This chapter is to appear in *The International Handbook of Language Development* (2019); edited by Jessica Horst & Janne Koss von Torkildsen; Publisher: Taylor & Francis, UK.

Speech perception and discrimination: from sounds to words

Caroline Junge[1], Natalie Boll-Avetisyan[2], & Titia Benders[3]

Most children listen to speech as their primary source of communication. Yet which language they learn depends on where they grow up (Cutler, 2012): for instance, babies growing up in the Netherlands will grow up speaking Dutch while those growing up in China might master Mandarin, or Cantonese. All these languages are markedly different: they differ in the repertoire of speech sounds, in the use of suprasegmental information (whether acoustic variations above the level of phonetics/phonemes can differentiate between words, such as word stress and tone), and in phonotactic information (which phoneme combinations are permissible in a language or not). When infants are born, their natural preferences and abilities in speech perception are hardly shaped by their native language. In other words, newborns are considered 'universal listeners' (Kuhl et al., 2008). Yet through repeated exposure to their native language, cross-linguistic differences between infants soon become apparent, which suggests that the first year of life sets the scene for language-specific listening. This chapter makes clear that speech perception is not a trivial task: speech is like a stream of sounds embedded in words combining into phrases, with no pauses that reliably signal where words begin or end (see Figure 1). Fortunately, speech contains and conveys cues to many linguistic elements simultaneously. Speech perception is the process of extracting cues from the speech stream, to recognize the message that a speaker is conveying. This process is further complicated by the fact that all speakers are different and therefore produce the cues slightly differently. In what follows next, we will first describe the input that children are exposed to (Section 15.1), before we turn to how children learn to recognize their native language from other acoustic signals (Section 15.2) and to decompose into meaningful units: into sounds (Section 15.3), into suprasegmental units (Section 15.4), and finally, into words (Section 15.5). In Section 15.6 we discuss the development of speech perception in relation to speech exposure and brain maturation. In our concluding section (Section 15.7) we underscore the relevance of early speech perception skill as crucial for language acquisition.

---

[1] 1 = Utrecht University, 2 = University of Postdam, 3 = Macquarrie University

15.1 What kind of speech do children hear?

The primary source of speech is vocal fold vibration, resulting in voiced sounds with a fundamental frequency, which is perceived as pitch. Speech can also vary in amplitude, perceived as loudness fluctuations, and in the duration of segments and phrases, which may signal speaking rate, among many other things. Pitch, amplitude and duration are all essential cues to suprasegmental structure, such as lexical stress, tone, and intonation (e.g., Fry, 1955; Pierrehumbert, 1980). To exemplify how stressed and unstressed syllables differ: Compare in Figure 1 the syllables in the disyllabic words /ɑlə/ ‘all’ and /əlkaːr/ ‘each other’. The first syllable in /ɑlə/ and the second syllable in /əlkaːr/ carry lexical stress, as signaled by their being higher in pitch, louder, and longer than the other syllable.

