Derek M. Houston and Andrea Warner-Czyz

2 Speech perception and auditory development in infants with and without hearing loss

1 Introduction

Despite the intricate link between the development of auditory skills and refinement of speech perception abilities, there have been few attempts to review these two perspectives simultaneously (although see Saffran, Werker, and Werner, 2006). This chapter examines the acquisition of infant hearing capacities from both perspectives, including implications for both infants with normal hearing (NH) and infants with hearing loss (HL). To understand how HL affects speech perception, we first must understand the complex relationship between auditory development and early speech perception.

Little is known about how auditory development and speech perception development interact. In some ways, they seem to develop quite differently. For example, hearing newborns have reduced primary auditory abilities (e.g., frequency, intensity, and localization) relative to older children and adults (Werner, 1996). The enhancement of auditory skills occurs in several dimensions, as described in the subsequent sections. In contrast, several reports show younger infants demonstrate the ability to discriminate phoneme contrasts that older infants and adults seemly cannot discriminate (e.g., Best, McRoberts, and Sithole, 1988; Werker and Tees, 1984). The goal of this chapter is to address that apparent discrepancy.

The chapter begins by reviewing auditory development, speech perception, and word recognition in NH infants. Next, we review current work assessing speech perception in infants with HL using hearing aids (HA) and cochlear implants (CI). Finally, we discuss future directions for better understanding speech perception in infants with HL.

2 Auditory development

2.1 Intensity

Auditory response to sound occurs along a continuum from sound awareness to comprehension of sound. This chapter will focus on the earliest developing auditory skills: *Detection*, the awareness of sound versus silence, and *discrimination*, the recognition of differences in stimuli contrasts.

Derek M. Houston, The Ohio State University Wexner College of Medicine
Andrea Warner-Czyz, The University of Texas at Dallas

DOI 10.1515/9781614514909-003

Infants exhibit elevated responses thresholds to auditory stimuli relative to adults (Northern and Downs, 2002; Werner and Gillenwater, 1990; Werner and VandenBos, 1993). These elevated responses thresholds may not represent the infant's true audiometric thresholds, but the lowest intensity at which the infant responds (i.e., *minimum response level*) (Matkin, 1977). Behaviorally, the minimal audible intensity for a hearing newborn to respond to sound is 40–50 dB poorer than adult thresholds, which converge at 0 dB Hearing Level (Eisele, Berry, and Shriner, 1975; Werner and Bargones, 1992; Werner and Gillenwater, 1990; Werner and VandenBos, 1993). Detection of sound improves to within 30–40 dB of adult thresholds at 1 month and within 15–30 dB of adult thresholds at 3 months (Olsho, Koch, Carter, Halpin, and Spetner, 1988; Tharpe and Ashmead, 2001; Trehub, Schneider, Thorpe, and Judge, 1991; Werner and Bargones, 1992; Werner and Gillenwater, 1990; Werner and Mancl, 1993). However, sound detection varies by frequency such that high frequencies have elevated response thresholds through the first three months of life (Olsho et al., 1988; Tharpe and Ashmead, 2001; Trehub et al., 1991; Werner and Bargones, 1992; Werner and Gillenwater, 1990). Minimum response levels for 6-month-old infants listening to pure-tone stimuli exceed adult thresholds by 10–15 dB, with similar response levels in the high frequencies (Spetner and Olsho, 1990; Tharpe and Ashmead, 2001; Werner and Bargones, 1992). Pure-tone thresholds more closely approximate adult thresholds between 8 and 12 months of age and, by 24 months, infant thresholds mirror those of adults across frequencies (Werner and Bargones, 1992; Parry, Hacking, Bamford, Day, and Parry, 2003). Electrophysiological measures confirm behavioral findings that auditory immaturities resolve as a function of age. Werner and colleagues (1993) report that 3- and 6-month-old infants exhibit adult-like auditory brainstem response thresholds at 1,000, 4,000, and 8,000 Hz. Other studies confirm mature electrophysiological responses by three years of age (Kaga and Tanaka, 1980; Klein, 1984; Schneider, Trehub, and Bull, 1979; Salamy and McKean, 1976; Teas, Klein, and Kramer, 1982).

