Introduction

For children, learning a new language starts with listening to its characteristic sounds and rhythms. Speech perception refers to how listeners hear and interpret these speech sounds. Very young infants are often called “universal listeners”. Their speech perception is broad enough to acquire any native language. With experience, the perceptual system undergoes reorganization: listeners become specialized at attending to just those sound contrasts that are relevant in their own language(s). This process is called perceptual narrowing (see Figure 8.1). While the most dramatic period of perceptual development occurs during the first year of life, listeners continue to tune their speech perception systems throughout childhood and into adolescence before attaining adult speech perception skills.

Speech perception is a complex problem for children who encounter two languages, whether simultaneously or sequentially. Let us first consider simultaneous bilingual infants, who must tune their perceptual systems to two native languages from birth. Byers-Heinlein and Fennell (2014) described four properties of early bilingual environments that are particularly important for understanding speech perception in bilingual infants. First, bilinguals have less exposure to each language than monolinguals, because their time is divided between the two languages. Second, bilingual infants must mentally represent two languages, increasing the number of perceptual categories they must learn. Third, bilingual exposure is “noisy”. For example, bilingual children often have bilingual caregivers who might speak with an accent in one or more languages. Finally, to learn both languages, simultaneous bilinguals must discriminate and separate their languages and engage in language-specific processing, a potentially challenging task. These differences between monolingual and bilingual environments contribute to how the two groups develop their speech perception skills.

Child second language (L2) learners, who learn their two languages sequentially, also follow a unique developmental path. When they begin learning a second language, their first language system has been largely acquired, but is not yet fully adult-like. This means that child L2 learners acquire a new perceptual system on top of their still-developing L1 system. In the domain of speech perception, children’s perceptual sensitivities to non-native sounds begin to decline sharply in the first year of life, and
this decline continues for some years (see Chapter 1 for a discussion of sensitive and critical periods). As a consequence, when children encounter a new language, their perceptual systems will neither be as open as infants’ nor as entrenched as adults.’ For this reason, we should expect child L2 learners to show perceptual development that is different from both simultaneous bilinguals and adult L2 learners.

This chapter will consider these issues as we explore speech perception in bilingual infants and child L2 learners. We will begin by discussing how simultaneous bilingual infants learn to discriminate and separate their two languages, and why this is important for their later language learning. Next, we will consider their phonological development, which is the perception of speech units smaller than a word. We will cover phonetic perception of vowel and consonant sounds, as well as lexical stress and lexical tone. Finally, we will turn to phonetic development and phonological awareness in preschool and school-aged L2 learners. We will pay special attention to the role of the language pair being learned, children’s language dominance, and comparisons with monolingual children.

Simultaneous Bilingual Infants

Language Discrimination and Separation

When infants are born into a bilingual environment, no one tells them when each language is being spoken. Instead, they must use their perceptual abilities to identify and discriminate their languages. Fortunately, infants are born with perceptual sensitivities that support language discrimination.

To investigate these sensitivities at birth, researchers must develop research methods that can be used with newborn infants. This is challenging, because newborns cannot
do much—they cannot answer an experimenter’s questions, press a button, or even move their heads very well. One procedure that can be used with newborn infants is called high-amplitude sucking, which capitalizes on their sucking reflex. In this procedure, newborns are offered a pacifier connected to a computer, which measures the strength and frequency of their sucking. Whenever newborns produce a strong or “high amplitude” suck, the computer plays a stimulus sound. Researchers can assess infants’ interest in different sounds by the number of sucks they produce: the more sucking, the greater the interest.

There are two variants of the high amplitude sucking procedure: one that tests newborns’ preference for different sounds and one that tests their discrimination of different sounds. To test newborns’ preference for different sounds, two types of sounds are played during alternating minutes. Researchers then measure which type of sounds elicits more sucks, indicating the one that newborns prefer. To test newborns’ discrimination, researchers use a high amplitude sucking procedure that includes a habituation phase and a test phase. During the habituation phase, newborns first hear one type of sound each time they produce a high amplitude suck. This continues until their sucking rate decreases, indicating that the newborns are habituated or “bored” with that sound. At this point, the test phase begins. Half of the newborns (the experimental condition) hear a new sound each time they suck, and the other half (the control condition) still hear the old sound each time they suck. If newborns can tell the difference between the new sound and the old sound, infants in the experimental group should suck more during the test (indicating they are interested in the new sound) than those in the control group (who are bored at still hearing the old sound).

Byers-Heinlein, Burns, and Werker (2010) used high-amplitude sucking to study a group of newborn infants whose mothers spoke both English and Tagalog (a Filipino language), and compared them to a group of newborns whose mothers spoke only English. Because fetal hearing is well developed in the third trimester of pregnancy, the babies already had either monolingual or bilingual experience from listening to their mothers’ voices in the womb. In a first study, the researchers tested whether infants could recognize their native language(s) using the preference version of the high-amplitude sucking procedure. On alternating minutes, newborns heard either English or Tagalog each time they sucked. The researchers compared the number of sucks the infants produced while hearing English versus Tagalog. English newborns sucked more during the English minutes, showing that they preferred listening to their native language (English). Bilingual newborns sucked similarly during English and Tagalog minutes, suggesting that they recognized both as their native languages.

In a second study, Byers-Heinlein et al. (2010) used the habituation variant of the high-amplitude sucking method to test whether the bilingual newborns could tell the difference between their two languages, even though they didn’t have a preference for one over the other. Using the habituation variant of the high-amplitude sucking procedure, English–Tagalog bilingual newborns were habituated to sentences from one of the languages, either English or Tagalog. At test, the language switched for infants in the experimental condition, but remained the same for infants in the control condition. Newborns who heard the language switch increased their sucking, suggesting that they noticed the switch and had therefore discriminated the languages. Those in the control group, who did not hear the language switch, continued to be disinterested. In sum, these studies of bilingual newborns provided two important insights. First, both
monolingual and bilingual infants are born ready to pay special attention to their native language(s). Second, at birth, bilingual newborns are able to discriminate their two native languages.

