similar (English and Dutch), suggesting they had a general sensitivity to rhythmic information in speech.

As discussed later in the chapter, infants' sensitivity to the rhythmic properties of speech appears to play an important role in their later speech segmentation skills, that is, their ability to locate word boundaries in the context of fluent speech. It is therefore tempting to assume that infants' ability to discriminate languages that are rhythmically different means that they are especially sensitive to rhythmic properties of speech. However, the rhythm of speech correlates strongly with its intonation, and it is not clear if infants rely mainly on rhythm or intonation to discriminate languages (Ramus, 2002). Whatever the case, infants' sensitivity to suprasegmental properties of speech seems to be general rather than limited to only the speech they were exposed to in utero. In the next section, we discuss the role that infant sensitivity to suprasegmental properties might play in their processing of speech.

Effects of Rhythm and Intonation on Early Speech Processing: Preference for Infant-Directed Speech

Not only are infants able to discriminate suprasegmental properties, several investigations suggest that these properties play a role in infants' attention to speech and what information they extract. With respect to attention, adults speak differently to infants than they do to other adults or even older children, especially when they want to engage infants' attention. *Infant-directed speech* (IDS) and *adult-directed speech* (ADS) differ in their rhythmic and intonational properties. IDS is characterized as being slower (longer durations of syllables and pauses), higher pitched, and having greater pitch excursions than ADS. Infants demonstrate greater attention to IDS than to ADS, at least during the first six months of life (Cooper & Aslin, 1990; Fernald, 1985; Fernald & Kuhl, 1987; Werker & McLeod, 1989).

The advantage of IDS over ADS in capturing infants' attention does not appear to depend much, if at all, on infants' experience with IDS. Cooper and Aslin (1990) assessed attention to IDS and ADS in 2-day-olds and 1-month-olds and found that both groups of infants demonstrated longer looking times when presented with IDS than when presented with ADS. Given our previous discussion of fetal sensitivities to speech, we might speculate that in utero experience plays a role in infants' attention to IDS. However, fetuses are exposed to ADS more than to IDS (unless the pregnant mother speaks more to infants than to adults and older children); thus, it seems very unlikely that newborns' increased attention to IDS could be due to their exposure to speech (which is mainly ADS) during the fetal period.

If infants' attention to IDS is not due to experience, then it is likely something about the acoustic properties of IDS that draws infants' attention to it. To investigate the relative contributions of the pitch (as measured by F0), intensity, and duration characteristics of IDS on capturing infants' attention, Fernald and Kuhl (1987) presented infants with one of three types of IDS and ADS computer-synthesized speech. Each type of synthesized speech preserved one characteristic that differentiated IDS and ADS and equalized the other two characteristics; all speech types were devoid of any lexical information. The investigators found that 4-month-olds showed greater attention to the IDS only when the pitch was preserved, suggesting that the pitch characteristics of IDS is what captures infants' attention (Fernald & Kuhl, 1987). Follow-up studies suggested that the aspect of pitch most important for infants' preference for IDS is intonation rather than the mean pitch height (Fernald, 1993). More recently, however, Singh, Morgan, and Best (2002) found that 6-month-olds preferred speech that conveyed positive affect to speech that conveyed neutral affect, regardless of whether it was IDS or ADS. Moreover, infants showed no preference for IDS over ADS when they controlled for affect. In fact, infants preferred ADS to IDS when presented with positive-affect ADS and neutral-affect IDS. Also, infants respond more positively to IDS that expresses approval than IDS that expresses disapproval (Fernald, 1993). These findings suggest that infants' prefer IDS because it generally conveys a positive affect, which is carried primarily through intonation (Fernald, 1989).

Taken together, research on young infants' sensitivity to suprasegmental information suggests that they are attuned to the rhythmic and intonational properties of speech at birth and even before. They prefer intonation that conveys emotional information. There is no strong evidence that infants have a similar kind of preference for a particular type of rhythm, but they are able to discriminate rhythmic differences at very young ages. It is not clear what drives infants to attend to rhythmic properties. It is possible that because intonation and rhythm are highly correlated, attention to intonation may contribute to sensitivity to rhythm. Infants' attention to rhythmic properties plays an important role in the development of more advanced speech perception skills, as we discuss later in this chapter. For now, however, we continue to focus on speech perception skills during early infancy. We now turn to their perceptual sensitivities to segmental information in speech.

Perception of Segmental Information

Segmental information refers to the acoustic properties of speech that differentiate phonemes. Given that much of the high-frequency acoustic information that distinguishes phonemes is not available to fetuses because of in utero filtering, we might expect young infants to be poor at discriminating phonemes. But despite the lack of experience with high-frequency segmental information in utero, newborns and young infants demonstrate sensitivities to fine-grained changes in segmental information.

Categorical Perception

The earliest infant speech perception research was motivated by findings that adults perceived some acoustic-phonetic dimensions categorically rather than continuously (Liberman, Harris, Hofman, & Griffith, 1957). For example, *voice onset time* (VOT) is the time between the release of an articulation for a stop consonant (e.g., [p, b, t, d, k, g]) and the onset of voicing. Although VOT varies along a continuum, adults perceive VOT categorically (Liberman, Harris, Kinney, & Lane, 1961). In English, stop consonants with relatively short VOTs (0–20 msec) are perceived as *voiced* (e.g., [b, d, g]) and those with relatively long VOTs

(>30 msec) are perceived as *voiceless* (e.g., [p, t, k]). They are considered to be perceived categorically because listeners are very poor at discriminating within-category changes in VOT (e.g., 0 and 20 msec VOTs or 40 and 60 msec VOTs) but can readily discriminate changes in VOT that cross VOT categories (e.g., 20 and 40 msec VOTs) even when the objective differences in VOT are identical. At issue in the late 1960s was whether or not categorical perception was due to innate auditory sensitivities or due to experience—learning the phonology of the ambient language.

To test whether or not infants were innately endowed with speech discrimination abilities that were attuned to language, Peter Eimas and his colleagues tested 1- and 4-month-old infants' ability to discriminate synthesized versions of [ba]-[pa] that varied along the VOT continuum. To test young infants, Eimas, Siqueland, Jusczyk, and Vigorito (1971) used the high amplitude sucking (HAS) paradigm. In the HAS infants suck on a non-nutritive pacifier that is connected to a computer that registers each suck. During the habituation phase, infants are presented with one stimulus each time they suck until their sucking rate decreases to a habituation criterion. They are then presented with a novel stimulus (experimental group) or the same stimulus (control group) and their sucking times are analyzed to determine if the presentation of the novel stimulus elicits an increase in sucking rate, suggesting infants can discriminate the two stimuli. Eimas and colleagues found that 1- and 4-month-old infants showed discrimination of the same VOTs as adults, suggesting that they also perceived VOT categorically. Numerous follow-up studies have shown that infants categorically discriminate voicing, place-of-articulation, and manner-of-articulation (e.g., Eimas, 1974, 1975; Eimas & Miller, 1980a, 1980b; Eimas et al., 1971), suggesting that infants are born with a perceptual system that is tuned to detect acoustic-phonetic properties important for identifying phonemes in many of the world's languages.