Infants need not only to detect sound, but also to discriminate sound based on intensity. Detection, which forms the lower level of the dynamic range, affects sound awareness. Intensity discrimination underlies the number of discriminable steps from the threshold of detection to the threshold of pain, and influences loudness growth. The terms *difference limen*, *just perceptible difference*, and *just noticeable difference* describe the smallest detectable change in a parameter such as intensity. Difference limens depend on several factors including signal intensity, duration, and type (e.g., pure tone versus narrowband noise). For example, the difference limen improves (i.e., gets smaller) with increased stimulus intensity levels, longer stimuli durations, and broadband versus narrowband stimulus frequency (Jesteadt and Bilger, 1974; Jesteadt, Wier, and Green, 1977; Steinhardt, 1936).

Infants exhibit immature intensity resolution relative to adults, who have an intensity difference limen of 0.5–2 dB (Sinnott and Aslin, 1985). Five- and 6-month-old infants can discriminate a 10 dB change in a 500 Hz tone (Moffitt, 1973). Slightly older infants (7–9 months) need a comparable intensity difference of 3–12 dB to perceive intensity change of a 1,000 Hz tone (Sinnott and Aslin, 1985). Intensity resolution improves throughout childhood, approximating adult norms by around age 12 years (Maxon and Hochberg, 1982). More recent electrophysiological studies using mismatch negativity (MMN) and magnetic mismatch responses (MMR) confirm statistically significant responses to intensity changes in the acoustic signal, though the direction of the deflection does not always match that of adults (Partanen et al., 2013; Sambeth et al., 2009).

In summary, both behavioral and electrophysiological response thresholds for infants require higher intensities, particularly in the high frequencies, but age-related changes asymptote around 6 months of age and reach adult levels at all frequencies by 24 to 36 months of age. Infants improve their ability to discriminate differences in intensity as a function of age, but do not reach adult performance levels until adolescence.

2.2 Frequency

Like intensity, infants' ability to detect and discriminate stimulus frequency improves over time, particularly during the first year of life (Abdala and Folsom, 1995; Olsho, 1984; Olsho, Koch, and Halpin, 1987). This section details development of frequency detection, specificity, and resolution.

Infants show less sensitivity to high frequency stimuli than adults (Tharpe and Ashmead, 2001; Werner and Gillenwater, 1990). Werner and Gillenwater (1990) report that very young infants (2–5 weeks) show behavioral thresholds elevated by 42–47 dB between 500 and 4,000 Hz relative to adults, with larger group differences at higher frequencies. High frequency thresholds improve to within 35 dB of adult norms at age 1 month, within 19–30 dB at 3 months, and within 15 dB at 6 months (Tharpe and Ashmead, 2001; Werner and Gillenwater, 1990; Folsom and Wynne, 1987). Low frequency thresholds also improve over time: 3-month-olds' thresholds exceed adult thresholds by 23 dB at 500 Hz (Olsho et al., 1988).

Not only does frequency affect detection, but also specificity. Psychophysical tuning curves offer a way to measure how an individual resolves a low-intensity pure-tone test signal in the presence of competing noise, called a masker. The intensity of the masker is increased until the individual no longer can detect the test signal. Several factors, including the frequency composition of the masker relative to the test signal, affect the masker intensity required to eliminate signal detection. Maskers similar in frequency to the test signal likely engage similar neural activity as the test signal itself, thereby requiring less intensity to eradicate detection of the test signal; conversely, spectrally disparate maskers require more intensity to eliminate response to the test signal. A narrow psychophysical tuning curve reflects good resolution, or tuning, of frequency; and a broad psychophysical tuning curve indicates poor frequency resolution. Infants exhibit poorer frequency specificity relative to adults. Three-month-old infants have broader psychophysical tuning curves than adults at 500 and 1,000 Hz (Spetner Olsho, 1990). Six-month-old infants possess psychophysical tuning curve widths and critical bands similar to adults, suggesting comparable levels of frequency specificity at suprathreshold intensities (Schneider, Morrongiello, and Trehub, 1990; Olsho, 1985). Thus, frequency specificity may reach maturity by six months of age.