What kind of information do infants use to recognize and discriminate languages? Research with monolingual newborns has suggested that they are especially sensitive to languages’ rhythm (Nazzi, Bertoncini, and Mehler 1998). English and Tagalog are rhythmically different, so they are relatively easy to discriminate. However, what about bilingual infants learning languages that are rhythmically similar?

While this question has not been investigated in bilingual newborns, it has been investigated in somewhat older infants, most often using looking-time methods. Infants’ looking is used as an index of their listening, as previous research has shown that looking and listening are tightly coupled in infancy. In a typical looking-time study, infants are seated in a darkened room (on their parents’ laps or in a high chair) in front of a monitor and loudspeakers. A researcher controls the study from a computer in an adjacent room, playing the infant different stimuli on different trials. Infants’ looking to the monitor is measured either automatically with an eye-tracker or by the researcher pressing a button whenever the infant is looking at the monitor.

To test infants’ discrimination of two sounds, infants are familiarized or habituated to one type of sound while seeing a neutral visual stimulus (often a checkerboard). Infants tend to look at the screen a lot initially as they are interested in the sound, but their looking time then declines as they get bored and are thus habituated to the sound. At this point, the test phase begins. During this phase, infants hear either a new sound (the experimental condition) or the old sound (the control condition). If they discriminate the new sound from the old sound, then they should look longer when hearing the new sound than the old sound.

Two studies using a looking-time method have shown that by age 3–4 months, bilingual infants can discriminate their languages even if they are rhythmically similar. One study tested Spanish–Catalan bilingual infants (Bosch and Sebastián-Gallés 2001) and another tested Spanish–Basque bilingual infants (Molnar, Gervain, and Carreiras 2014). Bilingual infants were familiarized to sentences from one of their languages. As expected, infants’ interest in the language declined over time, which they showed by looking less at the picture. At test, infants in one group heard sentences from the other language (so if they had been hearing Spanish, they now heard Catalan or Basque), while infants in a control group heard more of the same language. In both studies, infants regained interest when the language switched at test, showing that bilingual infants can discriminate even rhythmically similar languages. Together, studies of infants’ language discrimination suggest that within the first few months of life, infants have the perceptual ability to tell apart any language pair.

Intriguingly, bilingual infants are not only sensitive to differences between how their two languages sound but also to how they look. Using a looking-time method, Weikum et al. (2007) habituated French–English bilingual infants and monolingual English infants to videos of faces speaking either English or French. The videos’ sound was muted, so that infants only had access to visual information available on the face. At test, researchers switched the language for half of the infants and kept it the same for the other half. Bilingual infants could visually discriminate their languages at 4, 6, and 8 months, increasing their looking at the screen when the language switched. Monolinguals could also discriminate the languages at 4 and 6 months. However, unlike
bilinguals, they failed to discriminate the languages at 8 months. This pattern was found even when infants were tested using non-native languages: bilinguals could discriminate the languages at all ages, but monolinguals could no longer discriminate them at 8 months (Sebastián-Gallés et al. 2012). These results provide an example of perceptual narrowing. Initially, both monolinguals and bilinguals have broad perceptual sensitivities and notice the visual differences between languages. As monolingual infants seldom need to discriminate different languages, they lose sensitivity to this difference. For bilinguals, language discrimination is an important part of their everyday experience, and they maintain sensitivity to visual differences between languages.

Early language discrimination relies on infants’ perceptual skills and is foundational for successful bilingual development. However, note that a full, deeper understanding of the nature of their two languages – for example, that each language has a separate vocabulary, grammar, and so on – likely emerges more gradually across development (Byers-Heinlein 2014).

Research is only beginning to investigate language discrimination and separation after infancy. One experimental approach has been to measure event-related potentials (ERPs), which are electrical signals produced by the brain in response to stimuli. In ERP studies, infants wear a stretchy cap embedded with electrodes that passively pick up the brains’ electrical activity. To test discrimination, researchers can use what is called an oddball paradigm. Infants encounter one type of stimulus repeatedly (called the “standard”) and then occasionally encounter a new type of stimulus (the “oddball”). If the brain notices the difference between the standard and the oddball, then it will produce a characteristic pattern of electrical activity (the event-related potential) when the oddball sound is played. A different response to the oddball compared to the standard indicates that infants can discriminate the two types of stimuli.

One study of 2–3 year old English–Welsh bilinguals used ERPs to examine whether English–Welsh bilinguals could discriminate between words in each of their languages (Kuipers and Thierry 2012). They also tested English monolingual toddlers as a comparison group. Children saw a series of familiar pictures that were labeled in one of their languages (e.g., a bed, a duck). Infants heard several trials in which a picture was labeled in one language (e.g., English) and then an oddball trial where the picture was labeled in the other language (e.g., Welsh). Both the bilinguals and the monolinguals showed a brain response that indicated they detected the language change, although there were differences in the timing of the response, suggesting differences in processing.

Another interesting study linked bilingual infants’ speech perception with sensitivity to their languages’ grammatical structure. This study used a different type of looking-time paradigm called the head-turn preference procedure (see Figure 8.2). As in other looking-time procedures, the head-turn preference procedure uses infants’ looking behaviour to measure their interest in different types of stimuli. Here, screens (or sometimes lightbulbs) are placed to the infants’ left and right sides. On each trial, one of the lights begin flashing and an auditory stimulus begins playing as soon as infants turn their head to look at the light. Both the auditory and visual stimuli continue as long as infants are turned in that direction. Once infants become disinterested and look away, the trial stops, their attention is drawn back to the centre of the room, and the next trial begins. Researchers can play different stimuli on different trials, and measure which stimulus garners the most looking time and hence the greatest interest.
Gervain and Werker (2013) used the head-turn preference procedure to study 7-month-old bilingual infants who were learning two languages with different word orders. For example, in English one would say *eat an apple* (verb–object word order, VO) while in Japanese one would say *ringo-wo taberu* “apple eat” (object–verb word order, OV). Importantly, VO and OV languages have other systematic differences besides their word order. In VO languages, function words (which are highly frequent, and include determiners and prepositions such as *to*, *the*, *on*) tend to come before content words (which are less frequent, for example, nouns such as *banana*, *Paris*), while the opposite is true in OV languages. A final difference between VO and OV languages is their prosody. Elements in VO languages tend to alternate in length (short–long–short–long) while elements in OV languages tend to alternate in pitch (high–low–high–low).