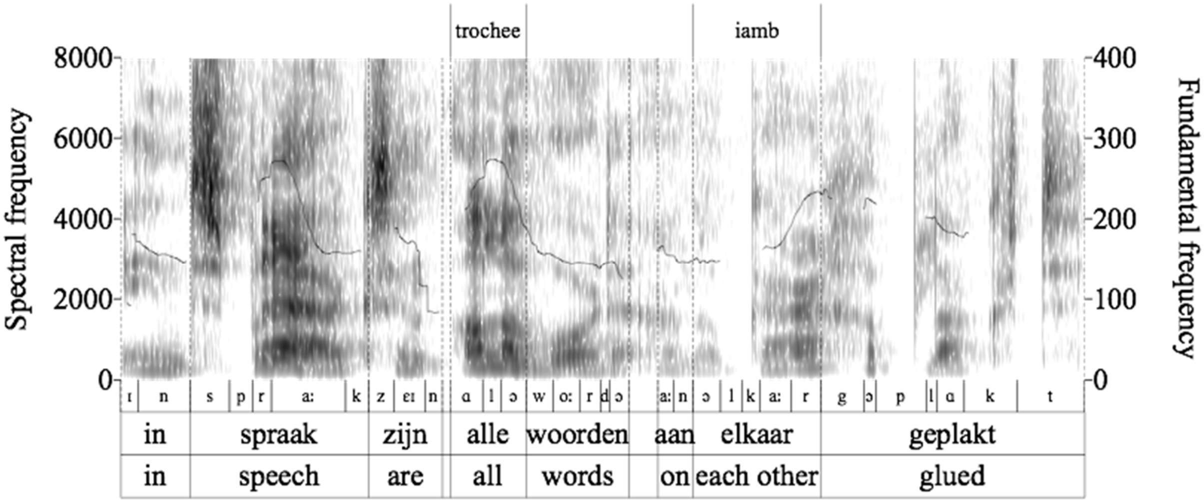

Figure 1: A figure to illustrate that speech is a continuous stream of sounds: a sound spectrogram of a Dutch utterance ‘in speech, all words are glued together’, with time on the x-axis, spectral frequencies from 0-8000 Hz (cues to vowel and consonant properties) on the left y-axis, and fundamental frequency (cue to intonation and word stress, indicated by means of the superimposed rising and falling line in the spectrogram) on the right y-axis. The tiers below the spectrogram show the speech stream segmented into relevant subunits of speech: the top tier displays segments. The second tier groups these segments into Dutch words. Note that pauses in the speech signal the onset of voiceless plosives (/k/, /p/, /t/); they do not align with the onset or offset of words. The third tier offers a translation of the words into English.

Most notably, however, speech consists of segmental information, that is, different consonants and vowels. Suprasegmental cues also play a role in distinguishing between these main segmental classes: vowels are typically voiced, and louder and longer than consonants. Further differentiation within these classes primarily draws from cues such as spectral information attenuated by the shape of the vocal tract (Fant, 1960). More specifically, the resonance in the vocal tract creates formant frequencies for voiced sounds and bands of noise energy in voiceless sounds.

The primary cues to differentiate between vowels are the first and second formant frequencies, which are generally carried by spectral frequencies between 100 and 3000Hz (Peterson & Barney, 1952). Other cues are the consonant-vowel formant transitions and cues such as duration and fundamental frequency (for a review see Rosner & Pickering, 1994). The primary cues to consonants are less easy to pinpoint, as consonants can be either voiced or voiceless and their cues are temporally distributed across the speech stream. For example, to recognize [k] in Dutch /əlkaːr/, a listener needs to recognize a brief moment of silence after the [l], followed by a burst. Despite the lack of formant frequencies in most consonants, consonant identification also relies critically on spectral information, including the spectral energy of the burst—at frequencies that can go up to 10,000Hz—, and the transitions of formant frequencies into the following vowel (see for a survey of cues: Repp, 1982; Wright, 2004)

Children begin "training" to become expert native language perceivers as 24-week-old fetuses, when their auditory system begins functioning (Di Pietro et al., 2013). Most sounds are transmitted to the fetus through the tissue and fluid of the womb, which attenuates frequencies over approximately 500Hz and strongly attenuates frequencies over 1000Hz (Querleu et al., 1988). The suprasegmental cues are thus somewhat audible to the fetus, whereas cues to vowels and consonants are much less audible or not at all (although describing the womb as a low-pass filter may be too simplistic, see Lecanuet et al., 1998). Before birth, infants are thus exposed to the suprasegmental aspects of their native language, as well as to some segmental cues (Querleu et al., 1988).

At birth, infants are for the first time exposed to speech transmitted through air with unattenuated spectral frequencies. In addition, they hear a special register, as adults tend to speak to infants with so-called Infant-Directed Speech (IDS). Suprasegmental cues, in particular fundamental frequency, are strongly modulated in IDS with a high pitch, large pitch range, and possibly a slower speaking rate (see for reviews Ferguson, 1977; Soderstrom, 2007; cf. Martin et al., 2016). Also the segmental contrasts appear to be changed in IDS, although there is debate whether the spectral cues to vowel contrasts are enhanced in IDS compared to adult-directed speech (ADS; e.g., Kuhl et al., 1997; Tang, Rattanasone, Yuen, & Demuth, 2017), remain unchanged (Benders, 2013; Englund & Behne, 2005) or may be reduced (Martin et al., 2015). Similarly, some studies suggest that temporal cues to consonant contrasts are enhanced in IDS (e.g., Malsheen, 1980; Fish, García-Sierra, Ramírez-Esparza, & Kuhl, 2017), whereas others find that these apparent enhancements may be side effects of the IDS prosody (McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013) or entirely absent or reversed (e.g., Baran, Laufer, & Daniloff, 1977; Narayan, 2013).

