Speech Perception☆

1998

Gerald W McRoberts, Haskins Laboratories, New Haven, CT, United States

© 2020 Elsevier Inc. All rights reserved.

Speech Perception

Speech perception refers to the ability to perceive linguistic structure in the acoustic speech signal. During the course of acquiring a native language, infants must discover several levels of language structure in the speech signal, including phonemes (speech sounds) which are the smallest units of speech. Although phonemes have no meaning in themselves, they are the building blocks of higher-level, meaningful linguistic units or structures, including morphemes, words, phrases and sentences. Each of the higher-level units are composed of units at the next lower level using rules that are specific to each language (i.e., phonology, morphology, grammar or syntax). Thus, sentences are made up of phrases, phrases are composed of words, and words are made up of morphemes. Each of the meaningful units are composed of one or more phonemes. In a very real sense, the ability to perceive differences between and categorize phonemes provides the underlying capacity for the discovery of the higher levels of language structure in the speech signal. In this way, infants' speech perception abilities play a fundamental role in language acquisition. Although infant speech perception has traditionally focused on discrimination and categorization at the phoneme level, research over the past two decades has shown that infants are also beginning to become sensitive to a variety of higher-level linguistic structures in speech. This chapter outlines the current state of knowledge about how infants begin to perceive linguistic structure in speech during the first year of life, and the methods used to study infant speech perception.

Why Speech Perception Is Difficult

Infants' discovery of language structure in speech is not a trivial task because phonemes lack acoustic invariance. That is, the acoustic properties of specific phonemes in fluent speech can vary dramatically based on several factors. The acoustic characteristics of speech sounds that listeners use in perception directly reflect the acoustic resonance properties of the vocal tract, which in turn are determined by moment to moment changes in the shape of the vocal tract during speech production. In addition, because the size and shape of speakers' vocal tracts vary, so do the specific acoustic properties of any specific phoneme. The rate at which speech is produced also introduces variations in the characteristics of speech sounds. Specifically, as the rate of speech increases, speakers' articulatory gestures (the movement patterns of articulators, such as the tongue tip) fail to reach the positions attained with slower rates of speech. This means that the vocal tract shape associated with a speech sound varies as a function of speech rate. Finally, although speech unfolds over time, speech sounds are not produced in a strictly serial manner. Rather, the production of speech sounds overlap in time due to a phenomenon referred to as coarticulation. One example of coarticulation can be seen when the word "two" is spoken. The vocal tract movements for the vowel /u/ in "two" include lip movement called rounding. However, it is common to see this lip movement occur throughout the word. Lip movement occurs during the /t/ even though it is not a normal part of the production of the that phoneme, nor is it necessary, since the word "two" can be produced without lip rounding during the /t/. One result of this coarticulation is that the sound of the /t/ is different when the lips are protruded compared to when they

☆*Change History:* February 2019. G. McRoberts has changed the entire section to this article and added reference citations.
This article is an update of G.W. McRoberts, Speech Perception, Editor(s): Marshall M. Haith, Janette B. Benson, Encyclopedia of Infant and Early Childhood Development, Academic Press, 2008, Pages 244–253.

are not. In general, because coarticulation is common in fluent speech production, the acoustic information that specifies any particular speech sound is highly context dependent. That is, the acoustic properties of speech sounds can depend significantly on the speech sounds that precede and follow it. Surprisingly, very young infants are sensitive to coarticulation effects (Fowler et al., 1990). Coarticulation, speaker, speaking rate and other sources of variation in the speech signal mean that there is no absolute acoustic signature (acoustic invariance) for any speech sound. In language acquisition, this lack of invariance in the acoustic specification of speech sounds is a complicating factor in speech perception. Additionally, it must also complicate the development of speech production, because infants' and young children's vocal tracts cannot physically produce many of the specific acoustic patterns they hear in adult speech.

Methodologies for Studying Infant Speech Perception

The limited behavioral repertoire of infants kept their ability to perceive speech unstudied until appropriate methods were developed beginning in the 1970s (Eimas et al., 1971). Unlike the case of vision, where observable behaviors, such as direction of eye gaze, are reliable indicators of perception, there is no overt behavior that indicates listening. Researchers interested in infants' auditory and speech perception capabilities had to develop methods that used behaviors infants had under their control as indirect measures of perception. For example, by coupling the presentation of speech (or other sounds) to behaviors that infants can control, such as sucking at a certain rate or pressure (e.g., Eimas et al., 1971) or fixating a visual target (Horowitz, 1975), these behaviors can be used as indices of infants' interest in the sounds. When infants look at a visual target more to listen to one sound playing longer than another, it is inferred that the greater looking is related to more interest in the sound. Thus, the duration of looking (or the amount of sucking) is used as an index of infant listening or attention to the sounds. Procedures that use contingencies between infant behavior and sound presentation are sometimes referred to as "infant-controlled" procedures because the infant controls the duration of sound presentation on each trial. There are other procedures used to study infants' speech perception that do not use contingencies, and thus are not infant-controlled. However, these procedures still require infants to produce an observable behavior, such as turning their head in a specific direction to either choose between two sounds, or to indicate they heard a change in the sound that was playing.

The focus of this chapter is on two aspects of infant speech perception: 1) infants' ability to discriminate between different speech sounds or categories of speech sounds; 2) infants' preference to listen to some forms of speech, or speech with specific types of structure, over others. Both sucking and gaze patterns have been used as behavioral indices in each of these approaches to infant speech perception. The next section describes several approaches to infant speech sound discrimination. A later section describes infant speech preference procedures and their uses in studying the development of speech perception.

Infant Speech Discrimination Procedures

Infants' ability to discriminate or categorize speech sounds has often been studied using procedures that involve habituation. These procedures are typically divided into two phases, habituation and test. Early studies of infant speech perception exploited infants' sucking reflex as a response measure in the High Amplitude Sucking (HAS) procedure (e.g., Eimas et al., 1971). In this procedure, infants suck on a non-nutritive nipple attached to a pressure transducer, which measured the amplitude of the sucking. Spontaneous sucking levels without sound presentation are measured during an initial period to establish each infant's baseline sucking amplitude. After the baseline is established, the habituation stimulus is presented contingent on the infants maintaining a sucking amplitude greater than the mean of the baseline period. In a major methodological advance, Best et al. (1988) introduced gaze as a response measure to the study of infant speech perception by adapting procedures previously used to study auditory discrimination (Horowitz, 1975) and perceptual constancy (Miller, 1983). In this case, infants are presented with a simple image, such as a checkerboard, as a visual target. When the infant fixates the target, the habituation stimulus is presented contingent on the infant maintaining fixation. When the infant breaks fixation for a period greater than 1 or 2 s, stimulus presentation stops and the trial ends (see Oakes, 2010 for a discussion of how habituation is instantiated in different infant testing situations).

For both sucking and gaze measures, a criterion is used to determine when infants have habituated. When sucking is the behavioral response, the criterion is typically a decline in the sucking rate by 20% for 2 consecutive minutes. When gaze is the behavioral response, the criterion is usually set at 50% of average looking on the first two trials, or the two longest trials. Habituation occurs when the behavioral response is below the criterion (e.g., 50% of the average of the first two trials) on two consecutive trials. Once the behavioral response indicates habituation, the habituation stimulus is changed to the test stimulus on the subsequent trial.