In Gervain and Werker’s (2013) study, infants were first familiarized to a stream of nonsense syllables that contained both frequent and infrequent items. For example, in the stream “…geofibugedefinagedofipe…”, “ge” and “fi” are frequent, but other syllables such as “fo”, “bu”, and “do” are infrequent. There were no pauses between the items, so it was ambiguous as to whether the stream should be parsed as frequent–infrequent (beginning with “ge” or “fi”, typical of VO languages) or infrequent–frequent (ending with “ge” or “fi”, typical of OV languages). However, different infants heard the stream played with different prosody (either short–long or high–low), which could cue infants as to whether they were hearing an OV or VO word order. The infants were then tested using the head-turn preference procedure. On each trial, researchers played infants syllable strings that were characteristic of either the OV or VO parsing strategy and measured their looking time. When bilinguals heard the stream with OV prosody, they parsed the stream in the OV way; they looked longer (or were surprised) when they heard infrequent–frequent strings. Bilinguals showed the opposite pattern of looking when they heard the stream with VO prosody. However, monolingual English infants
(who are learning a VO language) were not able to parse the stream when they heard OV prosody, looking equally to the two types of syllable strings. Even when they heard the stream with no prosody, they defaulted to a VO parsing strategy. Together, these results suggest that bilingual infants separate their languages based on prosody and use this information to form expectations about the word order they are hearing. This perceptual ability could help them acquire languages with different word orders.

**Consonant Perception**

A key achievement of speech perception in infancy is discovering a language’s phonemes, the speech sounds that form the building blocks of words. Each of the world’s languages has a different set of phonemes that are meaningful. For example, English makes a difference between /r/ and /l/, such that rake and lake have different meanings. However, in Japanese, the same sound difference does not change a word’s meaning. Children must learn what speech sound contrasts are phonemic (meaningful) in each of their languages. Because different languages use different phonemes, the consonant contrasts (discussed in this section) and vowel contrasts (discussed in the next section) that are meaningful in any given language pair will overlap to varying degrees.

Let us begin first with infants’ consonant perception, by considering the sounds /b/ and /p/. They are both bilabial consonants, produced with a closure of the lips. In technical terms, the defining difference between these two sounds is their voice onset time (VOT), which is the length of time between when the air behind the lips is released and when the vocal cords start vibrating. Here, /b/ has a short VOT (the vocal cords vibrate immediately, or even before the air is released from behind the lips) while /p/ has a long VOT (the vocal cords vibrate after the air is released). One can imagine a continuum of consonants that spans from /b/ to /p/, each with a slightly different VOT. Perception of such consonants tends to be categorical, in that at a certain point along the continuum, perception suddenly switches from /b/ to /p/. However, the exact location of this boundary differs across languages. This leads to an interesting situation for bilinguals whose languages locate the boundary between /b/ and /p/ in different places. Do bilinguals form one intermediate consonant boundary or do they maintain different consonant boundaries for each language? Three studies using three different methods – looking-time, event-related potentials, and magnetoencephalography – have investigated this question in infants.

Burns et al. (2007) tested French–English bilingual infants’ discrimination of /b/ and /p/ using a looking-time paradigm. Sitting on their parent’s lap, infants were habituated to an intermediate sound, which would be classified as /b/ by English-speaking adults but /p/ by French-speaking adults. As they listened to this repeated sound, their attention declined, as measured by their looking to an unrelated image. Infants then heard two types of test trials: an unambiguous /b/ and an unambiguous /p/. If infants increase their looking at these test trials, this suggests that that they perceive the boundary between the intermediate sound and the test sound. At age 6–8 months, infants’ experience did not matter much: both monolingual English and bilingual English–French infants showed similar responses. By age 10–12 months, monolinguals only perceived the English boundary, but bilinguals perceived both the English and the French boundaries. While monolinguals become specialized to perceive sound contrasts in their native language, bilinguals continue to perceive sound contrasts in both of their languages.
Garcia-Sierra et al. (2011) used similar logic in an event-related potentials (ERP) study that investigated infants’ brain responses to /t/, /d/, and a sound with an intermediate VOT value. This intermediate sound is perceived as /d/ by English speakers but as /t/ by Spanish speakers. Garcia-Sierra et al. used a double oddball paradigm, in that they presented two different oddball sounds, /t/ and /d/, and the intermediate sound served as the standard. At age 6–9 months, bilingual infants’ brains did not seem to detect either type of phonetic change, but by age 10–12 months they did. Monolingual infants’ brains appear to detect differences in native language consonants at a younger age, by 7 months (Rivera-Gaxiola, Silva-Pereyra, and Kuhl 2005). For the bilinguals, language dominance mattered: those infants with the most exposure to English or Spanish showed more changes over time in their ERP response to that language’s consonant.

Convergent results were found in a similar study that tested 11 month old Spanish–English bilinguals and English monolinguals on the same /t/ and /d/ consonants (Ferjan Ramírez et al. 2016). The researchers used a similar method, but this time infants’ brain responses were measured using magnetoencephalography (MEG; see Figure 8.3). While ERP measures electrical potentials on the scalp, MEG measures the magnetic fields produced by the brain’s electrical activity. The researchers measured infants’ brain responses in two time windows: 100–260 ms after the stimulus (early window) and 260–460 ms after the stimulus (late window). Based on previous research, a brain response in the early window is considered less mature than a response in the late window. For /t/ (the English contrast), there were no differences between monolinguals and bilinguals in either time window. However, for /d/ (the Spanish contrast), bilinguals showed a stronger neural response than the monolinguals in both time windows. In other words, infants’ brains responded most to the language(s) they were acquiring.

Both groups were acquiring English and so their responses to English did not differ, but only the bilinguals were acquiring Spanish and so their brains responded more strongly to Spanish. The study also found some evidence that bilinguals’ brains might take more time to become specialized in processing native phonetic contrasts. In the early time window, monolinguals did not show a different response to English and Spanish, while bilinguals showed a stronger response to English. Bilinguals’ sensitivity in the early time window might be related to a less mature, acoustically-based processing of the speech sounds, rather than the more mature phonetically-based processing shown by the monolinguals. Taken together, the authors concluded that encoding the sounds of two languages might take somewhat longer than encoding the sounds of a single language, but emphasized how this extended period of plasticity is adaptive for bilinguals.