Some more recent work suggests that infants' discrimination of consonants is not as fixed or as strictly categorical as previously thought. Maye, Werker, and Gerken (2002) tested the effects of input on infants' discrimination of VOT contrasts. They familiarized infants with repetitions of eight unaspirated alveolar stops that varied in VOT from [da] to [ta]. One group of infants was presented with relatively more tokens from the middle of the VOT range (unimodal distribution), while the other group was presented with relatively more tokens from the two endpoints of the VOT range (bimodal distribution). They then tested infants' discrimination of [da] and [ta] and found that only the group familiarized with the bimodal distribution demonstrated discrimination.

Whereas Maye et al. (2002) found that infants can fail to discriminate across category boundaries under some stimulus conditions, McMurray and Aslin (2005) found that infants can discriminate VOTs within category boundaries under some testing conditions. They used a head-turn preference procedure (described in more detail later) to assess 8-month-olds' discrimination of prototypical and nonprototypical tokens of [ba] and [pa]. Unlike previous studies that used a habituation/dishabituation procedure, McMurray and Aslin (2005) found that infants could discriminate tokens that fell within phoneme categories. Taken together, these findings suggest that while infants may have some initial auditory sensitivities to particular acoustic-phonetic cues, these sensitivities are not rigid and can be influenced by linguistic input.

Although the above work suggests that many consonant contrasts may be perceived categorically, investigations on the perception of vowels suggest that they are perceived more continuously. Unlike consonants, adults discriminate steady-state vowels in a continuous rather than a categorical manner (Fry, Abramson, Eimas, & Liberman, 1962; Pisoni, 1973; Stevens, Liberman, Studdert-Kennedy, & Ohman, 1969). Swoboda, Morse, and Leavitt (1976) discovered that 2-month-olds not only discriminated [i] and [ê] but also discriminated vowel sounds that fell within the same vowel category but differed with respect to formant frequencies, suggesting that, like adults, young infants also perceive vowels in a continuous manner.

Sensitivity to Phoneme Inventory

The above findings suggest that, like their ability to discriminate suprasegmental properties of speech, young infants are able to discriminate segmental properties after little to no experience with language. The above findings tell us very little, however, about infants' ability to encode phonemes into long-term memory. Evidence that they can encode suprasegmental information into long-term memory comes from studies reviewed in the previous section in which infants respond differently to the familiar rhythmic and intonational characteristics of their mothers' speech than to the speech of another woman and to a familiar nursery rhyme than to an unfamiliar one. Responding to familiarity requires not only the ability to discriminate familiar and unfamiliar stimuli, but also the ability to associate a familiar sample of speech to representations of that speech stored in long-term memory. In contrast, a finding of discrimination where one stimulus is presented until habituation is reached and then a novel stimulus is presented requires infants to store speech information into memory for only a very brief period of time.

If infants are able to encode segmental information of the ambient language into their long-term memory as they do suprasegmental information, they should show similar attentional preferences for native segmental information as they do for native suprasegmental information. Jusczyk, Friederici, Wessels, Svenkerud, and Jusczyk (1993) tested this possibility by assessing 6- and 9-month-old English-learning infants' preferences for English words versus foreign words that differed from English words in phoneme inventories, rhythmic properties, or both. Infants demonstrated longer looking times for their native rhythm versus a foreign rhythm (low-pass filtered Norwegian) but not for their native phoneme inventory versus a language with a similar rhythm but dissimilar phoneme inventory (Dutch). Nine-month-olds, in contrast, showed preferences based on both rhythm and phoneme inventories, suggesting that familiarity with native segmental characteristics emerges later than familiarity with native suprasegmental characteristics.

Effects of Language Experience on Speech Discrimination

Many phonemes are common across most languages, and early work on speech discrimination in infants focused on their ability to discriminate those common phonemes. However, some phonemes are particular to one or just a few of the world's languages. For example, in Hindi there are two types of "d" sounds. One is similar to the English "d" ([d]—produced by the tongue releasing down from the teeth); the other—a *retroflex* "d" [ɖ]—is produced by pulling the tongue back from the teeth. Non-Hindi speakers have difficulty hearing the difference between these two "d" sounds. Werker and Tees (1984) wanted to know if infants could discriminate phonemic contrasts that occurred in some languages but which were difficult for adults who did not speak those languages to discriminate. They tested three age groups (6–8 months, 8–10 months, and 10–12 months) from three language backgrounds (English, Hindi, and Nthlakapmx) on several consonant contrasts. Werker and Tees (1984) found that younger infants were able to discriminate all of the contrasts but 10- to 12-month-olds could discriminate only those that were linguistically relevant in their native language, suggesting that consonant discrimination is affected by language input.

The effect of language input on speech discrimination has been investigated with vowels as well. Kuhl, Williams, Lacerda, Stevens, and Lindblom (1992) tested English-learning and Swedish-learning infants' discrimination of variants of the English vowel [i] and the Swedish vowel [y]. The two groups of infants showed different patterns of results, suggesting that language background affected their discrimination of the vowels. Specifically, when infants were presented with variants of [i], English-learning infants were less likely to discriminate two variants that were acoustically similar to the prototypical English [i] than two variants that were less similar to the prototype, even though the variants in each pair were equally similar to each other. Swedish-learning infants, by contrast, were just as likely to discriminate both pairs of the English [i] variants. When infants were presented with variants of the Swedish vowel [y], the group differences were reversed: only the Swedish-learning infants' discrimination was affected by similarity to the prototype. Kuhl (1991, 1993) described these findings as representing a "perceptual magnet effect" in which the distribution of vowel variants in the ambient language shapes infants' perceptual systems such that they perceive variants within a vowel category to be more like the prototype of that category. Polka and Bohn (1996), however, found no evidence of a perceptual magnet effect when they tested English-learning and German-learning infants' discrimination of the German and English contrasts. Instead, these findings and others suggest that vowels on the periphery of the F1/F2 acoustic space serve as universal perceptual attractors (Polka & Bohn, 2003).

These and similar findings (Best, McRoberts, & Sithole, 1988; Trehub, 1976; Tsushima et al., 1994; Werker & Lalonde, 1988) led to a *universalist* view of infant speech discrimination—that infants are born able to discriminate any phonemic contrast that could potentially be relevant to any of the world's languages; and then, with experience, infants lose the ability to discriminate contrasts that are not relevant for their language (e.g., Eimas, Miller, & Jusczyk, 1987; Werker & Pegg, 1992). Since then, however, the picture of infants' speech discrimination abilities has become more complex. For example, while some non-native phoneme contrasts may fall into the same phoneme category in English (e.g., the Hindi [ɖ]) many speech sounds (e.g., African clicks) do not fall into any phonemic category of English speakers. Best, McRoberts, Lafleur, and Silver-Isenstadt (1995) found that such contrasts remain easy to discriminate for English-speaking adults and older infants. These findings provide evidence against a strong universalist view that infants lose the ability to discriminate all sounds that are not linguistically relevant.