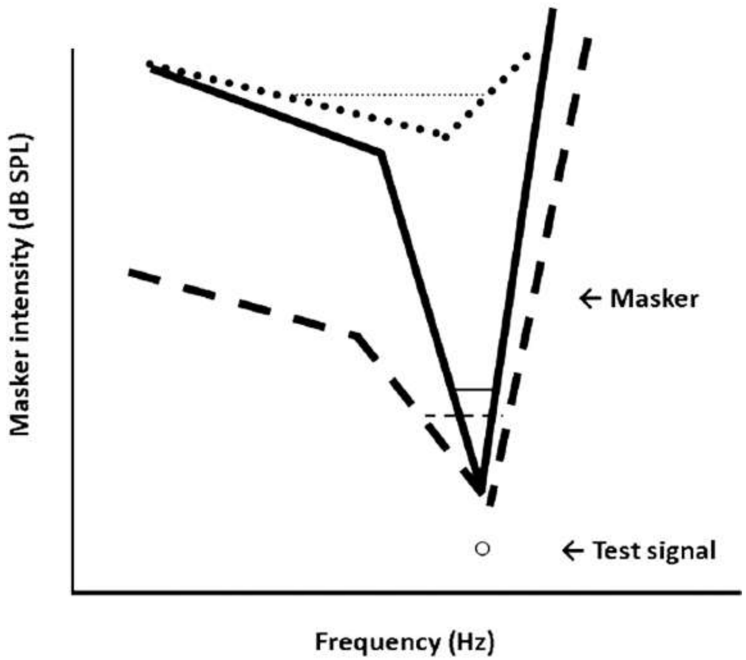

Figure 1: Schematic of psychophysical tuning curve
Figure 1 shows three psychophysical tuning curves plotted as a function of frequency on the x-axis and intensity on the y-axis. The small circle represents the frequency and intensity of a test signal. The three lines represent the intensity of the masker needed to eliminate detection of the test signal. The solid line depicts a normal psychophysical tuning curve that requires less intensity of the masker at frequencies similar to the frequency of the test signal and greater intensity of the masker at frequencies dissimilar to the frequency of the test signal. The width of the psychophysical tuning curve that leads to the tip specifies its resolution. A narrow tuning curve with a sharp point, as shown with the solid line, implies good frequency resolution. The broader tip of the tuning curve with the dashed line reflects poorer frequency resolution relative to the tuning curve with the solid line. The dotted line illustrates a psychophysical tuning curve with both an abnormal threshold, evident in the level of the masker required to eliminate perception of the test signal, and abnormal frequency resolution, evident in the width of the tuning curve

An individual's frequency resolution reflects attunement of their auditory system to specific frequencies, evidenced in the width of the psychophysical tuning curves.