Further evidence for bilinguals’ prolonged perceptual sensitivity comes from a study that used a neuroimaging technique called Functional Near Infrared Spectroscopy (fNIRS). In fNIRS, low-power infrared light is shone onto the head. The light is absorbed differently depending on how much oxygen is in the blood flowing through the brain areas underneath the light. Sensors measure the light that remains, so researchers can get an idea of brain activation in response to the different sounds. Petitto et al. (2011) studied monolingual infants learning English and bilingual infants learning English and one of a variety of other languages (e.g., French, Spanish, Chinese). The researchers investigated monolingual and bilingual infants’ brain responses to /b/–/d/, an English contrast that is also meaningful in most other languages, and to /t/–/t/, a Hindi contrast that is not meaningful in any of the languages being learned by the infants. The results were complex, but suggested that bilinguals may maintain sensitivity to non-native phonetic contrasts at a later age than monolinguals.

However, not all studies have found enhanced consonant discrimination abilities in bilinguals. Bosch and Sebastián-Gallés (2003a) tested monolingual Spanish, monolingual Catalan, and bilingual Spanish–Catalan infants’ discrimination of /s/–/z/ in a looking-time study. This consonant contrast is meaningful in Catalan, but not in Spanish. Monolinguals showed a pattern consistent with perceptual narrowing. At 4 months, both Catalan and Spanish monolinguals could discriminate the consonants, displaying a language-universal pattern. However, by 12 months, only Catalan monolinguals could discriminate the consonants. Since bilingual Spanish–Catalan bilinguals are also exposed to this contrast when they hear Catalan, they would be expected to discriminate it throughout development. Surprisingly, 12 month old bilinguals did not show discrimination. This is one of the rare studies of consonant perception that found less sensitivity in bilinguals compared to monolinguals. However, in the next section we will see several studies that have reported this pattern for vowels.

A final type of consonant contrast that has been examined in bilingual infants is one that occurs across their two languages. Sundara, Polka, and Molnar (2008) tested English–French bilingual 10–12 month old infants’ discrimination of two variants of the /d/ sound, using a looking-time habituation method. In English, /d/ is alveolar (produced with the tongue touching the ridge between the palate and the upper teeth), but in French, /d/ is dental (produced with the tongue against the upper teeth). Their results showed that bilinguals could tell the difference between the two /d/ sounds, which could be important as they pull apart their two languages.
In sum, by the end of the first year of life, bilingual infants perceive many of the consonant contrasts that are important in their native languages. There is some evidence that they are better than monolinguals at discriminating some non-native contrasts. How should we understand these differences? One interpretation is that bilinguals are somewhat delayed in their perceptual development. In other words, both monolinguals and bilinguals follow the same pattern of perceptual narrowing, but bilinguals take longer to reach a mature stage of speech perception. Another interpretation is that bilinguals have a perceptual advantage: they maintain enhanced sensitivity to speech sound contrasts as an adaptation to their environments. Deciding between these two alternative interpretations can be addressed with more research, which includes (a) testing other consonant contrasts, (b) directly comparing bilingual infants learning different language pairs, and (c) studying infants across an even wider range of ages.

Vowel Perception

Just as for consonants, different languages use different vowel contrasts and, again, some vowel contrasts are shared across languages while others are not. The first experiment to study bilingual infants’ phonetic perception investigated Spanish–Catalan learners’ perception of the /e/-/ε/ contrast (Bosch and Sebastián-Gallés 2003b). This contrast exists in Catalan, but not in Spanish. In a looking-time study, infants heard a repeated word containing one of the vowels (either [‘deði] or [‘deði]) until they had accumulated two minutes of attention. Next, infants moved to a test phase, where they heard different trials with either the same word as familiarization or the other word. If infants discriminate the two vowel sounds, they are expected to look longer when the word switches than when it remains the same. At age 4 months, monolingual Spanish, monolingual Catalan, and bilingual Spanish–Catalan infants all looked longer when the word switched, showing that they discriminated the two vowels. This fits well with the notion of perceptual narrowing: infants initially discriminated the contrast, whether it was meaningful in their native language or not. However, by 8 months, only the Catalan monolinguals still discriminated the contrast. It was not surprising that the Spanish monolinguals did not discriminate, as this contrast is not meaningful in Spanish: their perception had become attuned to their native language. Surprisingly, 8 month old bilinguals did not discriminate the contrast even though it is meaningful in one of their languages. To follow up on this finding, the researchers tested older bilinguals, age 12 months. This time, the infants did discriminate the contrast. Bilinguals’ developmental pattern can be described as U-shaped (discrimination, failure, then discrimination again). This does not follow the typical pattern of perceptual narrowing shown by monolinguals, who usually either discriminate speech sounds throughout development (for native contrasts) or else discriminate them at a young age but fail at older ages (for non-native contrasts). The authors posited a distributional account, pointing out that the Spanish /e/ actually falls somewhere in between the Catalan /e/ and /ε/. The overlapping distributions of these sounds in bilinguals’ environment may take some time for infants to sort out. Nonetheless, by the age of 1 year, bilingual infants do discriminate this vowel contrast.

Follow-up studies suggested that different experimental tasks might be more sensitive in revealing bilingual infants’ perceptual knowledge. One relatively new task for testing infants’ discrimination of different stimuli is called the anticipatory eye-movement
paradigm. Infants are seated in front of an eye-tracker and taught a simple rule about two types of stimuli. For example, hearing stimulus A predicts that a visual reward will appear on the left side of the screen, while hearing stimulus B predicts that it will appear on the right side of the screen. After familiarization to this rule, infants’ ability to anticipate the reward is measured. In other words, researchers ask whether they look towards the correct side of the screen after hearing the stimulus, even before the reward appears. If they can correctly anticipate the reward, this suggests that they can perceive the difference between stimulus A and stimulus B.