One limitation of a universalist view of speech discrimination is that it does not take into consideration subphonemic information that, while not relevant for distinguishing words, is relevant for other aspects of speech perception and language acquisition. Allophones (context-dependent variants of phonemes) specify details for how words are produced in the native language and can play a role in speech segmentation. For example, initial stops are aspirated in English but not in French—"port" is pronounced [pʰort] in English, but in French "porte" (door) is pronounced [port]. Stops are unaspirated in other positions in English (e.g., "sport"). To sound like a native English speaker, English-learning infants must encode allophonic information even though it does not differentiate words. Hohne and Jusczyk (1994) tested infants' discrimination of words and word pairs such as "nitrates" and "night rates." The same strings of phonemes comprise these sequences but differ with respect to some of the allophonic information: the "t" in "nitrates" is aspirated, released, and retroflexed; the "t" in "night rates" is unaspirated, unreleased, and not retroflexed. Also, the "r" is devoiced in "nitrates" but not in "night rates." Two-month-olds demonstrated discrimination of these allophones (Hohne & Jusczyk, 1994). Subsequent investigations of infants' use of allophonic information (discussed below) suggest that infants do not lose their ability to discriminate this fine-grained information.

Another challenge to a universalist view of speech discrimination is that some contrasts, rather than being discriminable universally during early infancy, require language experience before they can be discriminated. Lacerda (1993) found that Swedish-learning 6- to 12-month-olds could discriminate between [a] and [ʌ] but not between [a] and [ɑ], both of which are linguistically relevant in Swedish. Similarly, Lasky, Syrdal-Lasky, and Klein (1975) tested Spanish-learning 4.5- to 6-month-olds on three different VOT contrasts. They found that the Spanish-learning infants were able to discriminate a pair of speech sounds that was irrelevant for Spanish but relevant for English, but did not discriminate a contrast that is distinctive in Spanish. Recent investigations have demonstrated that discrimination of some contrasts improves with language experience from infancy through adulthood (Polka, Colantonio, & Sundara, 2001; Sundara, Polka, & Genesee, 2006; Tsao, Lui, & Kuhl, 2006). For example, English-learning infants and children improve in their ability to discriminate the [d]-[ð] contrast, whereas French-learning infants and children do not (Polka et al., 2001; Sundara et al., 2006). Taken together, the findings point to an early perceptual system that is able to discriminate most contrasts of the world's languages, and then through experience with language input, infants become more sensitive to sounds that are relevant for their language and less sensitive to contrasts that are not linguistically relevant.

Six Months to One Year: Demonstrations of Learning

Given that infants seem to learn something about the organization of sounds in their language by the second half of the first year of life, it is worth considering what kinds of learning mechanisms are required to allow this auditory-perceptual learning to occur. This section will describe some of the specific learning abilities of infants that may interact with their innate auditory and perceptual abilities to transition them from a universal perceiver to having a perceptual system tuned to the native language.

Mechanisms of Learning in Infants

Several learning mechanisms contribute to the development of language-specific speech perception skills. These learning mechanisms include (but are not limited to) recognition memory, associative learning, and statistical learning. Each of these learning mechanisms has been studied extensively by developmental scientists, and a full review of them is beyond the scope of this chapter. Instead, we will briefly describe what these mechanisms are and some evidence that infants possess these learning mechanisms.

Recognition Memory

Recognition is a very basic form of learning. In order to recognize something, it must be encoded into memory. Visual recognition memory has been investigated much more than auditory recognition memory. One way developmental scientists have investigated visual recognition memory is by using habituation/dishabituation paradigms (Colombo, Shaddy, Richman, Maikranz, & Blaga, 2004; Fagan & McGrath, 1981; Rose & Feldman, 1997; Rose, Feldman, & Jankowski, 2001). Infants are presented with an object or photograph of a face repeatedly until they habituate to it (i.e., decrease their looking time). Then they are presented with both a novel and the habituated object. Longer looking to the novel than to the habituated object indicates recognition of the object the child has already seen. Recognition memory improves significantly during the first year of life (Rose et al., 2001) and correlates with later cognitive and language outcomes (Rose, Feldman, & Wallace, 1992; Rose, Feldman, Wallace, & McCarton, 1991), suggesting that it is an important cognitive skill for cognitive and language development.

Some examples of infants' recognition memory for speech have already been reviewed above. Preferences for native language and mother's voice suggest recognition. Work with older infants suggests that infants' representations of speech sounds become more generalizable with experience and development. For example, Houston and Jusczyk (2000) tested infants' ability to recognize familiarized words when presented with a different voice. They found that 10.5-month-olds but not 7.5-month-olds were able to recognize words across talkers of the opposite sex, suggesting that 7.5-month-olds encode talker-specific information in memory and that this affects how they recognize words (see also Houston, 1999; Houston & Jusczyk, 2003). These findings illustrate how the development of recognition memory skills is important for correctly identifying words as novel and old, which is an important skill for learning the meaning of words across different contexts. Later, we discuss some additional examples of how the development of recognition memory skills affects speech perception in infants.

Associative Learning

Associative learning is highly relevant to language acquisition; word learning is a sophisticated type of associative learning. But well before infants utter their first words, their associative learning skills develop in nonlinguistic domains. In the visual domain, early associative learning plays an important role in forming categories of objects. Younger and Cohen (1986) investigated 4-, 7-, and 10-month-olds' use of correlated attributes (e.g., long neck associated with large ears and short neck associated with small ears) to form categories of novel animal drawings. They found that normal hearing 7-month-olds but not 4-month-olds could learn correlations among attributes when all attributes were perfectly correlated and that normal hearing 10-month-olds but not 7-month-olds could learn correlations among attributes when some of the attributes were correlated and others were not. Similar studies have found that older infants can learn correlations among objects' parts and their motion trajectories (Rakison & Poulin-Dubois, 2002).

In the auditory domain, young infants can learn simple associations, such as the relationship between vocal affect and facial expressions (Kahana-Kalman & Walker-Andrews, 2001; Walker, 1982; Walker-Andrews, 1986). Older infants learn to associate complex strings of speech sounds (i.e., words) with objects, actions, attributes, and experiences. The development of associative learning skills plays important roles in various aspects of language acquisition, some of which will be discussed later.