Consider an example in which a researcher wishes to test infants' ability to discriminate between two syllables differing in their initial consonant, such as [ba] and [pa], using the visual habituation procedure. During the habituation phase, infants would be presented with either a single token or multiple tokens of the habituation stimulus (e.g., [ba]) repeated at a short interval (e.g., 500 ms). When the infant fixates the visual target, the habituation stimulus is presented. Typically, a trial continues until the infant stops the target behavior for a specified duration (e.g., looks away for greater than 1 s) or the trial reaches a maximum time (e.g., 20 s). Because stimulus presentation is contingent on the infant maintaining fixation, the duration of the trial is taken as an index of their interest in the stimulus being presented. Over several trials the infant becomes more familiar with the stimulus and habituation occurs, resulting in less looking at the visual target. The looking time on each trial is compared to the predetermined habituation criterion. When the criterion is reached on consecutive trials, the test phase begins and the test stimulus (e.g., [pa]) is presented on the next trial. The number of trials in the test phase varies. Often only two or three trials are presented, but testing can also continue until the infants are habituated to criterion on the test stimulus.

Mean looking times for the last two habituation trials and the test trials are calculated for each group of infants. If infants noticed the difference between the habituation stimulus and the test stimulus, the expectation is that they will look (or suck) more to hear the novel (unhabituated) stimulus. An increase in looking during the test phase relative to the end of the habituation phase indicates response recovery, or dishabituation. To establish that discrimination did or did not occur, a comparison is made between the last two trials from the habituation phase and the first two trials from the test phase. Discrimination is inferred if infants look more to listen to a novel stimulus than they look to listen to the familiar stimulus. If their looking to the two stimuli is not statistically different, failure to discriminate is inferred. Because infants can exhibit some degree of spontaneous recovery after several short stimulus presentations, such as the final trials of the habituation phase, a no-change control group is often employed. Infants in the control group continue to hear the same stimulus after reaching habituation criterion. Thus, a second important comparison is between the test trials of infants in the test and control groups. To rule out the possibility that infants in the test group exhibited spontaneous recovery, their looking times in the test phase must be statistically greater than the no-change control group, as well as their own final habituation trials.

The Visually Reinforced Head Turn (VRHT) procedure differs from the HAS and infant gaze procedures in that it does not involve habituation and is not infant-controlled (Eilers et al., 1977; Kuhl, 1985). Rather, it requires infants to notice a sound change in a continuous stream of recurring syllables and look toward a visual reinforcer within a brief time window after the sound change. This procedure typically involves first training infants on the procedure, then testing them on the stimulus comparison of interest. During the training phase, a visual reinforcer, usually an animated toy in a smoked plexiglass box, is used to train infants to produce a head turn when a change occurs in a repeated background sound (e.g., from a high tone to a low tone). During training, the visual reinforcer is activated just prior to a sound change. Over the course of training, the interval between the sound change and activation of the reinforcer is reduced and finally reversed, so that activation occurs after the sound change. When the infant reliably anticipates activation of the reinforcer by turning toward the reinforcer after a sound change, but before the reinforcer is activated, the testing phase can begin with the stimuli of interest. During testing, the infant is distracted from the visual reinforcer by a research assistant displaying an interesting object. This reduces false positive responses, in which the infant looks to the reinforcer when no sound change has occurred.

Each of these procedures has both advantages and disadvantages. The habituation approach works well from the neonatal period into the second year of life. However, the sucking measure works best with infants up to 4-months of age, after which infants are prone to rejecting the nipple. The visual fixation method works well across a wide range of infant ages and is now the generally accepted method of choice. The Visually Reinforced Head Turn procedure requires adequate head, neck and postural control from the infant, and therefore does not work well with infants younger than about 4 months. In addition, some subjects fail to provide usable data because they do not reach criterion during the training phase. However, this procedure has the advantage of providing reliable data for individual infants, whereas the habituation procedures can only be used for group comparisons. Although it may be possible to modify habituation procedures to provide individual data, attempts to do so have not proven successful at this time.

Infants' Phonetic Discrimination

In order to understand important issues involved in the study of infant speech perception, it is helpful to have a bit of background on how speech sounds are made and classified (see Ladefoged, 1975 or Ladefoged and Johnson, 2015 for details of speech production). Speech sounds are produced by complex coordinated interactions among the components of the vocal tract. During speech production, air from the lungs induces the vocal folds to vibrate, producing a buzz-like sound source that is filtered by passing through several cavities in the vocal tract, including the pharyngeal, oral and nasal cavities. Speech production involves movement of the vocal tract articulators, including the tongue, velum and lips, which change the sizes and shapes of the cavities, altering their resonance properties. It is the resonance properties of these cavities that filter the sound source and account for most of the differences in the sounds used in speech.

Phonetics, the study of how speech is produced and perceived, generally distinguishes between two classes of phonemes, consonants and vowels. One basis for this distinction is related to differences in how these sounds are produced. Vowels are produced with a relatively open and unobstructed vocal tract, which allows air and sound to move freely through the vocal tract. Changes in the height and front-back position of the tongue, along with rounding or spreading of the lips, produce most of the differences in vowels in English. Other languages also distinguish vowels based on whether air flows only through the oral cavity, or through the nasal cavity, which results in a nasalized sound. Vocal fold vibration (voicing) occurs throughout vowels.

In contrast to vowels, consonants are produced by introducing a constriction of the airflow. Consonants can be classified by the type and location of the vocal tract constriction used in their production. Many types of constrictions, or manners of articulation, are used in languages around the world. The most common types involve a complete closure or one of several kinds of partial closure. Stop consonants (e.g., /p/ or /d/), are produced with a full closure of the vocal tract that results in a complete stoppage of air flow, whereas fricatives (e.g., /s/) involve a less-than-complete degree of constriction that results in a turbulent air flow. There are several other manners of articulation, some of which are not used in English. Two manners of articulation not used in English are click and ejective. Clicks are used in Zulu and some related languages in Southern Africa. Production of clicks involves a suction type closure and release (Ladefoged and Maddieson, 1996). Another common manner of articulation is called ejective. Ejectives are produced using a closure of the vocal tract in the mouth and also squeezing the vocal folds together. Air that is trapped between the vocal tract closure and vocal folds is compressed by moving the larynx upward. When the closure in the oral cavity is released, the built-up air pressure is released with a distinctive sound.

Each manner of articulation can be produced at various places within the vocal tract and can involve the tongue, lips, teeth, hard and soft palate, as well as other parts of the vocal tract. One common constriction location involves a closure at the lips and is referred to as bilabial. Examples of other common closures involve the tongue and any of several locations along the roof of the mouth, including the teeth (interdental), the ridge behind the teeth (alveolar), the hard palate (palatal) or the soft palate (velar).