One would expect frequency resolution to coincide with frequency discrimination such that individuals with better frequency specificity (i.e., narrower psychophysical tuning curves) can identify changes in frequency better than individuals with poorer frequency specificity. A mature human auditory system detects a difference limen equal to 1% change in frequency (Olsho, Schoon, Sakai, Turpin, and Sperduto, 1982). That is, adults can discriminate a 20 Hz change for a 200 Hz stimulus and a 200 Hz change for a 2,000 Hz stimulus. Better discrimination in the low versus high frequencies likely reflects attention to periodicity information in the signal rather than excitation information from the cochlea (Moore, 1973a; Moore, 1973b). Infants experience greater difficulty discriminating frequency compared to adults (Bargones, Werner, and Marean, 1995; Schneider, Trehub, Morrongiello, and Thorpe, 1989; Werner and Boike, 2001). Three-month-old infants require a frequency difference limen of 2–3% across the frequency range (Olsho et al., 1987). For example, 3-month-old infants require at least a 120 Hz change to discriminate a difference in a 4,000 Hz tone (Olsho et al., 1987). Five- to 12-month-old infants demonstrate frequency difference limens similar to 3-month-olds in the low and mid frequencies (250 Hz and 3,000 Hz), but comparable to adults in the high frequencies (4,000 and 8,000 Hz) (Olsho et al., 1982; Olsho et al., 1987; Sinnott and Aslin, 1985; Olsho, 1984).

Despite the fact that infants demonstrate better sensitivity to low versus high frequencies (Tharpe and Ashmead, 2001; Werner and Gillenwater, 1990; Olsho et al., 1988), young infants are able to discriminate both phonemes that differ primarily in low-frequency acoustic information (e.g., vowels; Kuhl, 1979; Trehub, 1973) and phonemes that differ primarily in high-frequency acoustic information (e.g., /s/ vs /ʃ/; (Eilers and Minifie, 1975) – at least under quiet laboratory conditions. Differences in fine-grained spectral resolution may not affect infant speech discrimination in quiet, but it may influence infant speech perception in adverse listening conditions such as background noise and perception of musical characteristics such as melody and timbre as shown in adults (Dorman, Loizou, Fitzke, and Tu, 1998; Fu, Shannon, and Wang, 1998; Won, Drennan, Kang, and Rubinstein, 2010).

Immaturities in frequency detection and discrimination likely relates to development of the middle ear system and higher order auditory pathways over the first few months of life. Differences in the efficiency and effectiveness of the middle ear system decrease its sensitivity to high frequencies, which match adult thresholds by 12 months of age (Keefe and Levi, 1996). Immaturity of the auditory brainstem and slower processing in higher order auditory pathways also may contribute to decreased sensitivity to high frequencies during infancy (Werner, Folsom, and Mancl, 1994).

2.3 Temporal

Temporal resolution underlies several aspects of auditory perception, including discrimination and recognition of speech, localization of sound, and listening in adverse conditions such as noise. Few studies have examined detection of temporal properties, instead focusing on the ability to discriminate differences in temporal aspects of sound.

Young infants respond reliably to gaps in stimuli, but need longer gap durations. For example, 3- and 6-month-olds require 40–60 ms longer gaps relative to adults at 500, 2,000, and 8,000 Hz (Werner, Marean, Halpin, Spetner, and Gillenwater, et al., 1992). Morrongiello and Trehub (1987) found similar results: 6-month-old infants detect gaps of 25 ms, 67% longer than gaps identified by five-year-old children (15 ms) and more than twice as long as gaps identified by adults (10 ms). Even though infants need longer gaps, they show a parallel pattern to adults with poorer gap detection in the low frequencies and better gap detection in the high frequencies (Werner et al., 1992). Three-month-old infants and adults perform similarly with a long temporal separation (200 ms), but differ with shorter temporal separations (5, 10, and 25 ms) (Marean and Werner, 1991). That is, shorter durations of temporal separation (5, 10, 25 ms) yielded a greater difference in masked versus unmasked thresholds for infants, but not adults.

Very young infants can discriminate voice onset time (VOT), the duration of time between the release of the stop consonant and the onset of vocal fold vibration (see Section 3.2.1. for details). Several researchers have shown that infants can discriminate voiced versus voiceless syllables (e.g., /ba/ vs. /pa/) within the first four months of life (Eimas, Siqueland, Jusczyk, and Vigorito, 1971; Eimas, 1974; Trehub and Rabinovitch, 1972). Infants' ability to perceive VOT cues emerges in spite of immature temporal resolution, as evidenced by measures such as gap detection and forward masking (Marean and Werner, 1991; Morrongiello and Trehub, 1987; Werner et al., 1992). Young infants can discriminate durational patterns, but their abilities do not quite match those of older children and adults. Progressive improvement throughout infancy suggests immaturity in temporal processing, but also may indicate immaturity in other realms such as attention and memory.