Statistical Learning

Statistical learning is related to associative learning. But rather than learning that *x* is associated with *y*, statistical learning involves learning the probability of *y* given *x*. In the visual domain, infants' statistical learning skills have been investigated by assessing their ability to learn visual sequences. Young infants (3- to 4-month-olds) can learn simple two- and three-location spatiotemporal patterns (Clohessy, Posner, & Rothbart, 2001; Haith, Hazan, & Goodman, 1988; Wentworth, Haith, & Hood, 2002), whereas older infants are able to learn more complex spatiotemporal sequences (Clohessy et al., 2001; Kirkham, Slemmer, Richardson, & Johnson, 2007).

In a seminal study on auditory statistical learning, Saffran, Aslin and Newport (1996) tested 8-month-olds' ability to detect syllable sequences within a two-minute continuous stream of synthetic CV syllables. The speech stream consisted of four three-syllable sequences with no pauses between sequences. Thus, the only way infants could learn the sequences was by encoding the transitional probabilities of the syllables. For example, if one of the four sequences was /da/ro/pi/ then the probability of /ro/ following /da/ and of /pi/ following /ro/ would both be 1.0. However, /pi/ would be followed by one of three syllables. Saffran et al. found that 8-month-olds showed a novelty preference for sequences that had lower transitional probabilities in the speech stream (e.g., /pi/go/) compared with sequences that had high transitional probabilities (e.g., /da/ro/). These findings and others suggest that infants are sensitive to the statistical properties of speech sounds in their language.

Motivation for Social Interaction and Exploration

Although children readily learn language, it does require some effort. Innate speech perception capacities and learning mechanisms do not by themselves explain language acquisition. Infants' motivations and intentions play an important role in language development (Bloom & Tinker, 2001). Infants are dependent on their caregivers for physical and emotional needs and are thus motivated to communicate with their caregivers (Locke, 1993). What infants attend to in speech depends on their needs, which change with development. At the beginning of life, infants may seek only social-emotional information from speech and may attend mainly to prosodic information. As infants become more sophisticated, they attend to other aspects of speech that are more relevant to language acquisition. Moreover, motivation to attend to one type of information (e.g., prosody for affect) may set the groundwork for acquiring knowledge useful for obtaining other types of information (e.g., word boundaries) that are useful for obtaining later goals (e.g., understanding what the caregiver is trying to communicate).

Infants' speech perception changes through the interaction of cognitive, social, and linguistic factors. With development, infants are motivated for increasingly more sophisticated communication. Attention to speech and increasingly sophisticated learning mechanisms result in infants forming mental representations that shape how speech is perceived. And because the input to the learning mechanisms differs across languages, infants form language-specific representations that result in language-specific perception of speech.

Organizing the Suprasegmentals

We learned earlier that newborns are able to discriminate the rhythmic properties of languages when languages fall into different rhythmic classes. That initial sensitivity to rhythmic information forms the basis for the ability to detect rhythmic information that relates to linguistic units in speech, such as clauses, phrases, and words. Being able to detect these linguistic structures may play a role in infants' ability to develop a vocabulary and acquire a grammar.

Utterance-Level Prosody

Utterances tend to contain one or more clauses, and clauses contain one or more phrases. An implicit understanding of clausal and phrasal organization in speech is important for language comprehension and production. Clauses and phrases influence the prosody of speech. For example, pauses tend to occur more often at clause and phrase boundaries than within clauses or phrases. It is possible that perceiving grammatical units within utterances (e.g., clauses) may be a first step in acquiring a grammar.

Hirsh-Pasek et al. (1987) investigated infants' sensitivity to prosodic markings of clause boundaries in fluent speech. They presented 6- and 9-month-olds with passages of natural infant-directed speech in which 1-second pauses were inserted either between or within clauses. Both groups of infants looked longer when pauses were between clauses than when they were within clauses, suggesting that by 6 months of age, English-learning infants have become familiarized with the prosodic cohesiveness of clauses in English.

Infants' sensitivity to the prosody of syntactic structures appears to also play a role in recognizing familiar sequences of words in the context of fluent speech. Nazzi, Kemler Nelson, Jusczyk, and Jusczyk (2000) investigated this in English-learning infants. Six-month-olds were familiarized with sequences of words (e.g., rabbits eat leafy vegetables) and then presented with passages in which the sequence of words occurred either within a clause (e.g., . . . rabbits eat leafy vegetables . . . .) or between clauses (e.g., . . . rabbits eat. Leafy vegetables . . .). Infants demonstrated recognition of the words only when they occurred within clauses (see also Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003). Six- and 9-month-old English-learning infants show similar encoding effects for the prosodic structure of phrases (Soderstrom, Kemler Nelson, & Jusczyk, 2005).

There are several possible prosodic cues that can play a role in infants' perception of prosodic structure cohesiveness (e.g., pauses, lengthening of vowels before clause boundaries, intonation). English-learning infants appear to use multiple cues fairly equally at 4 months of age (Seidl & Cristiá, 2008) but then rely mainly on pitch cues by 6 months of age (Seidl, 2007). The cues that infants rely on also seem to be language dependent (Johnson & Seidl, 2008). Taken together, the studies of infants' sensitivity to prosodic structure suggest that infants may begin parsing speech into prosodic units at a very young age using multiple cues and then eventually learn to rely on particular cues. Statistical and associative learning are involved such that infants learn via statistical learning which prosodic boundaries cues co-occur most often with other prosodic boundary cues and then associate those cues with clausal and phrasal boundaries.

Language discrimination studies provide additional evidence of infants' developing sensitivity to prosodic information. Recall that newborns are able to discriminate languages that differ rhythmically (Nazzi et al., 1998). Nazzi, Jusczyk, and Johnson (2000) found that by 5 months of age, English-learning infants can discriminate their native language (e.g., American English) from languages (e.g., Dutch) and dialects (e.g., British English) within their same rhythmic class but cannot discriminate two foreign languages (e.g., Dutch and German) from the same rhythmic class. The investigators concluded that because 5-month-olds do not show language discrimination based on segmental information in previous work (Jusczyk, Friederici, et al., 1993), their discrimination was most likely due to an increased sensitivity to prosodic information that allowed the infants to detect subtle differences in rhythmic properties between the languages and dialects.

The above findings suggest that infants develop increasing familiarity with the prosodic properties of speech, including prosodic cues to linguistic units. This development suggests that there are learning mechanisms that transform infants from having a universal sensitivity to prosodic information to having a perceptual system tuned to the prosodic properties of the ambient language. First, recognition memory is required to identify units of speech as having rhythmic structure consistent with being clauses or phrases. Second, statistical learning is required to learn that certain rhythmic units (e.g., clauses) tend to co-occur with pauses while other rhythmic units (e.g., a sequence of words across a clause boundary) are not associated with a pause.