Consonants are also classified as being either voiced or voiceless. Voiced means vocal fold vibration occurs during the constriction or constriction release. This contrasts with voiceless (or unvoiced) consonants, where vocal fold vibration begins some time after the constriction is released. The difference in when the vocal folds begin to vibrate is referred to as Voice-Onset-Time (VOT), or the time from when the constriction is released until the vocal folds begin to vibrate. In English, voiced consonants have a VOT from about 0 to 40 ms (or thousandths of a second), meaning the onset of vocal fold vibration can be simultaneous with, or up to 40 ms after, release of the closure or constriction (e.g., /b/ or /z/). In unvoiced consonants, vocal fold vibration begins from about 60 to 100 ms after the constriction is released (e.g., /p/ or /s/). Thus, the consonant inventory of a language is dependent on which combinations of articulatory features, including voicing, manner of articulation and places of articulation, are used. The phoneme inventory of languages differ substantially in the number of vowels and consonants that are used. Anyone acquiring a language must be able to distinguish among the phonemes, as well as produce them.

Infants' Discrimination of Phonemes

Early research on infant speech perception focused mainly on the ability of young (e.g., 1 to 6 months), English-learning infants to categorize and discriminate between pairs of speech sounds from the phonetic inventory of English. These early studies established that young infants are very good at discriminating a wide variety of speech sounds. Later research investigated how infants perceive speech sounds that do not occur in their native language (e.g., Werker and Tees, 1984; Best et al., 1988), the discrimination of longer segments of speech (e.g., Karzon, 1984), and the role of visual information in infants' speech perception (Rosenblum et al., 1977).

Eimas and his colleagues (Eimas et al., 1971) reported the first study of infant speech perception. This study used the HAS method described previously to show that infants between 1 and 4 months of age were able to discriminate between two stop consonants that differed in VOT ([ba-pa]). Other studies soon followed that investigated infants' discrimination of sounds differing in place of articulation. These studies showed infants discriminated [ba]-[ga], [bae]-[dae], [fa]-[øa], [va]-[øa], and [ma]-[na]. Infants were also shown to discriminate between sounds that differ in the manner of articulation, including [ra]-[la], [ba]-[wa], and [ba]-[ma] (see Jusczyk, 1997 for an in-depth review).

After several early studies demonstrated that infants could discriminate various consonant contrasts, other researchers investigated infants' perception of vowels (Trehub, 1973; Swoboda et al., 1976; Kuhl, 1983). These studies showed that 1- to 4-month-old infants also discriminate among a variety of vowels, including [a]-[I] (e.g., hod vs. hid), [I]-[i] (e.g., hid vs. heed), [a]-[aw] (e.g., hod vs. hawed). Thus, young infants also appear to be very good at discriminating among vowels. As noted earlier, the lack of acoustic invariance in speech sounds is a potential problem in discrimination, since the same sound produced by different talkers will have different acoustic characteristics. Several experiments by Patricia Kuhl and her colleagues have shown that infants are able to categorize and discriminate among vowels, even when they are spoken by different speakers (e.g., males and females; Marean et al., 1992), or have irrelevant acoustic variation (e.g., pitch contour; Kuhl and Miller, 1982).

Overall, these early studies established that young infants up to about 6 months of age are able to perceive the differences between consonants that differed along articulatory dimensions including VOT, place of articulation and manner of articulation, as well as discriminate between many vowels. Initially, infants' discrimination of consonants was tested with syllables that differed in their initial consonants (e.g., [ba]-[pa]). Other studies showed that infants could also perceive differences between syllables that differed in their final consonants, or when consonants were between two vowels, and in multi-syllable contexts. For example, young infants were shown to discriminate consonants in syllable-final position (e.g., pat vs. pad), between two vowels (e.g., [aba] vs. [apa]), and in two-syllable sequences (daba vs. daga]) (see Jusczyk, 1997). In order to more closely approximate natural speech, discrimination of consonants embedded in longer stretches of speech was also investigated (e.g., Goodsit et al., 1984). In a study by Karzon (1984), 1- to 4-month-old infants were shown to discriminate between sequences of three syllables when only a single consonant differed (e.g., [marana] vs. [malana]).

Categorical Perception

As noted earlier, the acoustic properties of the same speech sound can vary significantly when produced by different speakers, or by the same speaker in different phonetic contexts. Faced with this variation, human perceivers "categorize" speech sounds. That is, they are able to ignore irrelevant acoustic variation and focus on the properties that identify a phoneme as a member of a specific speech sound category. Because attention is focused on similarities among items of the same category, discrimination of different tokens or versions of a speech sound that are within a single category is usually difficult compared to speech sounds that are from different phonetic categories. That is, two spoken versions of [ba] sound more alike and are thus more difficult to discriminate than a [ba] and a [pa].

The sounds [ba] and [pa] in English differ in many ways, but one of most prominent differences is VOT. When producing [ba], the vocal folds begin to vibrate almost simultaneously with the opening of the lips, whereas with [pa] there is a noticeable lag in the onset of vocal fold vibration. It is possible to produce a VOT continuum of equal steps (e.g., 10 ms) from zero (i.e., vocal fold vibration simultaneous with lip opening) to 80 ms (i.e., vocal fold vibration starts 80 ms after lip opening). On such a continuum, the endpoints (i.e., zero and 80 ms) are heard clearly as [ba] and [pa], respectively. When native English-speaking adults are asked to label each token from the continuum, they typically show a rather abrupt shift from one category to another rather than a gradual change. That is, they may label items up to 40 ms as [ba], and items over 60 ms as [pa], while items at 50 ms might be labeled as [ba] half the time and [pa] half the time. Thus, a category boundary appears to exist between 40 and 60 ms. If asked to discriminate between pairs of sounds at 20 ms intervals along the continuum, it will be difficult to distinguish between two tokens on the same side of the category boundary, such as +10 and +30 ms, but easy to discriminate tokens that are from different sides of the boundary, such as +40 and +60. In other words, discrimination is poor within categories (e.g., two [pa]'s or two [ba]'s), but good when the pairs cross a category boundary (i.e., one [ba] and one [pa]), even when the physical difference between the items in the two pairs is equal. The degree to which native-language phonemes are perceived categorically varies. For example, stop consonants are usually perceived very categorically, while vowels are perceived less categorically (see Repp, 1984 for a detailed review of categorical perception).

Early studies showing young infants' ability to discriminate among various speech sounds were often followed by studies of categorical perception of the same sounds (reviewed in Miller and Eimas, 1983). These studies suggest that young infants, like adults, seem to have categorical perception of consonants that differ in voice-onset-time, place of articulation and manner of articulation. However, it should be noted that the procedures for establishing categorical perception are necessarily somewhat different for adults and infants, because infants are unable to directly identify speech sounds. Thus, the conclusion that infants perceive speech categorically relies on results showing that infants have more difficulty discriminating items that are within the adult categories than items that fall across category boundaries.

Audio-Visual Speech Perception

Speech is usually considered to be an acoustic event, and speech perception is typically seen as perceiving the acoustic structure of speech. However, there is strong evidence that under some circumstances, visual information from the face of a speaker can influence the perception of speech. Two kinds of influences of visual speech information have been studied in infants. In some studies infants have been shown to look longer at a video of a speaking face that matches a speech sound than at a face that is mismatched. For example, 6-month-old infants have been shown to look longer at a matching face for some consonant-vowel syllables (MacKain et al., 1983), even when the distinction is not in the native language (Pons et al., 2009). Four- and six-month-old infants have been shown to look more to a face that matches spoken vowels from the native language (Kuhl and Meltzoff, 1982; Patterson and Werker, 2003), and at least one study suggests the ability to match visual and acoustic speech information may be present at birth or very early in the post-natal period (Aldridge et al., 1999).