Regardless of the precise differences between the two registers, we do know that infants treat them differently: infants prefer listening to IDS over ADS (Cooper & Aslin, 1990; Fernald, 1985), possibly because of its emotional valence and/or arousal (Singh, Morgan & Best, 2002). Because of its special properties, IDS has been suggested to be an ideal source for children to discover their native language's prosody, contrastive speech sounds, and words (Liu, Kuhl, & Tsao, 2003; Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989; Ramírez-Esparza, García-Sierra, & Kuhl, 2014; Thiessen, Hill & Saffran, 2005).

15.2 How do children recognize their language?

In order to initiate language acquisition, before everything, the infant needs to be able to separate language in the acoustic environment from non-speech input. This may already be a non-trivial task. Could it be that infants are intrinsically biased to listen to speech? If so, is this bias triggered by specific aspects of speech that facilitate language recognition and help to initiate language acquisition? Which are a language's properties that infants can hold on to recognize speech as language, maybe even as *their* language? Such questions have been focus of studies with newborns, which highlights that neonates already have elaborate speech perception abilities.

To explore whether infants differentiate between speech and non-speech acoustic signals, it is necessary that the type of non-speech signal compared to speech is of similar acoustic complexity, so that it cannot be claimed that speech is simply more interesting. Controlling for acoustic complexity, studies have found that newborns prefer human speech over non-speech analogues, may this be sinusoidal speech (Vouloumanos & Werker, 2007) or music (Dehaene-Lambertz et al., 2010). Such results provide a strong indication that from birth infants are biased to listen to speech. Moreover, the left hemisphere already appears dominant for language processing (see also von Koss Torkildsen, this volume): newborns show greater left hemispheric brain activity when listening to forward speech compared to backward speech (Peña et al., 2003), or compared to other non-speech sounds (emotional voices, monkey calls, scrambled speech; Minagawa-Kawai et al., 2011) – an effect that persists until older age (e.g. Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002).

Infants' initial perception abilities are, probably as a result of prenatal experience (see Section 15.1), even more specialized than discriminating speech from non-speech. It appears that newborns are already sensitive to the rhythm of their language, as they can discriminate between rhythmically distinct languages (Mehler et al., 1988), that is, between stress-timed languages such as English and syllable-timed languages such as French (cf. Cutler, 2012; Chapter 5, this volume). Moreover, they show greater left hemisphere activation when processing their native language compared to an unfamiliar one with a different rhythm (Minagawa-Kawai et al., 2011). Furthermore, when listening to two unfamiliar languages, they prefer the one whose rhythm resembles that of their native language (Mehler et al., 1988). This suggests that infant speech perception initially is focused on prosodic information, as evidenced by similar effects when newborns are exposed to low pass filtered speech (Nazzi, Bertoncini, & Mehler, 1998). Also, infants remain unable to discriminate between languages of comparable speech rhythm until the age of four months (Bosch & Sebastián-Gallés, 1997; Mehler et al., 1988).

As described in Section 15.1, infants' prenatal experience with speech is largely limited to its suprasegmental information, as segmental information is not so well transmitted. Hence, it is likely that prosody is a characteristic of speech that catches infants' special attention from the start (Gleitman & Wanner, 1982). Newborns' preferences for forward over backward speech may also point towards their preference for their native languages' speech rhythm. Forward and backward speech are of comparable physical complexity and share many acoustic characteristics, in particular with regards to its segmental properties (i.e., fricatives, nasals, and vowels sound similar when they are reversed, as their formant structures are relatively symmetrical around their nucleus). Differences between the signals should be most notable in their rhythmic structure (e.g., across languages, speech typically displays a steady decrease in pitch over the envelope of a sentence, and final syllables are lengthened - properties that are reversed in backward speech). Indeed, neonates can only discriminate between two rhythmically distinct languages when speech is played forward but not backward (Mehler et al., 1988; Ramus, Hauser, Miller, Morris, & Mehler, 2000).