Albareda-Castellot, Pons, and Sebastián-Gallés (2011) once again tested 8 month old Spanish–Catalan bilingual infants’ discrimination of the /e/–/ε/ contrast, but this time using an anticipatory eye-movement paradigm. Infants watched an animated Elmo character disappear into a t-shaped tunnel, eventually emerging on one side or the other. Hearing the /e/ sound predicted Elmo’s reappearance on the right side, while hearing the /ε/ sound predicted Elmo’s reappearance on the left side (or vice versa). This time, both monolingual Catalan and bilingual Catalan–Spanish infants succeeded, which they could only do if they perceived the difference between the two sounds.

These two studies – using two different methods (familiarization versus anticipatory eye-movements) – found very different results. Why would this be? It may be that methods relying on “surprise” (e.g., increasing attention to a switched stimulus such as in habituation studies) are not ideal for measuring speech perception in bilingual infants, particularly those learning Spanish and Catalan. These languages have many cognate words (similar-sounding words with the same historical root) that share consonants, but differ in their vowels. The authors suggest that bilingual infants may use an adaptive strategy of paying more attention to consonants than vowels, and thus may not be particularly surprised by vowel changes, even when they perceive them.

Other studies support the idea that the particular language pair being learned matters for how bilingual infants perceive vowel sounds. Sundara and Scutellaro (2010) tested Spanish–English bilinguals’ discrimination of this same /e/–/ε/ contrast, in a looking-time habituation method similar to that used by Bosch and Sebastián-Gallés (2003b). This time, bilingual infants looked longer to the novel switch trial than the familiar same trial, both at 4 and 8 months. Why did 8 month Spanish–English bilingual infants succeed in the same situation where Spanish–Catalan infants failed? The authors pointed out a key difference between these language pairs: Spanish and Catalan are rhythmically similar, while Spanish and English are rhythmically distinct. As discussed in an earlier section, rhythm is a highly salient cue for young infants as they discriminate their languages. While refining their vowel categories in the first months of life, Spanish–English bilinguals may have used rhythm to sort and categorize the vowel sounds in their two languages. This might have allowed them to perceive a vowel contrast at age 8 months that was more difficult for Catalan–Spanish bilinguals. Another important difference could be that English and Spanish share fewer cognates than Catalan and Spanish, so that English–Spanish infants might react more to vowel changes than Catalan–Spanish infants.

While most studies of infants’ vowel perception have used looking-based methods, some researchers are beginning to use ERPs to measure infants’ brain responses to native and non-native vowel contrasts. These studies are just beginning to document the typical brain responses that bilinguals show to changes in vowel sounds and how
these compare with those that monolinguals show (Shafer, Yu, and Garrido-Nag 2012; Shafer, Yu, and Datta 2011), and will be foundational for future studies in this area.

**Lexical Tone**

In languages like English, pitch is used to convey information such as emotion and emphasis. However, in tonal languages, changing the pitch of a word changes its meaning. By some estimates, up to 70% of the world’s languages are tonal, and these languages differ from each other in the number and type of tones they use (Yip 2002). As an example, Mandarin Chinese has four tones. A syllable such as ‘mā’ (produced with a high tone) means mother, while ‘mǎ’ (produced with a dipping tone) means horse. Children learning a tone language must perceive and encode tone as they learn new words. Conversely, children learning non-tone languages must learn that pitch is irrelevant to the meaning of a word.

Bilingual children could find themselves in several different types of situations with relation to lexical tone. Children might be learning two tone languages, each with a different set of tones to master. Alternately, they might be learning one tone and one non-tone language, in which case they should pay attention to tone in one language but not in the other. Finally, children might be learning two non-tone languages, in which case tone is irrelevant in both languages. The bulk of the research to date on bilingual infants’ tone perception has focused on infants learning two non-tone languages.

Liu and Kager (2016) studied a group of bilingual infants and toddlers learning Dutch and a variety of non-tone languages. They tested infants’ sensitivity to different tone contrasts used in Mandarin Chinese. Like previous studies with monolinguals (Liu and Kager 2014), bilingual infants showed early language-general sensitivity to the tone contrasts, and a decline in sensitivity at around 8–9 months. Eventually, both groups regained some sensitivity to the tone information, which could be related to their growing ability to process other types of pitch information (e.g., emotion, emphasis). However, bilinguals showed a perceptual advantage. Bilinguals regained sensitivity by age 11–12 months, while monolinguals regained sensitivity later, around 17–18 months. The authors raise several possible explanations for this difference. One possibility is that as bilinguals disentangle their two languages, they gain enhanced sensitivity to acoustic information generally or to pitch information more specifically. Another possibility is that the differences in performance come from bilingual cognitive advantages in information encoding, memory, or novelty detection.

**Lexical Stress**

In English, stress is phonemic, meaning that words can change meaning depending on which syllable is stressed. For example, the noun REcord refers to a large plastic disk, whereas the verb reCORD refers to the process of capturing the sound that can be played from the disk. Stressed syllables are louder, higher, and longer than unstressed syllables, and English speakers use these cues to process lexical stress. However, speakers of languages such as French, which does not use lexical stress, often find stress difficult to perceive.

One study examined whether 10 month old monolingual and bilingual infants could discriminate words with different stress patterns (Bijeljac-Babic et al. 2012). Using the
head-turn preference procedure, monolingual and bilingual infants were familiarized to bisyllabic words with either a trochaic pattern (stressed–unstressed, as in REcord) or an iambic pattern (unstressed–stressed, as in reCORD). At test, infants heard new words from each of the two stress patterns. If infants discriminate the two patterns, they are expected to pay more attention when they hear the new pattern than when they hear the old pattern. The monolingual French-learning infants did not discriminate the two stress patterns, which was expected because French does not use lexical stress to contrast word meanings. The bilingual infants were all learning French and one of 15 different languages that does use lexical stress (e.g., Spanish, English, Swedish, Polish, etc.). Bilinguals’ performance depended on their language dominance. Those bilinguals who were dominant in the stress language discriminated the stress patterns, but those who were dominant in French did not. The results suggest that, at least at 10 months, bilingual infants need a certain amount of exposure to lexical stress to be able to perceive it.