A second way visual information can influence speech perception is to alter how speech sounds are perceived. When a video of a speaker's face saying [ba] is combined with audio [ga], many (though not all) adults and children perceive [da] or [tha]. Thus, the auditory and visual information have been integrated to form a novel percept. Other combinations are also possible. This phenomenon was first demonstrated by Harry McGurk and John MacDonald (McGurk and MacDonald, 1976), and is referred to as the McGurk Effect.

A fascinating study by Larry Rosenblum and his colleagues (Rosenblum et al., 1977) demonstrated that five-month-old infants also appear to show the McGurk Effect. Rosenblum and his colleagues habituated infants with video clips that contained a matched face (visual) and voice (auditory) saying [va], and then tested them on three kinds of trials that paired different auditory information with the original visual [va]: 1) the original auditory [va]; 2) auditory [ba]; 3) auditory [da]. Infants differed in the rate at which they habituated to the three test stimuli. They were slower to habituate to audio [da] paired with visual [va] (perceived as [da] by adults) than when either auditory [va] or [ba] was paired with the visual [va] (both perceived as [va] by adults). These results suggest that the infants perceived the auditory [da] with visual [va] as different from either of the other two. That is, it appears the infants heard the visual [va] and auditory [da] as something other than [va], just as adults do, though it is unclear from these results that they heard the same thing ([da]) that adults hear. Other studies have confirmed that infants appear to be susceptible to the McGurk Effect (e.g., Burnham and Dodd, 2004). Thus, this form of auditory-visual speech perception appears to emerge early in infancy.

Infants' Perception of Non-native Speech Sounds

Early studies of infant speech perception clearly demonstrated that infants are quite adept at discriminating and categorizing a wide range of speech sounds. Some hypothesized that infants' speech perception acumen was evidence of an innate ability (e.g., Eimas et al., 1971). However, as a wider range of phoneme contrasts, language environments and ages of infants were studied, it became clear that infants didn't always discriminate all speech contrasts. For example, several studies (e.g., Eilers, 1977; Eilers et al., 1977) suggested that infants have difficulty distinguishing between some fricative sounds, although others reported success with some of these same fricative contrasts as young as 2-months of age (Levitt et al., 1988). Other studies found that some speech contrasts were discriminated by infants in one language environment but not another, such as the English [ba]-[pa] contrast which is well discriminated by English-learning infants, but not by infants learning Kikuyu (Streeter, 1976). Still other studies clearly show that infants can discriminate between some speech sounds not present in their ambient language environment which adult speakers of their native language could not discriminate (e.g. Trehub, 1976). Thus, the accumulating evidence suggested that the speech perception abilities of young infants must eventually become attuned to their native language. Beginning in the 1980s, the issue of how infant speech perception develops from a language-general ability that prepares infants to acquire any language to a language-specific ability became a central theme of research.

In a seminal series of experiments, Janet Werker and her colleagues (Werker and Tees, 1984; Werker and Lalonde, 1988) showed that infants' attunement to their ambient language environment begins by the end of the first year of life. Werker and Tees (1984) tested English-learning infants between 6 and 12 months of age on a native language phonetic contrasts ([ba] vs. [da]), as well as two non-native phonetic contrasts that adult and child native-English speakers could not discriminate. One of the non-native contrasts included two stop-consonants from Hindi. The phonemes were [ta], with a dental place of articulation (similar to English [ta]), and [ʈa] which has a more posterior place of articulation referred to as retroflex. Also included was a velar-uvular ejective contrast, [k'ae] vs. [q'ae], from the Native American language Nthlakapmx (also referred to as Puget Salish). The 6- to 8-month-old English-learning infants discriminated all three contrasts. However, the 8- to 10-month-old infants discriminated the non-native sounds less well than the younger infants, and the 10- to 12-month-old infants were generally unable to distinguish the non-native sounds. Both groups of older infants retained the ability to discriminate the native language [ba]-[da] contrast. This showed that infants' sensitivity to at least some non-native speech contrasts declines significantly by 10–12 months.

These findings initially seemed to suggest that the decline in discriminability was due to a lack of experience with the speech sounds (Aslin and Pisoni, 1980). That is, while young infants could discriminate speech sounds that did not occur in their native language, exposure or experience with the speech sounds seemed to be necessary to maintain that ability. However, research by Catherine Best and her colleagues has shown that the decline in perceptual sensitivity at 10 to 12 months only occurred for some non-native speech sounds. In one study (Best et al., 1988), English-learning infants from 6 to 14 months of age, and English-speaking adults, were tested on their ability to discriminate click contrasts from the Zulu language. The Zulu click sounds are very different from anything in the inventory of English speech sounds, and thus represent an example of speech sounds that English-learning infants and English-speaking adults would never experience in a speech context. The results of this study showed that native English-speaking adults were able to discriminate among several click consonant contrasts. Both older and younger infants were tested on a subset of the clicks the adults discriminated and were also able to discriminate the clicks. Thus, although it seems clear that exposure to language influences the development of speech perception, the results with the Zulu clicks demonstrated convincingly that it is not necessary to have experience with specific speech sounds in order to be able to continue discriminating them. The results of this study reframed the issue of the development of speech perception to why discrimination of some non-native speech sounds declines, while others continue to be well discriminated.

Best and her colleagues have investigated this question in a series of studies, in which the discrimination abilities of younger and older infants were compared to that of adults on a variety of non-native speech contrasts (Best and McRoberts, 2003; Best et al., 2001; Best et al., 1995). In these studies, the discrimination pattern for adults is usually established first, followed by infant tests. To establish the pattern of discrimination ability, each infant was tested three times, using two non-native phonetic contrasts, and a native-language contrast. Across several studies, English-learning infants and English-speaking adults have been tested on a wide variety of non-native sounds, including several additional Zulu contrasts (lateral voiced-voiceless fricatives [ɮa]-[ɬa]; voiceless aspirated velar stop-ejective [k^ha]-[k'a]; plosive-implosive bilabial stop [bʉ]-[ɓʉ]), a bilabial-alveolar ejective distinction from Tigrinya (Ethiopian) ([p'a]-[t'a]), the Nthlakapmx velar-uvular ejective contrast [k'ae] vs. [q'ae] from Werker's earlier study (Werker and Tees, 1984), as well as English bilabial-alveolar stop consonants [ba]-[da] and English alveolar voiced-voiceless fricatives [sa]-[za]. Younger infants discriminated all of the non-native contrasts. Older infants discriminated the clicks and fricatives from Zulu, the Tigrinya ejectives, and the stop-consonants and fricatives from English. They failed to discriminate the Nthlakapmx velar-uvular ejective contrast, confirming Werker's earlier results. However, even among the contrasts that were discriminated at both ages, the younger infants performed better than older infants on all tests except the Tigrinya [p'a]-[t'a] and the English [ba]-[da], which were discriminated equally well at both ages.