However, prosody cannot be the exclusive feature of language that infants are attending to. The type of segmental information also appears to direct attention. For example, sinusoidal speech (as used in Vouloumanos & Werker, 2007) only removes most segmental information but preserves prosodic properties like duration relationships between units, pitch contours, amplitude envelopes; still, neonates show a clear preference for natural speech. Moreover, May, Gervain, Carreiras, and Werker (2017) found that newborns showed greater activation in areas related to language-processing when perceiving two spoken languages (English and Spanish) compared to perceiving a whistled language (Silbo Gomero). Interestingly, the latter language patterns like Spanish regarding its prosodic and rhythmic structure; the main acoustic difference is that Silbo Gomero has acoustically simple whistled tones instead of spectrally complex phonemes.

Future studies should explore in more detail what the properties of language are that neonates are attracted to. By disentangling the role of acoustic complexity, its rhythmic and maybe even segmental characteristics of the native language we will get a better understanding how infants learn to recognize their language from other acoustic signals.

15.3 How do children recognize native speech sounds?

The first step to native-language speech perception is perceptual attunement: learning to discriminate better between two segments that one's native language maps onto two different segment categories than between sounds that the language maps onto a single category (Aslin & Pisoni, 1980). Emerging evidence suggests that newborns respond differently to native than non-native vowels and have thus acquired some aspects of native segment perception prenatally (Moon, Lagercrantz, & Kuhl, 2013). However, the received accounts of perceptual attunement still describe the newborn as a “universal listener” (Kuhl et al., 2008). The universally listening newborn discriminates well between some speech sound contrasts and more poorly between some other contrasts, and exposure to the full speech spectrum as transmitted through air will enable her to perceptually attune to her native language.

The pathways of perceptual attunement differ between sound contrasts, depending on children's (in)ability to discriminate the contrast at birth, and the status of the contrast in their native language. Some consonant contrasts are discriminated well at birth, and that ability is *maintained* if the contrast is phonemic in the native language, and *lost* if it is not phonemic. A seminal example is the contrast between velar [k] and uvular [q], discrimination of which is *maintained* by Salish-learning infants and *lost* by English-learning 10-to-12-month-olds (Werker & Tees, 1984). Other consonant contrasts are not discriminated well at birth, and discrimination is *facilitated* (i.e., shows improvement) if the contrast is phonemic in the children’s language (Kuhl, Conboy, Padden, Nelson & Pruitt, 2005; Tsao, Liu, & Kuhl, 2006), but *remains undeveloped* if the contrast is not contrastive. To illustrate, take the contrast between alveolar [n] and velar [ŋ] in onset position: discrimination is *facilitated* for Filippino 10-to-12 month olds, but *remains undeveloped* in English infants (Narayan, Werker, & Beddor, 2010). Exceptions to perceptual attunement occur if infants maintain full sensitivity to a contrast that is not phonemic in their native language, such as the discrimination between Zulu clicks by infants exposed to English (Best, McRoberts, & Sithole, 1988).

Infants do not only become perceptually attuned to consonant contrasts, but also to vowel contrasts. For example, English infants may *lose* their initial ability to discriminate between [y] and [u] after six months listening experience, which is still present in four-month-olds (Polka & Werker, 1994). An apparent difference between perceptual attunement to consonants and vowels is that full loss may not occur, and all vowel contrasts may remain (to some extent) discriminable by all infants (Tsuji & Cristia, 2014). Discrimination within, rather than between, vowel categories has frequently been studied in the context of perceptual magnet effect, which describes that human adults and six-month-old infants (but not monkeys) have more difficulties detecting changes around "good" vowel examples compared to changes around "poor" vowel examples (Kuhl, 1991). Six-month-old infants may only display this perceptual-magnet effect around the good examples of native, but not non-native vowels (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Although perceptual attunement primarily refers to discrimination between sounds that belong to different vowel categories, whereas the perceptual magnet effect describes gradient discrimination within a vowel category, both describe that infants typically learn not to discriminate between sounds that their language maps unambiguously onto one single category.