Infants with NH demonstrate immature auditory sensitivity across intensity, frequency, and temporal domains, but improve to more adult-like auditory capacities within the first 2 to 3 years of life. During the same period of life, infants develop sophisticated speech perception skills. It would be tempting to assume that improvements in auditory sensitivities will lead to increased perceptual sensitivity of speech. However, as we will see below (Section 3), the relationship between psychoacoustic abilities and speech perception abilities is much more complex than that.

3 Infant speech perception

At the same time that their auditory system is developing increased sensitivity to intensity, frequency, and temporal domains, infants develop perceptual skills to accomplish a seemingly simple but truly challenging task: To recognize acoustically different instances of the same word as phonologically equivalent, and, at the same time, differentiate words that may be acoustically very similar. The word *oil* produced by a man from Georgia in an angry mood is likely very different acoustically from the same word produced by a woman from New Hampshire in a happy mood, whereas two words like *peak* and *beak* produced by the same talker in the same mood are acoustically very similar. How humans recognize phonological differences and equivalences from the acoustic signal is far from straightforward.

This section reviews the current state of the field of infant speech perception as it relates to word recognition. Two speech perception processes form the foundation for recognizing words in natural speech: (1) identifying where words begin and end in continuous speech (i.e., *speech segmentation*) and (2) comparing candidate words segmented from continuous speech to stored representations of words in the mental lexicon (i.e., *discrimination/categorization*).

3.1 Speech segmentation

Most speech directed to infants occurs in the form of multiword utterances rather than isolated words (van de Weijer, 1998). Even when explicitly instructed to teach words to their infants, parents rarely provide words in isolation (Woodward and Aslin, 1990). Because no reliable acoustic cues to word boundaries exist in natural speech, infants segment words from fluent speech using a combination of several probabilistic cues informed by the ambient language (Cole and Jakimik, 1980).

3.1.1 Early-learned words

Young infants take advantage of existing knowledge to segment words from fluent speech. Six-month-olds use recognition of early-learned, highly familiar words (e.g., *Mommy*), which often occur in isolation, as boundary markers for candidate words (Bortfield, Morgan, Golinkoff, and Rathbun, 2005). This finding contrasts with earlier research showing that six-month-olds could not segment words from fluent speech without such markers (Jusczyk and Aslin, 1995).

3.1.2 Rhythmic properties

Although early-learned words may support segmentation by infants in some cases, infants consistently use cues related to the acoustic-phonetic organization of the ambient language to separate words from the speech stream. Infants show sensitivity to one suprasegmental aspect of speech, the rhythmic properties of words, at very young ages – even before birth. Fetuses at 33–41 weeks gestational age respond to a change from their ambient language to another language, likely reflecting sensitivity to rhythmic properties of speech transmitted *in utero* (Kisilevsky et al., 2009). Newborns can discriminate languages that differ in rhythmic structure even without *in utero* experience with either language (Nazzi, Bertoncini, and Mehler, 1998).

Sensitivity to rhythmic properties of speech influences speech segmentation between 6 and 12 months, at least for English-learning infants. Most bisyllabic words in English begin with a strong stressed syllable (e.g., *doctor*) versus a weak unstressed syllable (e.g., *guitar*). Jusczyk et al. (1999) found that 7.5-month-old English-learning infants could segment strong/weak but not weak/strong words from fluent speech, suggesting that English-learning infants treat strong syllables as beginnings of words and a strong-weak bisyllables as a cohesive unit. Infants learning other languages use other types of rhythmic information for word segmentation (Nazzi, Iakimova, Bertoncini, Fredonie, and Alcantara, 2006). Thus, experience drives awareness of rhythmic properties within words, which subsequently influences word segmentation.