Infants' perception of non-native vowels has been studied less than non-native consonants. Discrimination of two non-native vowels by English-learning infants has been reported by 10 months of age (Polka and Werker, 1994). However, the development of vowel perception proceeds differently from the pattern seen for consonants. One difference is that the influence of the native language environment on vowel perception may occur as early as 6-months of age (Kuhl et al., 1992; Polka and Werker, 1994), whereas similar effects for consonants do not typically emerge until 10 to 12 months. In addition, there appear to be directional asymmetries in vowel discrimination (Polka and Bohn, 1996, 2003) that are not commonly reported with consonants (but may nonetheless exist). These asymmetries reflect the fact that infants discriminate better when they are habituated to one vowel (e.g., vowel A) and tested on a second vowel (e.g., vowel B), compared to when habituated to vowel B and tested on vowel A. As described earlier, different English vowels are produced by changing the height and front-back position of the tongue, as well as rounding, protruding or spreading the lips. A vowel space can be defined based on the extremes of tongue and lip positions. For example, for /i/ (as in "heed") the tongue is high and to the front of the oral cavity, and the lips are spread. The vowel /u/ (as in "who'd") is produced with the tongue high and to the back of the oral cavity, with the lips rounded and protruded. The vowels /ae/ (as in "had") and /a/ (as in "hod") are produced with the tongue low (on the floor of the oral cavity), and more to the front and back of the oral cavity, respectively. Other vowels are produced with less extreme positions. The vowels /i/, /u/, /ae/, /a/ define the limits of the English vowel space, and are thus considered most peripheral (i.e., nearest the periphery) in the space. There is some evidence that the asymmetries in vowel discrimination are related to the degree of peripherality within this vowel space. Vowels that are more peripheral in the space appear to act as reference points, so that discrimination is easier when infants are habituated to vowels that are less peripheral in the space and then tested on more peripheral vowels.

Tone languages use tone levels and patterns in addition to consonants and vowels to differentiate between words. This use of tone to differentiate lexical items is sometimes referred to as lexical tone to distinguish it from the use of prosodic changes in pitch, such as rising intonation for questions. Tone languages are common, especially in Asia and parts of Africa (Ladefoged and

Maddieson, 1996). The development of infants' perception of lexical tone has begun to receive research attention in the last decade. For example, Mattock and her colleagues (Mattock et al., 2008; Mattock and Burnham, 2006) demonstrated that discrimination of Thai tones declined between 6 and 9 months of age for infants learning English and French, non-tone languages, but not for infants learning Chinese, a tonal language. Thus, the developmental trajectory of tone appears to be similar to other aspects of phonetic perception in infants.

In conjunction with earlier studies, research on the perception of non-native phonemes show that infants enter the world as "language-general" speech perceivers, able to discriminate among most of the speech sounds of the world's languages, and over the first year or so of life, become attuned to many of the specifics of their language environment to become "language-specific" speech perceivers. Several theoretical models have been proposed to account for the results of infant non-native speech perception studies, including Kuhl's Native Language Magnet models (Kuhl, 1993, 2004; Kuhl et al., 2008), Jusczyk's Word Recognition and Phonological Structure Acquisition model (Jusczyk, 1997) and Werker's Processing Rich Information from Multidimensional Representations model (Werker and Curtin, 2005; Curtin et al., 2011). The model that most adequately accounts for the results, especially with non-native consonants, is Catherine Best's Perceptual Assimilation Model, or PAM (Best et al., 1988; Best, 1993; Best, 1994, Best, 1995). This model assumes listeners hear non-native speech sounds in terms of their native language phonetic categories whenever possible. This is called perceptual assimilation; that is, listeners assimilate non-native speech sounds into their native language phonological categories whenever there appears to be a reasonable fit. Within PAM, listeners can assimilate non-native phoneme contrasts into their native categories in several ways. Two contrasting non-native speech sounds can: 1) be assimilated into a single native language category (Single Category, or SC assimilation); 2) be assimilated into two different native language phonetic categories (Two Category, or TC assimilation); 3) be assimilated into two native language categories, but with different degrees of goodness (Category Goodness, or CG assimilation); 4) fail to be assimilated into any native language category (Non-Assimilable, or NA). Based on the earlier discussion of categorical perception, it should be clear that discrimination for SC assimilation will be very difficult (at or near chance levels) under most circumstances, but discrimination will be quite easy (near ceiling levels) in TC assimilation. In the case of CG assimilation, discrimination is intermediate between SC and TC. Finally, when non-native speech sounds are so different from any native language phoneme that assimilation does not occur (i.e., a Non-Assimilable or NA contrast), discrimination will be very good because the sounds will be perceived as non-linguistic, allowing perceivers to compare them on acoustic dimensions that would be irrelevant or unavailable in phonetic perception. Predictions from PAM about the degree of discriminability of non-native sounds (especially consonants) have been tested extensively with English-speaking adults, who have also provided descriptions of their assimilations of non-native phonemes into native language categories. In general, the predictions have been upheld. Although older infants often show the same pattern of discrimination as adults, Best and her colleagues believe that infants' speech sound categories are not fully developed by the end of the first year. Rather, they suggest that infants' perception of speech sounds becomes increasingly sensitive to information about how the sounds are produced (i.e., phonetic or articulatory information). Thus, near the end of the first year of life, infants are beginning to perceive many, but not all, of the details that specify how native language speech sounds are produced. Therefore, their pattern of discrimination (assimilation) of non-native speech sounds is becoming increasingly adult-like, but further development occurs as more fine-grained phonetic detail is achieved through infancy and into early childhood (Best et al., 2016).

Speech Preferences in Infancy

The research on phonetic discrimination abilities provides clear evidence of infants' underlying speech perception capabilities, as well as the timeline of some of the developmental processes involved in acquiring a native language phonology. The development of infants' speech perception from language-general to language-specific over the course of the first year focuses attention on the early influence of the ambient language environment on language development. One question of interest is whether infants attend preferentially to some aspects of the language environment over others. This is tested using the auditory preference procedure, which differs from the discrimination procedures described earlier. Rather than assessing infants' ability to detect differences between smaller units of speech, such as phonemes or syllables, preference procedures allow researchers to study how infants respond to longer samples of speech that more realistically approximate what they normally hear. As a result, studies of infant speech preferences have taken a significant role in research on the development of speech perception and language development. Early studies of speech preferences focused on infants' preference to listen to speech over non-speech sounds, and Infant-Directed Speech (IDS, or Child-Directed Speech, CDS) over Adult-Directed Speech (ADS). Other studies used speech preferences to investigate the development of sensitivity to various aspects of language structure in speech.

Auditory and Speech Preference Procedures

Auditory preferences in infants were first demonstrated by Bernard Freidlander (1968). The procedure involved an apparatus with two large knobs, a speaker and an activity recorder. By manipulating the knobs, infants could activate one of two recorded audio samples. The activity recorder collected the amount of time each sample was played. Infants aged 11 to 15 months showed a preference by listening more (longer) to some sounds, such as their mothers' voice, over other sounds, such as simple musical passages. Friedlander's apparatus was later modified for use with younger, less mobile infants. Anne Fernald (1985) developed a related procedure, that could be used with young infants. This procedure required only a head-turn response to activate sounds that played from speakers located to the infant's right and left. No contingency was required to continue sound presentation. Eventually, infant-controlled variants were developed using both gaze and sucking.