Rather than testing infants’ discrimination of different stress patterns, another study tested 6 month old infants’ preference for different stress patterns, also using the head-turn preference paradigm (Bijeljac-Babic, Höhle, and Nazzi 2016). The researchers compared French monolingual, German monolingual, and French–German bilingual infants. Note that German uses lexical stress, while French does not. Infants repeatedly heard the nonsense word “gaba” with either trochaic stress (GAba) or iambic stress (gaBA). Monolinguals showed the classic pattern of perceptual narrowing: German-learning infants preferred the trochaic words (this pattern is most common in German), while French-learning infants showed no particular preference (Höhle et al. 2009). French–German bilingual infants also preferred the trochaic words, showing an identical pattern to the monolingual German-learning infants. Unlike in the previous study, there was no evidence that language dominance played a role. This research suggests that despite less exposure to German than monolingual infants, by age 6 months bilinguals recognize the predominant stress pattern of one of their native languages.

Phonotactics

Beyond learning which types of sound changes are meaningful in their language (phonetics), infants must also learn about which sound combinations can and cannot occur in their languages, called phonotactics. For example, in English the combination “str” is phonotactically legal at the beginnings of words, occurring in words such as “strong” and “strike”. However, the combination “tsr” is illegal as no English words begin with this combination of sounds. Each of the world’s languages has its own rules about which sound combinations are phonotactically legal and which are illegal, and infants must learn these patterns as they acquire their native language(s).

One study to date has looked at the emergence of phonotactic knowledge in bilingual infants. Sebastián-Gallés and Bosch (2002) tested 10 month old Spanish–Catalan bilingual infants using the head-turn preference procedure. Infants heard lists of made-up words that were either phonotactically legal or phonotactically illegal in Catalan, and their looking time was measured. Bilinguals, particularly those with the most exposure to Catalan, looked longer when they heard the phonotactically legal than when they heard the illegal words. This suggested that bilingual infants gain early sensitivity to phonotactic knowledge, but that language dominance might also play a role.
Child L2 Learners

While infants’ phonological knowledge must be inferred indirectly from brain responses and looking time, children can follow instructions, push buttons, and answer verbally. This makes it unsurprising that studies of phonological development in child L2 learners predate similar studies of infants.

Some of the earliest research was conducted by Snow and Hoefnagel-Höhle (1978), who studied native English-speaking children and adults who had recently moved to the Netherlands. The researchers were interested in testing the critical period hypothesis. Theories positing a critical or sensitive period predict that second language acquisition should be fastest and most successful for younger children, particularly those who have not yet reached puberty (see Chapter 1). To investigate the critical period hypothesis, the researchers tested L2 learners aged 3–15 years, as well as adults. They tested many aspects of participants’ L2 knowledge, including grammar, pronunciation, story-telling, and most relevant here, auditory discrimination.

In the auditory discrimination task, learners heard a word and had to point at which of two pictures had been labeled. The names of picture pairs differed in a single consonant or vowel sound, so participants could only succeed if they detected the phonetic difference. Each participant was tested three times, with 4–5 months between each test session. While all groups improved over the course of testing, children aged 12–15 did the best of any group, and performed similarly to native Dutch-speaking children, even during the first testing session—a short 6 months after arriving in the Netherlands. The authors concluded that, contrary to the critical period hypothesis, second language acquisition, including phonetic perception, is most rapid for children age 12–15 years (see also Chapter 1 for a discussion of how language-learning environments and motivation can vary for younger versus older learners). Since this pioneering study, numerous other researchers have continued to examine phonological development in child L2 learners. Using more nuanced techniques and targeted investigation of specific phonetic contrasts, their findings have sometimes supported and sometimes contradicted this early conclusion. In the next sections, we will examine subsequent findings in this area, first discussing consonant perception, then vowel perception, and finally global measures of phonological awareness.

Consonant Perception

How long does it take L2-learning children to become good at perceiving L2 consonants? Some research has suggested that child language learners are quick to achieve mastery. McCarthy et al. (2014) studied a group of English-learning children who had grown up in a Sylheti-speaking community in London (Sylheti is a language from Bangladesh). The English-learners were compared to a group of native English-speaking monolingual children. Children were first tested at around 4.5 years of age, which was about 7 months after they had begun attending nursery school in English, and were retested one year later. The study focused on two consonant pairs: /k/–/g/ and /b/–/p/, which differ in voice onset time. English and Sylheti place these consonant boundaries in different locations, such that certain sounds that would be categorized as /g/ and /b/ in English would be categorized as /k/ and /p/ in Sylheti. Children played a game where they helped a panda learn new words. Across several trials, children heard
a word, and then had to point to a picture representing the word they had heard, for example, choosing between a picture of a coat and a goat or a pea and a bee. Children were presented with words across a range of voice onset time values, so that researchers could pinpoint where they placed their boundaries. Surprisingly, the native English speakers and the L2 learners showed very similar performances overall. Both groups located the consonant boundaries in approximately the same place and became both more consistent and more adult-like in their categorization abilities over time. These results would suggest that children quickly acquire the perceptual skills to process L2 consonants.

However, other research suggests that some L2 consonant contrasts are difficult for children to acquire. A study of French–English bilingual 4 year olds included children who had acquired their languages either simultaneously or sequentially (Sundara, Polka, and Genesee 2006). The bilinguals, as well as French and English monolingual children, were tested on their discrimination of a consonant contrast: /d/–/ð/. While /d/–/ð/ is a meaningful difference in English (e.g. “doze” and “those”), /ð/ does not exist in French and is difficult for French learners of English to master. Children were introduced to a robot who was learning to say different words. The children could help the robot by pressing a button when they heard the robot correctly say a target word (e.g., “those”). Children heard several repetitions of a distractor word (“doze, doze, doze”) and then a single instance of the target (“those”). Their correct and incorrect button presses, as well as missed button presses, were recorded. The results showed that while the English monolingual children readily discriminated the contrast, French monolingual children and bilingual children performed significantly worse, and similarly to each other. Bilingual adults tested in the same procedure showed performance that was intermediate between English monolingual and French monolingual adults. Together, this study suggested that bilingual 4 year olds have more difficulty perceiving some consonant contrasts than their monolingual counterparts, although these differences lessen with age.