Word-Level Prosody

The research reviewed so far has informed us about infants' sensitivity to the organization of large prosodic units and intonational patterns, which may be an important first step in children's acquisition of syntax. But we have said very little so far about infants' sensitivity to smaller rhythmic units. Young infants are able to discriminate isolated words that differ in stress pattern (Jusczyk & Thompson, 1977) just as they are able to discriminate isolated words and syllables that differ by one phoneme (reviewed above). But to what extent are they sensitive to the rhythmic properties of words in the real world? In other words, do infants encode the rhythmic properties of words in the ambient language and build up implicit knowledge—via statistical learning—of the frequencies of different rhythmic patterns of words? Before we review the research that has addressed this question, we should first consider why sensitivity to the rhythmic properties of words might be important for speech comprehension. One important role for the rhythmic properties of words in speech comprehension is the role it plays in the process of segmenting words from the context of fluent speech (i.e., speech segmentation).

Speech segmentation is a major topic in speech science because natural speech does not contain obvious acoustic cues to word boundaries (Cole & Jakimik, 1980). We perceive word boundaries because we are able to segment continuous speech into words (listening to someone speak an unfamiliar language is an easy way to appreciate this fact). Although fluent speech does not reliably contain obvious word boundaries, people are, of course, able to segment fluent speech once they know a language. Knowing the words of a language is probably the most important factor for segmenting speech—recognizing a word informs the listener where the onset of the following word is. But there are also acoustic cues that become useful for segmentation as listeners gain implicit knowledge of the language, including the rhythmic properties of words.

One model of speech segmentation that emphasizes the importance of word rhythm is Anne Cutler's metrical segmentation strategy (MSS) model. The MSS asserts that in some languages, including English, listeners are attuned to strong[1] syllables as the primary acoustic cue for speech segmentation. In languages like English strong syllables can serve as cues for segmentation because of their distribution in the language. Cutler and Carter (1987) conducted a corpus investigation of English and found that approximately 90% of content words in English begin with a strong syllable. Thus, if listeners assumed that every strong syllable they heard marked the onset of a word, they would be correct most of the time. Cutler and colleagues tested this idea experimentally in a number of studies with adults and found that English speakers do indeed tend to perceive strong syllables as word onsets (Cutler & Butterfield, 1992; Cutler & Norris, 1988; McQueen, Norris, & Cutler, 1994).

As important a role that word rhythm may play in adults' ability to segment speech, it may play an even greater role in the development of speech segmentation during infancy. Unlike adults, infants do not have a developed lexicon to help them identify words in fluent speech. And while infants may learn some words from hearing them often in isolation (e.g., "hi," "daddy"), the vast majority of words are not uttered in isolation,

[1]The term "strong syllable" is nearly synonymous with the term "stressed syllable." A strong syllable is any syllable that has a fully realized (i.e., nonreduced) vowel. A stressed syllable is a syllable that is more perceptually salient than its neighboring syllables. To illustrate, take the spondee "mailman." Both syllables are strong because their vowels are fully realized. However, neither syllable has lexical stress because they have similar perceptual salience.

even to infants. An analysis of speech to an infant over a three-week period found that 90–95% of utterances contained more than one word (van de Weijer, 1998). Even when caregivers are instructed to teach words, they present the novel words in isolation only 20% of the time (Woodward & Aslin, 1990). Thus, being able to segment words from the context of fluent speech is an important skill for language acquisition.

Understanding the role word rhythm might play in infant speech segmentation returns us to the question of infants' sensitivity to the rhythmic properties of words. To address this question, Jusczyk, Cutler, and Redanz (1993) presented 6- and 9-month-old English-learning infants with lists of strong/weak words and lists of weak/strong words. They found that 9-month-olds oriented longer to the words that follow the predominant stress pattern of English (strong/weak), but 6-month-olds did not. Similarly, Echols, Crowhurst, and Childers (1997) presented infants with trisyllabic weak/strong/weak sequences that contained a pause either before or after the strong syllable. They found that 9-month-old English-learning infants preferred sequences with the pause before the strong syllable, which preserved the strong/weak structure. These findings suggest that over the course of at least six months of exposure to English, infants build up the implicit knowledge that strong/weak words are more common than weak/strong words. In other words, their statistical learning skills enable them to acquire sensitivity to the predominant stress pattern of words in their language.

Findings that English-learning infants become sensitive to the predominant stress pattern of words in their language led Peter Jusczyk and his colleagues to investigate the role of lexical stress in the development of speech segmentation skills. They did this by using a variant of the headturn preference procedure (HPP) to directly assess infants' ability to detect different types of familiarized words in the context of fluent speech. In the HPP infants are first familiarized with two words—one word per trial repeated up to 20 times. Then during a test phase, they are presented with four passages—two of which contain the familiarized words. Their attention to each passage is measured by the amount of time they orient to a light that is located in front of the loudspeaker presenting the passages and which blinks during the presentation of each passage. Seminal work using this methodology suggests that infants orient longer to the presentation of passages containing the familiarized words than to other passages when they are able to segment and recognize the familiarized words from the context of fluent speech (Jusczyk & Aslin, 1995). An alternative version of the HPP involves presenting two of the four passages during the familiarization period and then presenting the two familiarized words and two unfamiliar words during the test phase. These two variants of the HPP have been found to produce identical results (Jusczyk & Aslin, 1995; Jusczyk, Houston, & Newsome, 1999).

Using the HPP, Jusczyk, Houston, and Newsome (1999) assessed 7.5- and 10.5-month-old English-learning infants' ability to segment strong/weak and weak/strong words from fluent speech. They found that 7.5-month-olds were able to segment strong/weak but not weak/strong words from fluent speech. Instead, infants showed evidence of segmenting only the strong syllable of weak/strong words. When 7.5-month-olds were familiarized with just the strong syllable of weak/strong words (e.g., tar from guitar) they oriented longer to passages containing the corresponding weak/strong whole words (e.g., guitar). However, when they were familiarized with the whole words (e.g., guitar), they did not orient longer to the passages containing the familiarized words.[2] In other words, *tar* matched better to what 7.5-month-olds heard in passages containing *guitar* than did *guitar*. By 10.5 months of age, however, English-learning infants show the opposite pattern of results.

Jusczyk, Houston, and Newsome (1999) interpreted the findings with the strong/weak and weak/strong words as follows: English-learning infants begin segmenting words from fluent speech using a type of metrical segmentation strategy: They assume that strong syllables mark word onsets; when a strong syllable is followed by the same weak syllable consistently—as is the case when a strong/weak word occurs often in a passage—then infants will treat the strong/weak word as a cohesive unit. To test this interpretation, they created new passages for the weak/strong words in which each weak/strong word was consistently followed by the same function word (e.g., guitar is). Infants were presented with two of the passages and then tested on either the strong syllables of the target words or with strong/weak nonwords formed from the strong syllable of the weak/strong words and the following function word (e.g., tar-is). Unlike the previous experiment with weak/strong words, 7.5-month-olds did not demonstrate segmentation of the strong syllable from the weak/strong words in the passages. Instead, they demonstrated recognition of

[2]For both conditions, identical results were found when infants were tested with the passages-first variant of the HPP.

the strong/weak pseudowords (e.g., taris). These findings suggest that 7.5-month-old English-learning infants use a strong version of the MSS to segment words from fluent speech such that they segment strong/weak units even when they cross word boundaries.