Two versions of the auditory preference procedure are in common use today. Both are infant-controlled procedures that use gaze as the behavioral measure. In one variant, a single visual target is located directly in front of the infant (Central or Single Fixation Preference; Cooper and Aslin, 1990). The other approach is a direct descendent of Fernald's procedure (Fernald, 1985), and requires a head turn to one of two visual targets located on the infant's right and left (Head-Turn Preference Procedure, or HPP; Kemler Nelson et al., 1995). In the central fixation procedure, infants fixate a centrally presented visual target, such as a checkerboard. Stimulus sounds are played through a speaker located directly below the visual target. Sound presentation on each trial is contingent on maintaining fixation of the target. When fixation is broken for more than one or 2 s, the sound ends, the visual target is removed, and the trial ends. After a brief delay, the visual target returns, signaling the availability of the next trial.

The head turn procedure is somewhat more complicated than the central fixation procedure. It uses two speakers, each 90° to the infants' right and left. Small red lights are usually placed near the speakers, and directly in front of the infant. A test typically begins with several presentations of stimuli from each speaker to familiarize infants with the procedure and train them to turn toward the lateralized lights and speakers in order to initiate a trial. Trials begin with the central red light flashing. When the infant orients to the flashing light, it is extinguished and one of the lateral lights blinks to indicate a stimulus is available. Stimulus types can be associated with a specific side or can be randomized to either side. Infants are required to make a criterial head turn (e.g., at least 45°) to initiate a trial, and they must maintain the head turn to continue sound presentation on each trial.

Both procedures are infant-controlled, and typically have trials evenly divided between two types of speech presented on alternate or randomized trials. As in the discrimination procedures discussed previously, fixation time on each trial is used as an index of interest or listening. Fixation times are averaged across trials of each stimulus type, resulting in a mean for each stimulus type.

Although both procedures are in current use, the central fixation procedure seems to have several advantages over the head turn procedure: 1) it eliminates the effects of lateral biases evident in some infants; 2) it does not require training or familiarizing infants with head-turning to initiate trials; 3) observers are not required to judge whether infants maintain a sufficient degree of head-turn to continue a trial; 4) it can be used across a wider age range (newborns to 2 years or older).

Infants' Speech Preferences

The first study to address young infants' speech preferences was performed by Anne Fernald (1985), who investigated 4-month-old infants' listening preference for IDS by female speakers compared to speech by the same speakers to another adult, or ADS. Fernald used the non-contingent head turn procedure described earlier. Her results showed that infants turned to the side that activated the IDS speech samples more often than the ADS speech samples. In other studies infants have showed more positive affect when listening to IDS than ADS (e.g., Fernald, 1993; Werker et al., 1994).

Other studies have explored a variety of aspects of infants' preference for IDS over ADS. For example, it appears that infants' preference for IDS is present from very early in the post-natal period (e.g., Cooper and Aslin, 1990). And although some studies suggested that the preference for IDS might decline or disappear in older infants (Hayashi et al., 2001; Newman and Hussain, 2006), more recent studies show that infants as old as 14 months of age continue to show a preference for IDS over ADS (McRoberts et al., 2009). Still other studies have demonstrated that infants will attend to IDS in an unfamiliar language over ADS from the same language (Werker et al., 1994), and also prefer male speakers' IDS over the same speakers' ADS (Werker and McLeod, 1989).

The acoustic basis for infants' preference for IDS over ADS is somewhat in dispute. An early study (Fernald and Kuhl, 1987) showed that 4-month-old infants listened more to sine-wave analogues of the F0 contours of IDS than ADS, but not to sine-wave analogues of the amplitude envelope or temporal structure of IDS. Thus, it appeared that the higher F0, wider F0 range, and expanded intonation contours typical of IDS were the acoustic basis of infants' preference. Several recent studies have noted that early studies of IDS preference often confounded the prosody of IDS with affect, because IDS typically contains expressions of positive emotion. These studies have shown that IDS prosody by itself, in the absence of expressions of positive affect, is not sufficient to result in a preference by infants at 6-months of age (Singh et al., 2002; Kitamura and Burnham, 1998). In addition, when prosodic factors such as F0 range are controlled, infants prefer to listen to ADS containing positive affect over IDS that does not contain positive affect. Thus, it appears that by 6 months of age, infants may be attending to the positive affective expressiveness of IDS, even if other prosodic characteristics of IDS are absent. Nonetheless, in spontaneous interactions with infants, the typical prosodic characteristics of IDS, including higher F0 and wider F0 range, will normally be highly correlated with positive affect. It is also possible that the basis for infants' IDS preference changes as infants mature and become more sophisticated perceivers of speech.

The possibility that the basis for IDS preferences changes over the first year of life was investigated by McRoberts and his colleagues in a series of experiments testing infants at several ages with IDS directed to younger (4–6 months) and older (12–14 months) infants vs. ADS by the same speakers (McRoberts et al., 2009). The stimuli were contiguous utterances from naturally produced, spontaneous speech by mothers interacting with their infants (IDS) and another adult (ADS). In two initial experiments, 6-, 10- and 14-month-old infants preferred IDS directed toward younger infants. In addition, 4-, 8-, 10- and 14-month-old infants preferred IDS directed toward older infants. However, in both experiments six-month-old infants failed to prefer IDS to older infants. One difference between speech to older and younger infants is the frequency of repeated utterances, which appears to reach a maximum between 4 and 6 months of age (e.g., Stern et al., 1983). The IDS stimuli to older infants was manipulated (digitally edited) to have the same frequency of repeated utterances as IDS directed toward younger infants. A third experiment revealed that 6-month-old infants preferred this repetition-enhanced IDS over the ADS by the same speakers. The manipulation that produced the repetition-enhanced IDS in this experiment confounded verbal repetition and prosodic repetition. In a final experiment, new IDS stimuli were produced that contained verbal repetition without prosodic repetition. Two new stimulus sets were constructed from this IDS. One set contained immediately repeated utterances (e.g., AABBCCDDEE) and one contained the same utterances without immediate repetition (e.g., ABCDEBCAD). Six-month-old infants preferred the IDS that contained immediately repeated utterances compared to the same IDS utterances presented without immediately repeated utterances. However, 4-month-olds did not demonstrate a preference. Thus, between 4- and 6-months of age infants shift their attention from prosodic aspects of IDS to linguistic structure in the form of repeated utterances. The results also confirm that older infants continue to listen preferentially to IDS.

Beyond establishing infants' preference to listen to IDS over ADS, auditory preference procedures have been used to study infants' detection of a wide variety of linguistic structures in speech, such as patterns of lexical stress (e.g., pres'-ent vs. pres-ent') (Jusczyk et al., 1993), locating words in sentences (Jusczyk and Aslin, 1995), sensitivity to rhyming (Braze, et al., 2011), and native language phonotactic patterns (Frederici and Wessels, 1993). Preference studies have also demonstrated very young infants' ability to perceive differences between languages based on prosodic patterns. Several studies have shown that newborn infants listen longer to speech from languages that have rhythmic patterns that are similar to their ambient language (e.g., French) over languages with different prosodic patterns (e.g., Russian) (e.g., Mehler et al., 1988). However, not until 4–5 months do infants show a preference for a familiar language over languages with similar prosodic structure (e.g., English vs. Dutch; Nazzi et al., 2000).