Rather than testing learners at a single point in time, other research has looked at how children’s L2 perception changes with experience. One interesting study compared Japanese speaking children’s (aged 6–14) and adults’ (parents of the children) learning of English consonant sounds (Aoyama et al. 2008). Both groups were tested twice: approximately six months after arrival in the United States and approximately 1.5 years after arrival. Most of the adults had taken English courses at school in Japan and so already had some knowledge of English, whereas only one of the children had any experience with English before arrival. The consonants of interest were the sounds /s/ (as in “sink”) and /θ/ (as in “think”). Both English and Japanese have /s/, but /θ/ does not exist in Japanese and so might be hard for Japanese learners of English to master. The participants heard three sounds and had to indicate which of the three was different from the other two. Both the Japanese children and adults showed difficulty with this task compared to native English-speaking adults and children. Further, although they showed improvement with time, Japanese children performed much worse than all the other groups. This again suggests that it may take some time for child L2 learners to accurately perceive some L2 speech sounds.

A different study of Japanese learners of English investigated their perception of the /r/, /l/, and /w/ sounds. As a reminder, Japanese does not have the /l/ sound, but does have an /r/ sound, although it is produced somewhat differently from the English /r/.
Native Japanese listeners tend to perceive both English /l/ and /r/ as either /r/ or /w/. Aoyama et al. (2004) tested perception of these contrasts by native Japanese learners of English, both adults and children. Participants were tested in two separate sessions a year apart, and had lived in the United States for about 6 months at the time of the first test. In a categorization task, participants had to indicate whether three consonants were all the same or whether one was different. On the initial test, the Japanese adults did better than the Japanese children, but the children improved more over time such that this difference disappeared at the second test. Even though adults had some experience studying English in school before moving to the United States, children were quickly able to learn to perceive the English contrasts, albeit not at the same level as native English-speaking children and adults.

In sum, studies of children's L2 consonant perception have revealed several different developmental patterns. In some cases, children seem to acquire non-native contrasts rather quickly, while in other cases they show more difficulty. At the same time, evidence shows that children's perception of L2 consonants improves rapidly, particularly compared to adult L2 learners.

Vowel Perception

Studies of vowel perception in child L2 language learners have converged on an intriguing result: sometimes children accurately produce L2 sounds that they have seem to have difficulty perceiving. For example, Darcy and Krüger (2012) looked at vowel perception and production in children growing up in Germany who spoke Turkish at home. Participants were ages 9–12 and had started learning German in preschool (age 2.5–4 years). Children were tested on their discrimination of several different German vowels that previous work had classified as either difficult or easy for monolingual Turkish adults to discriminate. Children heard three different robots each produce a nonsense word (e.g., “kak”, “kak”, “kek”), and had to decide which robot was saying something different. Compared to a control group of German monolingual children, bilingual children were less able to discriminate the difficult contrasts, but were equally able to discriminate the easy contrasts. This result suggested that despite their early and ongoing exposure to German over many years, they still perceived speech sounds differently than monolinguals.

Intriguingly, when measured on their ability to produce the same vowels in real German words, monolinguals and bilinguals showed highly similar production. Indeed, bilinguals seemed to be able to produce vowel contrasts that they were not able to perceive! How could this be explained? One possibility is that the production task used familiar German words, whereas the perception task used nonsense words (which are unfamiliar). If L2 learners have German categories that are present but less stable than those of monolinguals, they may have more difficulty applying them in the contexts of nonsense words than familiar words. This implies that in everyday speech, these children's vowels can “pass” for native, even though their perceptual categories may not be identical to those of native speakers.

A study of Korean learners of English has also reported differences in how learners perceive versus produce vowels. Tsukada et al. (2005) tested children (age 9–17 years) and adults’ production and perception of several different English vowel pairs, comparing learners who had been in an English-speaking country for either 3 or 5 years.
Bilingual children perceived the vowels more accurately than bilingual adults, but were less accurate than monolingual English-speaking children. Further, children who had been in the country for a longer duration did significantly better than those who had been in the country for a shorter duration, while this had no effect for adults. The authors concluded that the bilingual children were in the process of attuning their English vowels to be native-like, while the adults were not undergoing such development.

The study also revealed an apparent perception–production asymmetry: unlike their performance on the perception task, bilingual children's productions of the same vowels were just as accurate as monolingual children's productions. Does this mean that bilingual children's vowel productions outpace their perceptions? This is only one possible interpretation of the data. As in the previous study, the authors of this study also point out that the production task could have been easier than the perception task.

These first two studies investigated learners who had already been immersed in a new language for several years, but what about learners with less L2 experience? Baker et al. (2008) looked at native Korean-speaking children (aged 7.5–14 years) and adults who had been in the United States for only 6 months. The researchers focused on learners' perception on the /i/-/ɪ/ contrast (as in “beat” and “bit”) as well as the /u/-/ʊ/ contrast (as in “boot” and “book”). In a first task, they played participants different English words containing one of these vowels and had them decide which of 10 different Korean vowels the stimulus vowel was most similar to the one that was played. Participants rated the goodness-of-fit of each token, from 1 (sounded very dissimilar) to 7 (sounded very similar). Compared to adults, children’s categorizations were more varied and their goodness-of-fit ratings were somewhat lower (although not significantly so). This result demonstrates that children and adults perceive L2 sounds differently, at least at the initial stages of learning.

In a second task, researchers played word pairs containing different vowels and participants judged whether they heard the same sound or different sounds. Neither group performed very well at the task, and, surprisingly, children were not more accurate than adults in their judgments. In fact, adults outperformed children on /u/-/ʊ/, although the two groups performed similarly on /i/-/ɪ/.

In a third task, they had the children and adults produce words with these target vowels. This time, the Korean children outperformed the Korean adults when producing the /i/ and /ɪ/ vowel, and did similarly to the adults on the /u/ and /ʊ/ vowels.

Finally, the authors tested whether performance on the three tasks correlated, but no strong correlations were found. However, there was a significant relationship between participants’ age of exposure to English and their performance on both the perception and production tasks. The authors suggested that younger L2 learners rely on different language processing and learning mechanisms. They concluded that it is the age at which L2 is learned, rather than the age at which learners are tested, that is the most important factor in bilinguals’ L2 perception.