Using a strong version of the MSS allows English-learning infants to segment most words from fluent speech correctly. However, as seen in the findings of Jusczyk et al. (1999), a strong version of MSS results in mis-segmenting words that do not follow the predominant stress pattern of English. Thus, if infants use a strong version of the MSS, they must eventually adopt a less strong version and incorporate other information in their segmentation strategy. To investigate the use of this strong version of the MSS in older infants, Jusczyk et al. (1999) also tested 10.5-month-olds with the passages in which the weak/strong target words were always followed by the same function word. Unlike 7.5-month-olds, 10.5-month-olds did not demonstrate recognition when presented with the pseudowords (e.g., taris). However, they did demonstrate recognition when presented with the weak/strong words (e.g., guitar) during testing, suggesting that 10.5-month-old English-learning infants correctly segment weak/strong words from fluent speech even when they are consistently followed by the same function word.

Taken together, the above findings and others (Houston, Jusczyk, Kuijpers, Coolen, & Cutler, 2000; Houston, Santelmann, & Jusczyk, 2004; Johnson & Jusczyk, 2001) suggest that infants use statistical learning to infer the rhythmic structure of words in their language, and then that learning influences their processing of fluent speech. The initial segmentation strategy that develops from acquiring knowledge about the rhythmic properties of words does not always result in correct segmentation, so infants must acquire knowledge about other segmentation cues. We discuss what some of those other cues are next.

Meaningful Segmental Information

We learned earlier that young infants demonstrate discrimination of phonetic segments even when the differences between those segments do not differentiate words in their language. We also learned that toward the end of the first year of life, infants lose the ability to discriminate some non-native contrasts. On the surface, it appears that infants become less sensitive to segmental information. However, investigations of infants' sensitivity to segmental properties in their language suggest that infants learn much about how phonetic segments are distributed and organized in the ambient language during the first year of life. This statistical learning about the distributional properties of segmental information, in turn, shapes more advanced speech perception processes, such as infants' perception of word boundaries in fluent speech (i.e., speech segmentation). The following sections review investigations of older infants' sensitivity to several types of segmental information and what role these acquired sensitivities play in segmenting words from fluent speech.

Phonotactic Probabilities

The term "phonotactics" refers to the ordering of segments in languages. Languages differ greatly with respect to what clusters of phonemes are permissible in various positions within and between words and syllables. In English syllables, for example, each consonant before a vowel must be more sonorous than the previous segment and less sonorous than the subsequent segment ([s] is an exception). Thus the word "plan" is possible but the word "lpan" is not in English. Moreover, the word "pkan" is not possible because [p] and [k] have equal sonority. In other languages, such as Polish, syllable-initial consonant clusters can contain two voiceless stops in a row. In addition to phonotactic rules, there are phonotactic probabilities. Phonotactic probabilities refer to the occurrence of segment pairs within and between words and syllables. For example, the pair [s]-[d] occurs more often between words than within words in English whereas the pair [s]-[t] occurs more often within words than between words. Sensitivity to these kinds of phonotactic probabilities can provide information about likely word boundaries in fluent speech.

Infants appear to become sensitive to phonotactic rules and probabilities at around the same time they show sensitivity to the predominant stress pattern of words in their language. Jusczyk, Friederici, Wessels, Svenkerud, and Jusczyk (1993) tested English-learning and Dutch-learning infants with lists of words that were either phonotactically legal in English and not in Dutch or vice versa. Dutch- and English-learning 9-month-olds both oriented longer to lists of words that were legal in their native language. Six-month-olds showed no preferences. Similarly, Friederici and Wessels (1993) found that Dutch-learning 9-month-olds but not 6-month-olds attended longer to nonwords with phonotactically permissible word onsets and offsets than to nonwords with phonotactically impermissible onsets and offsets, even though the impermissible onsets were permissible as offsets and vice versa. Jusczyk, Luce, and Charles-Luce (1994) found that 9-month-old English-learning infants attended longer to lists of words with higher phonotactic probabilities than to lists of words with relatively lower phonotactic probabilities even though none of the words had any sequences that were phonotactically impermissible.

Like sensitivity to lexical stress, sensitivity to phonotactic probabilities appears to play a role in infants' segmentation abilities. Mattys and Jusczyk (2001) used the HPP to investigate 9-month-old English-learning infants' segmentation of words from fluent speech that had either easy or difficult phonotactic boundary information. Similar to Jusczyk, Houston, and Newsome (1999), infants were familiarized with two passages, each containing several instances of a target word. In this experiment, however, they manipulated the words that surrounded the target words such that they provided either good or poor phonotactic boundary information. Infants were able to segment the words from fluent speech only when the phonotactic boundary information for the target words was good at either the onset, offset or both. These findings suggest that by 9 months of age, English-learning infants acquire knowledge about the frequency of occurrence of phoneme sequences within words and can use this information to segment words from fluent speech.

Subphonemic Cues

So far, we have reviewed speech cues infants use for segmenting words from fluent speech that are at the clause, phrase, syllable, and phoneme levels of acoustic-phonetic information. As discussed earlier, infants appear to be sensitive not only to these levels of acoustic-phonetic information; they are also sensitive to subphonemic information in speech (Hohne & Jusczyk, 1994; McMurray & Aslin, 2005). It is possible that subphonemic information may play a role in infant speech segmentation. Using the HPP, Jusczyk, Hohne, and Bauman (1999) investigated English-learning infants' sensitivity to allophonic information as a cue to word segmentation. Infants were familiarized with two-syllable items (e.g., nitrates) and tested for recognition of those sequences in fluent speech. Jusczyk et al. found that 10.5-month-olds but not 9-month-olds listened longer to the passages containing the exact match (e.g., nitrates) than to passages containing an allophonic variant (night rates) of the familiarized items, suggesting that only the older infants relied on allophonic information to segment words from fluent speech.

Subsequent investigations have provided additional evidence that infants are sensitive to subphonemic cues to word segmentation. By 8 months of age, English-learning infants' segmentation of three-word sequences is affected by whether they are produced as single words (e.g., catalog) versus three-word phrases (e.g., cat a log; Johnson, 2003; Johnson & Jusczyk, 2001). By 12 months of age, English-learning infants demonstrate sensitivity to subtle acoustic-phonetic word boundary information (e.g., [toga][lore] versus [toe][galore]) when segmenting strong/weak sequences (e.g., toga) from fluent speech (Johnson, 2008). Taken together, these studies suggest that a variety of subphonemic cues play a role in infants' speech segmentation, especially by the end of their first year of life.