3.1.3 Statistical learning

Noticing that one syllable consistently follows another syllable exemplifies "statistical learning" (Saffran, Aslin, and Newport, 1996). Learning the probability that syllables co-occur represents a powerful domain-general learning skill that plays a significant role in segmenting words from fluent speech.

3.1.4 Other acoustic properties

Infants refine their sensitivity to rhythmic cues and syllable co-occurrences to include segmental and subsegmental (e.g., coarticulatory) properties of their native language. Nine-month-old English-learning infants are sensitive to phonotactic properties such as the likelihood of consonant clusters occurring within or between word boundaries in their ambient language (Mattys, Jusczyk, Luce, and Morgan, 1999). Older infants (11 months) use allophonic and coarticulatory cues to identify word boundaries (Johnson, 2003; Jusczyk, Hohne, and Bauman, 1999). By the end of the first year of life, infants exhibit proficiency at segmenting words from fluent speech regardless of their rhythmic properties (Jusczyk, Houston, and Newsome, 1999).

3.2 Word recognition

To correctly recognize a word, an infant must discriminate it from other words and categorize it with other variants of the same word. As indicated above, this task goes beyond a simple determination of overall acoustic similarity to include identification of the acoustic-phonetic properties that afford discrimination of different words and categorization of the same word.

3.2.1 Word discrimination

The earliest infant speech perception work focused on which aspects of speech infants could discriminate (Eimas et al., 1971). Within the first weeks of life, infants can differentiate syllables based on single features such as consonant voicing, place of articulation, and manner of articulation (Eimas, 1974; Eimas, 1975; Eimas and Miller, 1980b; Eimas and Miller, 1980a; Eimas et al., 1971). Infant discrimination patterns often approximate those of adults. For example, both infants and adults demonstrate categorical perception of voicing contrasts when differences in voice onset time (VOT) cross a category boundary (e.g., 20 vs. 40 ms VOT) but not when VOT differences fall within the same category (e.g., 0 vs. 20 ms) (Eimas et al., 1971; Liberman, Harris, Kinney, and Lane, 1961).

Adult-like discrimination of speech contrasts by young infants challenge the notion that ambient language experience guides perception. Werker and Tees (1984) investigated English-learning infants' discrimination of a Hindi contrast, [da] vs. [ḍa] (retroflex "d", produced by pulling the tongue back from the teeth) which has no linguistic relevance in English. In contrast to Hindi-learning infants, who discriminate the native language contrast regardless of age, English-learning infants show age differences in discrimination. Younger infants (6–8 months) discriminate the non-native language contrast, but older infants (10–12 months) – like adults – do not.

Multiple studies confirmed this age-related pattern in discrimination of linguistically irrelevant language contrasts across multiple studies (Best et al., 1988; Kuhl, Williams, Lacerda, Stevens, and Lindblom, 1992; Trehub, 1976; Tsushima et al., 1994; Werker and Lalonde, 1988). This convergence of results led to widespread adoption of the *Universal Theory* of infant speech perception, which posits that infants emerge with the ability to discriminate all linguistically relevant contrasts of all the world's languages (Aslin and Pisoni, 1980). Exposure to the ambient language(s) causes infants to lose the ability to discriminate non-native, linguistically irrelevant contrasts but retain the ability to discriminate contrasts linguistically relevant to the ambient language(s) (e.g., Eimas, Miller, and Jusczyk, 1987; Werker and Pegg, 1992).

The fact that discrimination of non-native speech sounds becomes poorer seems to contrasts with the fact that psychoacoustic skills become more refined over the course of infancy. How can the auditory and speech perception systems develop in opposite directions? More recent work reviewed below suggests that while the auditory system becomes more acute over development, the complexity of the speech perception system exceeds that proposed by the Universal Theory.