Another use of the preference procedure is to pair it with familiarization to stimuli for which infants initially have no preference. Peter Jusczyk and Richard Aslin (1995) used this approach in a landmark study showing that by 7.5 months of age, infants begin to segment words from fluent speech. In their initial study, infants were familiarized with a repeated list of single words spoken in isolation (either cup or bike). The infants were then tested for a listening preference with sentences containing the familiarized word (e.g., cup) and similar sentences containing the unfamiliar word (e.g., bike). Although 6-month-old infants showed no preference, 7.5-month-old infants preferred the sentences containing the familiarized word. Follow-up experiments reversed the procedure, familiarizing the infants with the sentences that contained the word, and then testing on the familiarized word versus the unfamiliar word, each repeated in isolation. The outcome confirmed the initial results, showing that infants listened longer to the familiarized words. Additional experiments showed that infants failed to prefer words that differed from the familiarized target by one phoneme, such as "zeet" and "gike" instead of "feet" and "bike" (Jusczyk and Aslin, 1995). Taken together, these results provide evidence that soon after the middle of their first year, infants can extract and remember phonetic strings that occur in running speech.

Both speech discrimination studies and speech preference studies provide important insights into how infants become attuned to the properties of their native language. Perhaps as early as birth, infants can distinguish their native language and similar languages from others based on prosodic patterns. Over the course of the first year, infants become attuned to the acoustic and likely to the articulatory aspects of their native language. By the middle of the first year, infants begin to attend to the linguistic structure of speech at the level of utterances and begin to extract potential words. In general, research over the past 40-plus years demonstrates both how competent even young infants are at perceiving speech and how rapidly their speech perception capabilities develop over the first 12 to 18 months.

References

Aldridge, M.A., Braga, E.S., Walton, G.E., Bower, T.G.R., 1999. The intermodal representation of speech in newborns. Dev. Sci. 2, 42–46.

Aslin, R.N., Pisoni, D.B., 1980. Some developmental processes in speech perception. In: Yeni-Komshian, G., Kavanagh, J., Ferguson, C. (Eds.), Child Phonology: Perception and Production. Academic Press, New York, pp. 67–96.

Best, C.T., 1993. Emergence of language-specific constraints in perception of non-native speech contrasts: a window on early phonological development. In: de Boysson-Bardies, B., de Schonen, S., Jusczyk, P., MacNeilage, P., Morton, J. (Eds.), Developmental Neurocognition: Speech and Face Processing in the First Year of Life. Kluwer Academic, Dordrecht, The Netherlands, pp. 289–304.

Best, C.T., 1994. The emergence of native-language phonological influences in infants: a perceptual assimilation model. In: Nusbaum, H.C. (Ed.), The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words. MIT Press, Cambridge, MA, pp. 167–224.

Best, C.T., 1995. A direct realist perspective on cross-language speech perception. In: Strange, W., Jenkins, J.J. (Eds.), Cross-language Speech Perception. York Press, Timonium, MD, pp. 171–204.

Best, C.T., McRoberts, G.W., Goodell, E., 2001. American listeners' perception of nonnative consonantcontrasts varying in perceptual assimilation to English phonology. JASA 109, 775–794.

Best, C.T., McRoberts, G.W., 2003. Infant perception of non-native consonant contrasts that adults assimilate in different ways. Lang. Speech 46, 183–216.

Best, C.T., McRoberts, G.W., Sithole, N., 1988. Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. J. Exp. Psychol. Hum. Percept. Perform. 14, 345–360.

Best, C.T., McRoberts, G.W., LaFleur, R., Silver-Isenstadt, J., 1995. Divergent developmental patterns for infants' perception of two non-native contrasts. Infant Behav. Dev. 18, 339–350.

Best, C.T., Goldstein, L.M., Nam, H., Tyler, M.D., 2016. Articulating what infants attune to in native speech. J. Ecol. Psychol. 28, 216–261.

Braze, D., McRoberts, G.W., McDonough, C., 2011. The role of early phonological sensitivity in the development of phonological awareness. In: Braze, D., Fowler, C., Brady, S. (Eds.), Explaining Individual Differences in Reading: Theory and Evidence. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ.

Burnham, D., Dodd, B., 2004. Auditory-visual speech integration by prelinguistic infants: perception of an emergent consonant in the McGurk effect. Dev. Psychobiol. 45, 204–220.

Cooper, R., Aslin, R., 1990. Preference for infant-directed speech in the first month after birth. Child Dev. 61, 1584–1595.

Curtin, S., Byers-Heinlein, C., Werker, J.F., 2011. Bilingual beginnings as a lens for theory development: PRIMIR in focus. J. Phonet. 39, 492–504.

Eilers, R.E., 1977. Context sensitive perception of naturally produced stop and fricative consonants by infants. J. Am. Acoustic. Soc. 61, 1321–1336.