Phonological Awareness

Phonological awareness refers to an individual's awareness of the sound structure of words, for example, that words can be broken down into phonemes and syllables. Unlike the research discussed in the previous sections, researchers in this field are primarily interested in links between these skills and early literacy. Many studies have shown that
preschoolers with better phonological awareness become better readers, and improving children’s phonological awareness can improve their reading skills (Ziegler and Goswami 2005). Children also improve their phonological awareness as they learn to read, pointing to a reciprocal relationship between these two skills. Just as strong phonological awareness skills can boost reading, phonological difficulties can have the opposite effect. Problems with encoding and representing phonological information contribute to reading difficulties and disorders such as dyslexia (Ziegler and Goswami 2005).

Researchers use a number of different tasks to study children’s phonological awareness. For example, in deletion tasks, children are asked to produce a word without one of its phonemes, for example, “Say neep. Say it without the /n/”. Such production tasks combine several different skills, including processing the relevant speech information, representing it, encoding it in working memory, manipulating it, and then producing a response. Our interest here is in bilingual children’s speech perception, so we will focus on phonological awareness tasks that involve perception but not production. For example, in a detection task, children might hear three words and have to identify which word does not rhyme with the other two, or has a different onset. In a matching task, children might judge whether two words match in their onset or rhyme. Importantly, while the studies of speech perception that we have discussed so far focused on specific speech sound patterns or contrasts (e.g., /p/–/b/), most studies of phonological awareness studies test children on a broad range of speech sounds and syllables. In this section, we will focus on children’s receptive phonological awareness skills within two themes. First, we will discuss a study that has compared monolingual and bilingual children’s phonological awareness. Second, we will look at research that has investigated links between phonological awareness skills in each of bilingual children’s two languages.

How do monolingual and bilingual children compare in their phonological awareness? There is some evidence for a bilingual advantage. For example, Mumtaz and Humphreys (2001) compared 7–8 year old Urdu–English bilinguals to English monolinguals. They presented children with a series of three words and children had to indicate whether all three words rhymed. The bilinguals performed better than the monolinguals. The authors posited that this was because the bilinguals had all learned to read Urdu before learning to read English. In English, the correspondence between letters and sounds is complex (consider the different pronunciations of “oo” in the words “book”, “door”, and “noon”), while in Urdu it is straightforward. Learning to read Urdu might promote earlier phonological awareness, which bilinguals can then transfer to English. However, in this study there was no monolingual Urdu group and no bilingual group that had learned to read English prior to Urdu, so this conclusion remains tentative.

Beyond simply comparing monolinguals to bilinguals, it is perhaps even more interesting to investigate relationships between phonological awareness in bilinguals’ two languages. A meta-analysis investigated whether bilingual children’s performance on phonological awareness tasks in one language was related to their performance in their other language (Branum-Martin et al. 2012). A total of 38 studies were amalgamated, all investigating bilingual children aged 3–14 years learning English and a variety of other languages (Urdu, Greek, Spanish, French, Arabic, Hebrew, Korean, Cantonese, Mandarin). The meta-analysis included studies that used both perception and production tasks, as the authors argued that these two types of tasks measure the same underlying ability. Nearly every study reported a positive correlation across the two languages; that is, children who performed well in their first language tended to
perform well in their second language. However, the strength of the relationship varied widely across languages, with correlations ranging from 0.38 in Mandarin–English bilinguals to 0.86 in French–English bilinguals (1 indicates a perfect relationship). Why is this? One possibility is that certain language pairs (for example, those that are more closely related) make it easier to transfer phonological skills from one language to the other. At the same time, studies of different language pairs were not always comparable in terms of their experimental tasks, the ages and backgrounds of the children tested, and the types of educational programs they attended. This makes it difficult to know whether it was the language pair or some other differences that mattered. To begin to understand some of these issues, we will explore a few of these studies in-depth.

Gottardo and Mueller (2009) studied 5–6 year old Spanish learners of English. In addition to several production measures, the researchers used an initial phoneme detection task in English (children indicated which of three words began with a different sound) and three phoneme-matching tasks in Spanish (rhyming, initial phoneme matching, and final phoneme matching). They also measured children’s reading abilities to see whether this would be related to their phonological awareness. A significant correlation of 0.43 was found between English and the Spanish phoneme detection tasks and positive but non-significant correlations (0.24) between English and the other Spanish tasks, providing some evidence that phonological awareness in the two languages is related. A larger analysis that included the production measures suggested that phonological awareness in the two languages should be viewed as separate but related skills in terms of how they predict reading comprehension.

Wang, Perfetti, and Liu (2005) tested Mandarin–English second and third grade (about 8 years old) bilingual children growing up in the United States. They had children complete both English and Chinese onset and rhyme-matching tasks. Uniquely, they also included a Mandarin tone-matching task. The children showed near-ceiling performance on the onset and rhyme tasks in both languages, with average scores of between 85 and 95%. However, the tone task was much harder, with an average score of 64%. This suggests that children might have more difficulty perceiving the tone of a word than perceiving its other phonological properties.

Finally, while most studies have found that phonological awareness is related across bilingual children’s two languages, there is some research suggesting that the two languages are not always equal. In a study that tested Punjabi–English 6–7 year olds raised in the United Kingdom, children showed greater phonological awareness in English than in Punjabi for both onset and rhyme-detection tasks (Stuart-Smith and Martin 1999). These results could signal that phonological awareness develops earlier in English than in Punjabi. Another explanation is that children perform better in English because of practice with English at school, for example, with literacy activities that promote phonological awareness in English.

**Summary**

In this chapter, we have explored how bilingual infants and children perceive the speech stream and how this changes with experience. This overview has shown that even very young bilinguals are equipped with perceptual skills that help them acquire two
languages, such as abilities to discriminate both sentences and sounds from their different languages. Child L2 learners rapidly improve their L2 speech perception, although they may lag behind native-speaking children for some years. At times, bilingual children seem to show more native-like speech production than would be predicted by their performance on speech perception tasks. Phonological awareness develops largely in parallel across bilingual children's two languages and bilingual children sometimes show enhanced phonological awareness relative to monolinguals.

References


Tsukada, Kimiko, David Birdsong, Ellen Bialystok, Molly Mack, Hyekyung Sung, and James Flege. 2005. “A developmental study of English vowel production and perception by...