Infants' sensitivity to segmental information as cues to word segmentation is acquired as a result of their experience to language and their developing learning mechanisms. Most of the segmental cues to segmentation discussed in this section are not universal across languages. So for these cues to be useful for segmentation, statistical and associative learning is necessary in order to learn which phoneme sequences and subphonemic variants are associated with word boundaries and which are not. However, in order to learn which segmental cues are associated with word boundaries, infants must be able to segment at least some words from fluent speech. As a solution to this apparent chicken-and-egg problem, Jusczyk (1997, 2002) posited that English-learning infants use a divide-and-conquer strategy: they first segment strong/weak units from fluent speech and then analyze the strong/weak units to discover other segmentation cues. This is a plausible strategy for English-learning infants to use because of the rhythmic properties of English words. Recent evidence suggests that in languages with different rhythmic properties, infants adopt other segmentation strategies (Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006).

Nonphonetically Based Segmentation Cues

Most work on infant speech segmentation has focused on the role that acoustic-phonetic properties play. However, there are other types of information infants can exploit to segment words from fluent speech. Earlier we reviewed work by Saffran et al. (1996), which found that statistical learning skills enable 8-month-olds to compute transitional probabilities of syllables. For example, if an infant notices that occurrences of the syllable [ma] are usually followed by the syllable [mi] and that what precedes [ma] and what follows [mi] is highly variable, this statistical information may contribute to helping the infant segment [ma]-[mi] as a cohesive unit from fluent speech. And then if an infant is able to recognize a familiar word like "mommy" in an utterance, then that word can serve as a wedge to help segment the surrounding words. A recent investigation found just that. Bortfeld, Morgan, Golinkoff, and Rathbun (2005) tested 6-month-olds' segmentation of words from fluent speech that were preceded by words they already knew (e.g., "mommy"). In contrast to earlier studies showing that 6-month-olds could not segment words from fluent speech (Jusczyk & Aslin, 1995), Bortfeld et al. (2005) found that 6-month-olds could segment words from fluent speech, but only when preceded by a word they already knew. These findings suggest that word recognition plays an important role in segmenting novel words from fluent speech.

Some Afterthoughts

The field of infant speech perception has grown immensely over the last forty years, and this review of the work is, as a consequence, very incomplete. Many of the most important studies in the field were left out. More thorough reviews of the field can be found elsewhere (Jusczyk, 1997; Saffran, Werker, & Werner, 2006). The purpose of this chapter is to provide some understanding of how our hearing instrument becomes tuned to process speech. We first reviewed some of what is known about infants' early speech perception abilities and then described some of the developments in speech perception during the second half of the first year of life that put infants in a position to segment words from fluent speech. Finally, I described some general learning mechanisms that play a role in the development of speech perception skills.

Gaining a better understanding of how cognitive mechanisms affect speech perception development in typically developing normal-hearing infants contributes to our general understanding of language development. This knowledge also may have clinical implications for infants and children with congenital hearing loss. With impoverished auditory input, general cognitive skills may be particularly important for hearing-impaired infants' ability to achieve successful language outcomes. Future work exploring the links between specific cognitive and speech perception skills in both normal-hearing and hearing-impaired infants could provide valuable information to clinicians, especially if methods of improving cognitive skills in infants can be developed. Moreover, comparing normal-hearing and hearing-impaired infants' speech perception skills can provide insight into the effects of early auditory experience on the development of early speech perception and language skills (Horn, Houston, & Miyamoto, 2007; Houston, Pisoni, Kirk, Ying, & Miyamoto, 2003; Houston, Ying, Pisoni, & Kirk, 2003).

Acknowledgements. Preparation of this chapter was facilitated by a Research Grant from NIDCD (DC006235) and support from the Philip F. Holton Fund. The author would like to thank Carissa Shafto for helpful comments on a previous version of this chapter.

References

Behme, C., & Deacon, S. H. (2008). Language learning in infancy: Does the empirical evidence support a domain specific language acquisition device? *Philosophical Psychology*, *21*(5), 641–671.

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

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of the perceptual re-organization for speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 345–360.

Birnholz, J. C., & Benacerraf, B. B. (1983). The development of human fetal hearing. *Science*, *222*, 516–518.

Bloom, L., & Tinker, E. (2001). The intentionality model and language acquisition: Engagement, effort, and the essential tension in development. *Monographs of the Society for Research in Child Development*, *66*(4), 1–91.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *15*(4), 298–304.

Chomsky, N. (1968). *Language and mind.* New York, NY: Harcourt Brace.

Chomsky, N. (1975). *Reflections on language.* New York, NY: Pantheon Books.

Clohessy, A. B., Posner, M. I., & Rothbart, M. K. (2001). Development of the functional visual field. *Acta Psychologica*, *106*(1–2), 51–68.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–63). Hillsdale, NJ: Erlbaum.

Colombo, J., Shaddy, D. J., Richman, W. A., Maikranz, J. M., & Blaga, O. M. (2004). The developmental course of habituation in infancy and preschool outcome. *Infancy*, *5*(1), 1–38.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584–1595.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*(2), 218–236.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2*, 133–142.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113–121.

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

DeCasper, A. J., Lecanuet, J. P., Busnel, M. C., Granier-Deferre, C., & Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. *Infant Behavior and Development*, *17*, 159–159.

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

Echols, C. H., Crowhurst, M. J., & Childers, J. (1997). Perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, *36*, 202–225.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, *16*, 513–521.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception and Psychophysics*, *18*, 341–347.

Eimas, P. D., & Miller, J. L. (1980a). Contextual effects in infant speech perception. *Science*, *209*, 1140–1141.

Eimas, P. D., & Miller, J. L. (1980b). Discrimination of the information for manner of articulation. *Infant Behavior and Development*, *3*, 367–375.

Eimas, P. D., Miller, J. L., & Jusczyk, P. W. (1987). On infant speech perception and the acquisition of language. In H. Stevan (Ed.), *Categorical perception: The groundwork of cognition* (pp. 161–95). New York, NY: Cambridge University Press.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(968), 303–306.

Fagan, J. F., & McGrath, S. K. (1981). Infant recognition memory and later intelligence. *Intelligence, 5,* 121–130.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*, 181–195.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message. *Child Development*, *60*, 1497–1510.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, *64*(3), 657–674.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for Motherese speech. *Infant Behavior and Development*, *10*, 279–293.

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception and Psychophysics*, *54*, 287–295.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*, 171–189.

Haith, M. M., Hazan, C., & Goodman, G. S. (1988). Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Development*, *59*, 467–479.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*(3), 269–286.

Hohne, E. A., & Jusczyk, P. W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception and Psychophysics*, *56*(6), 613–623.

Horn, D. L., Houston, D. M., & Miyamoto, R. T. (2007). Speech discrimination skills in deaf infants before and after cochlear implantation. *Audiological Medicine, 5*, 232–241.

Houston, D. M. (1999). *The role of talker variability in infant word representations.* Doctoral dissertation. Johns Hopkins University, Baltimore, MD.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582.

Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(6), 1143–1154.

Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review*, *7*, 504–509.

Houston, D. M., Pisoni, D. B., Kirk, K. I., Ying, E., & Miyamoto, R. T. (2003). Speech perception skills of deaf infants following cochlear implantation: A first report. *International Journal of Pediatric Otorhinolaryngology*, *67*, 479–495.

Houston, D. M., Santelmann, L., & Jusczyk, P. W. (2004). English-Learning infants' segmentation of trisyllabic words from fluent speech. *Language and Cognitive Processes*, *19*(1), 97–136.

Houston, D. M., Ying, E., Pisoni, D. B., & Kirk, K. I. (2003). Development of pre word-learning skills in infants with cochlear implants. *Volta Review*, *103 (Monograph* 4), 303–326.

Johnson, E. K. (2003). *Word segmentation during infancy: The role of subphonemic cues to word boundaries*. Doctoral dissertation, The Johns Hopkins University, Baltimore, MD.

Johnson, E. K. (2008). Infants' use of prosodically conditioned acoustic-phonetic cues to extract words from speech. *Journal of Acoustical Society of America*, *123*(6), EL144-148.

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

Johnson, E. K., & Seidl, A. (2008). Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy*, *13*(5), 440–455.

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

Jusczyk, P. W. (2002). How infants adapt speech-processing capacities to native-language structure. *Current Directions in Psychological Science*, *11*(1), 18.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of english words. *Child Development*, *64*(3), 675–687.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*, 402–420.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics*, *61*, 1465–1476.

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

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*(5), 630–645.

Jusczyk, P. W., & Thompson, E. (1977). Perception of a phonetic contrast in multisyllabic utterances by 2- month-old infants. *Perception and Psychophysics*, *23*(2), 105–109.

Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development*, *72*(2), 352–369.

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development*, *78*(5), 1559–1571.

Kisilevsky, B. S., Hains, S. M., Brown, C. A., Lee, C. T., Cowperthwaite, B., Stutzman, S. S., . . . Wang, Z. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behavior and Development*, *32*(1), 59–71.

Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*(3), 220–224.

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

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–274). Dordrecht: Kluwer.

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

Lacerda, F. (1993). Sonority contrasts dominate young infants' vowel perception. In *PERILUS XVII* (pp. 55–63). Stockholm University, Sweden.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, *20*, 215–225.

Lecanuet, J. P., Gautheron, B., Locatelli, A., Schaal, B., Jacquet, A. Y., & Busnel, M. C. (1998). What sounds reach fetuses: Biological and nonbiological modeling of the transmission of pure tones. *Developmental Psychobiology*, *33*(3), 203–219.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. L. (1961). The discrimination of relative-onset time of the components of certain speech and non-speech patterns. *Journal of Experimental Psychology*, *61*, 379–388.

Locke, J. L. (1993). *The child's path to spoken language.* Cambridge, MA: Harvard University Press.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91–121.

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

McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, *95*(2), B15–B26.

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 621–638.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178.

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

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756–766.

Nazzi, T., Iakimova, Bertoncini, J., Frédonie, & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, *54*, 283–299.

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

Nazzi, T., Kemler Nelson, D. G. K., Jusczyk, P. W., & Jusczyk, A. M. (2000). Six-month-olds detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy*, *1*(1), 123–147.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, *13*, 253–260.

Polka, L., & Bohn, O. S. (1996). A cross-language comparison of vowel perception in English-learning and German learning infants. *Journal of the Acoustical Society of America*, *100*(1), 577–592.

Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221–231.

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/ - /ð/ perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America*, *109*(5), 2190–2201.

Rakison, D. H., & Poulin-Dubois, D. (2002). You go this way and I'll go that way: Developmental changes in infants' detection of correlations among static and dynamic features in motion events. *Child Development*, *73*(3), 682–699.

Ramus. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition*, *2*, 85–115.

Rose, S. A., & Feldman, J. F. (1997). Memory and speed: Their role in the relation of infant information processing to later IQ. *Child Development*, *68*(4), 630–641.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2001). Visual short-term memory in the first year of life: Capacity and recency effects. *Developmental Psychology*, *37*(4), 539–549.

Rose, S. A., Feldman, J. F., & Wallace, I. F. (1992). Infant information processing in relation to six-year cognitive outcomes. *Child Development*, *63*, 1126–1141.

Rose, S. A., Feldman, J. F., Wallace, I. F., & McCarton, C. (1991). Information processing at 1 year: Relation to birth status and developmental outcome during the first 5 years. *Developmental Psychology*, *27*(5), 723–737.

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

Saffran, J. R., Werker, J., & Werner, L. (2006). The infant's auditory world: Hearing, speech and the beginnings of language. In R. Siegler & D. Kuhn (Eds.), *Handbook of child development.* (pp. 58–108). New York, NY: Wiley.

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, *57*(1), 24–48.

Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, *11*(4), 596–606.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants listening preferences: Baby talk or happy talk? *Infancy*, *3*(3), 365–394.

Soderstrom, M., Nelson, D. G. K., & Jusczyk, P. W. (2005). Six-month-olds recognize clauses embedded in different passages of fluent speech. *Infant Behavior and Development*, *28*(1), 87–94.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, *49*(2), 249–267.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M. G., & Ohman, S. E. G. (1969). Cross-language study of vowel perception. *Language and Speech*, *12*, 1–23.

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-th/ in monolingual and bilingual acquisition of English. *Cognition*, *100*(2), 369–388.

Swoboda, P., Morse, P. A., & Leavitt, L. A. (1976). Continuous vowel discrimination in normal and at-risk infants. *Child Development*, *47*, 459–465.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, *47*, 466–472.

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *Journal of the Acoustical Society of America*, *120*(4), 2285–2294.

Tsushima, T., Takizawa, O., Sasaki, M., Siraki, S., Nishi, K., Kohno, M., . . . Best, C. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6–12 months: Language specific developmental changes in speech perception abilities. In *International conference on spoken language processing.* Yokohama, Japan: The Acoustical Society of Japan.

van de Weijer, J. (1998). *Language input for word discovery.* Nijmegen, The Netherlands: Kluwer.

Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology*, *33*(3), 514–535.

Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice. *Developmental Psychology*, *22*(3), 373–377.

Wentworth, N., Haith, M. M., & Hood, R. (2002). Spatiotemporal regularity and interevent contingencies as information for infants' visual expectations. *Infancy*, *3*(3), 303–321.

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

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*, *43*, 230–246.

Werker, J. F., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 285–311). Timonium, MD: York Press.

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

Woodward, J. Z., & Aslin, R. N. (1990). *Segmentation cues in maternal speech to infants*. Paper presented at the 7th biennial meeting of the international conference on infant studies. Montreal, Quebec, Canada.

Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development*, *57*(3), 803–815.