Eilers, R.E., Wilson, W.R., Moore, J.M., 1977. Developmental changes in speech discrimination in infants. J. Speech Hear. Res. 20, 766–780.
Eimas, P.D., Siqueland, E.R., Jusczyk, P.W., Vigorito, J., 1971. Speech perception in infants. Science 171, 303–306.
Fernald, A., 1985. Four-month-old infants prefer to listen to motherese. Infant Behav. Dev. 8, 181–195.
Fernald, A., 1993. Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages. Child Dev. 64, 657–666.
Fernald, A., Kuhl, P.K., 1987. Acoustic determinants of infant preference for motherese speech. Infant Behav. Dev. 10, 279–293.
Fowler, C.A., Best, C.T., McRoberts, G.W., 1990. Young infants' perception of liquid coarticulatory influences on following stop consonants. Percept. Psychophys. 48, 559–570.
Frederici, A.D., Wessels, J.M.I., 1993. Phonotactic knowledge and its use in infant speech perception. Percept. Psychophys. 54, 287–295.
Freidlander, B.Z., 1968. The effect of speaker identity, voice inflection, vocabulary and message redundancy on infant's selection of vocal reinforcement. J. Exp. Child Psychol. 6, 443–459.
Goodsit, J.V., Morgan, J.L., Kuhl, P.K., 1984. Perceptual strategies in prelingual speech segmentation. J. Child Lang. 20, 229–252.
Hayashi, A., Tamekawa, Y., Kiritani, S., 2001. Developmental change in auditory preferences for speech stimuli in Japanese infants. J. Speech Lang. Hear. Res. 44, 1189–1200.
Horowitz, F.D., 1975. Visual attention, auditory stimulation, and language discrimination in infants. Monogr. Soc. Res. Child Dev. 39, 5–6. Serial No. 158.
Jusczyk, P.W., 1997. The Discovery of Spoken Language. The MIT Press, Cambridge, MA.
Jusczyk, P., Aslin, R.N., 1995. Infants' detection of sound patterns of words in fluent speech. Cogn. Psychol. 29, 1–23.
Jusczyk, P.W., Cutler, A., Redanz, N., 1993. Preference for the predominant stress patterns of English words. Child Dev. 64, 674–687.
Karzon, R.G., 1984. Discrimination of a polysyllabic sequence by one- to four-month-old infants. J. Exp. Child Psychol. 39, 326–342.
Kemler Nelson, D.G., Jusczyk, P.W., Mandel, D.R., Myers, J., Turk, A., Gerken, L., 1995. The head-turn preference procedure for testing auditory perception. Infant Behav. Dev. 18, 111–116.
Kitamura, C., Burnham, D., 1998. The infant's response to vocal affect in maternal speech. In: Rovee-Collier, C., Lipsett, L., Hayne, H. (Eds.), Advances in Infancy Research, vol. 12. Ablex, Stamford, CT, pp. 221–236.
Kuhl, P.K., 1983. Perception of auditory equivalence classes for speech in early infancy. Infant Behav. Dev. 6, 263–285.
Kuhl, P.K., 1985. Methods in the study of infant speech perception. In: Gottlieb, G., Krasnegor, N.A. (Eds.), Measurement of Audition and Vision in the First Year of Postnatal Life: A Methodological Overview. Ablex, Norwood, NJ.
Kuhl, P.K., 1993. Innate predispositions and the effects of experience on speech perception: the native language magnet theory. In: de Boysson-Bardies, B., de Schonen, S., Jusczyk, P., MacNeilage, P., Morton, J. (Eds.), Developmental Neurocognition: Speech and Face Processing in the First Year of Life. Kluwer Academic, Dordrecht, The Netherlands, pp. 259–274.
Kuhl, P.K., 2004. Early language acquisition: cracking the code. Nat. Rev. Neurosci. 5, 831–843.
Kuhl, P.K., Meltzoff, A.N., 1982. The bimodal development of speech in infancy. Science 218, 1138–1141.
Kuhl, P.K., Miller, J.D., 1982. Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. Percept. Psychophys. 31, 279–282.
Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by six months of age. Science 255, 606–608.
Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T., 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLMe). Philos. Transac. R. Soc. 36, 979–1000.
Ladefoged, P., 1975. A Course in Phonetics. Harcourt Brace, Orlando.
Ladefoged, P., Johnson, K., 2015. A Course in Phonetics, seventh ed. Cengage, Stamford, CT.
Ladefoged, P., Maddieson, I., 1996. The Sounds of the World's Languages. Blackwell, Oxford.
Levitt, A., Jusczyk, P.W., Murray, J., Carden, G., 1988. The perception of place of articulation contrasts in voiced and voiceless fricatives by two-month-old infants. J. Exp. Psychol. Hum. Percept. Perform. 14, 361–368.
MacKain, K., Studdert-Kennedy, M., Spieker, S., Stern, D., 1983. Infant intermodal speech perception is a left-hemisphere function. Science 219, 1347–1349.
Marean, G.C., Werner, L.A., Kuhl, P.K., 1992. Vowel categorization by very young infants. Dev. Psychol. 28, 396–405.
Mattock, K., Burnham, D., 2006. Chinese and English infants' tone perception: evidence for perceptual reorganization. Infancy 10, 241–265.
Mattock, K., Molnar, M., Polka, L., Burnham, D., 2008. The developmental course of lexical tone perception in the first year of life. Cognition 106, 1367–1381.
McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746–748.
McRoberts, G.W., McDonough, C., Lakusta, L., 2009. The development of infant speech preferences from 4- to 14-months of age. Infancy 14, 162–194.
Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, N., Bertoncini, J., Amiel-Tison, C., 1988. A precursor to language acquisition in young infants. Cognition 29, 144–178.
Miller, J.L., Eimas, P.D., 1983. Studies on the categorical perception of speech by infants. Cognition 13, 135–165.
Nazzi, T., Jusczyk, P.W., Johnson, E.K., 2000. Language discrimination by English- learning 5-month-olds: effects of rhythm and familiarity. J. Mem. Lang. 43, 1–19.
Newman, R., Hussain, I., 2006. Changes in preference for infant-directed speech in low and moderate noise by 4.5- to 13-month-olds. Infancy 10, 61–76.
Oakes, L., 2010. Using habituation of looking time to assess mental processes in infancy. J. Cogn. Dev. 11, 255–268.
Patterson, M.L., Werker, J.F., 2003. Two-month-old infants match phonetic information in lips and voice. Dev. Sci. 6, 191–196.
Polka, L., Bohn, O., 1996. A cross-language comparison of vowel perception in English-learning and German-learning infants. J. Acoust. Soc. Am. 100, 577–592.
Polka, L., Bohn, O., 2003. Asymmetries in vowel perception. Speech Commun. 41, 221–231.
Polka, L., Werker, J.F., 1994. Developmental changes in perception of nonnative vowel contrasts. J. Exp. Psychol. Hum. Percept. Perform. 20, 421–435.
Pons, F., Lewkowicz, D.J., Soto-Faraco, S., Sebastián-Gallés, N., 2009. Narrowing of intersensory speech perception in infancy. Proc. Natl. Acad. Sci. Unit. States Am. 106, 10598–10602.
Repp, B., 1984. Categorical perception: issues, methods, findings. Speech Lang. 10, 243–335.
Rosenblum, L.D., Schmuckler, M.A., Johnson, J.A., 1997. The McGurk effect in infants. Percept. Psychophys. 59, 347–357.
Singh, L., Morgan, J.L., Best, C.T., 2002. Infants' Listening Preferences: Baby Talk or Happy Talk? Infancy 3, 365–394.
Stern, D., Spieker, S., Barnett, R.K., MacKain, C., 1983. The prosody of maternal speech: infant age and context-related changes. J. Child Lang. 10, 1–15.
Streeter, L.A., 1976. Language perception of 2-month old infants shows effects of both innate mechanisms and experience. Nature 259, 39–41.
Swoboda, P., Morse, P.A., Levitt, L.A., 1976. Continuous vowel discrimination in normal and at-risk infants. Child Dev. 49, 332–339.
Trehub, S.E., 1973. Infant sensitivity to vowel and tonal contrasts. Dev. Psychol. 9, 91–96.
Trehub, S.E., 1976. The discrimination of foreign speech contrasts by infants and adults. Child Dev. 47, 466–472.
Werker, J.F., Curtin, S., 2005. PRIMIR: a developmental framework of infant speech processing. Lang. Learn. Dev. 1, 197–234.
Werker, J.F., Lalonde, C.E., 1988. Cross-language speech perception: initial capabilities and developmental change. Dev. Psychol. 24, 672–683.
Werker, J.F., McLeod, P., 1989. Infant preference for both male and female infant-directed talk: a developmental study of attentional and affective responsiveness. Can. J. Psychol. 43, 230–246.
Werker, J.F., Tees, R.C., 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behav. Dev. 7, 49–63.
Werker, J.F., Pegg, J., McLeod, P., 1994. A cross-language investigation of infant preference for infant-directed communication. Infant Behav. Dev. 17, 323–333.

Relevant Websites

Haskins Laboratories, http://haskins.yale.edu.
University of British Columbia Infant Studies, http://infantstudies.psych.ubc.ca.
Juscyk Lab Final Report, http://hollich.psych.purdue.edu/Jusczyk.
Patricia K. Kuhl, Ph.D., http://ilabs.uw.edu/institute-faculty/bio/i-labs-patricia-k-kuhl-phd.
Stanford University Center for Infant Studies, https://web.stanford.edu/group/centerinfstudies.
The McGill Infant Development Centre, http://www.psych.mcgill.ca/labs/bebe/english/index.html.
Professor Catherine (Cathi) Best, https://www.westernsydney.edu.au/marcs/our_team/researchers/professor_catherine_best.