What determines the exact attunement pattern for a contrast within a language? The initial state of attunement can be related to acoustic salience: Infants appear better in discriminating between salient than non-salient contrasts (Eilers, Wilson, & Moore, 1977; Narayan et al., 2010). Acoustic salience may also account for the end state of attunement, as it might explain why adults maintain sensitivity to some non-native contrasts (Repp, 1984). The rate of attunement can be related to the frequency of segments in the infants' environment, with contrasts between frequent consonants being acquired before those involving infrequent consonants (Anderson, Morgan, & White, 2003). In addition, it is currently debated whether attunement is primarily driven by exposure to speech, or also shaped by word knowledge (Feldman, Griffiths, Goldwater, & Morgan, 2013; Swingley, 2009; Werker & Curtin, 2005) or social context (Conboy, Brooks, Meltzoff, & Kuhl, 2015; Kuhl, Tsao, & Liu, 2003).

Discrimination between speech sounds has so far been described as a symmetric behaviour, being equally strong from one member of the contrast to the other as vice versa. However, asymmetric discrimination may also exist. For example, infants are able to discriminate a change from labial consonants (such as [p]) to coronals (such as [t]), but not vice versa (Dijkstra & Fikkert, 2011; Tsuji, Mazuka, Cristia, & Fikkert, 2015). As discrimination is better in the direction of the more salient and frequent coronals, saliency and frequency effects may both play a role in the existence or development of asymmetric perceptual behavior. Asymmetrical discrimination also exists between vowel categories, as described by the native referent vowel framework (Polka & Bohn, 2011): Universal listeners, at the outset of language acquisition, are good at discriminating spectral changes in the direction of salient vowel sounds that serve as natural referents, but not in the opposite direction. Later in development, these asymmetries may be attenuated by the category frequencies (Pons, Abareda-Catellot, & Sebastian-Galles, 2012). Ultimately, native speech perception requires the acquisition of symmetrical discrimination between phonemically contrastive speech sounds. Adult speech perception is then characterized by an acute sensitivity to cue changes that signal a native phonemic contrast, and by steep categorical boundaries between the sounds along an acoustic continuum that are perceived as one versus the other category (Liberman, Safford Harris, Hoffman, & Griffith, 1957).

15.4 How do children acquire prosodic knowledge?

Across languages, words are typically characterised not only by their phoneme inventories and their phonotactics, but also by their prosody, that is, by suprasegmental characteristics. There are three different types of prosodic phenomena that characterise words: lexical stress, lexical tone, and lexical pitch accent. The acquisition of language-specific word-level prosody is an important milestone in language development, as it gives cues to speech segmentation and lexical acquisition (see Section 15.5). Regarding prosody acquisition, similar developmental trajectories can be observed as in the acquisition of sounds (see Section 15.3): Usually it is observed that infants are relatively good at discriminating between prosodic patterns at birth, and that ability is *maintained* (and possibly improved) if they are contrastive in the native language, and *lost* if they are not.

Lexical stress. Not all languages have lexical stress (e.g., French and Korean do without), but if they do, lexical stress comes in different forms. Some languages have fixed lexical stress, where acoustic prominence has to fall always on one specific position in the word (e.g. the first syllable in Finnish, the second last syllable in Polish). Fixed stress is assumed to have a demarcative function that serves speech segmentation (Trubetzkoy, 1939/1969). Other languages have variable lexical stress (e.g., German, English, Spanish), where stress does not always occur in the same position of a word. In such languages, stress position can change the meaning (e.g. in German, TEnor means ‘common sense’, while teNOR refers to a singer) or the syntactic class of a word (e.g., in English, /IMport/ is a noun while /imPORT/ is a verb; in Spanish, /FAbrica/ is a noun ‘Factory’ while faBRIca is a verb conjugated in 3rd person present tense). Consequently, in these languages stress has a contrastive function relevant for word recognition.

Several studies have indicated that infants are able to discriminate between different lexical stress patterns from birth (e.g., Sansavini, Bertoncini, & Giovanelli, 1997). Also French-learning infants are able to discriminate between lexical stress patterns at four months (e.g., Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009), which suggests that the discrimination ability for lexical stress patterns begins language-independently. However, by nine months, studies found French-learning infants’ sensitivity to lexical stress to decline, which indicates effects of language-specific reorganization (Bijeljac-Babic, Serres, Höhle, & Nazzi, 2012; Skoruppa et al., 2009).

A characteristic of variable lexical stress language is that prominence can practically occur on any position in a word, but usually some stress patterns are more frequent than others. In Germanic languages like Dutch, most bisyllabic words are trochaic, that is, they have stress on the first syllable, whereas it is much rarer for bisyllabic words to be iambic that is, with stress in the final syllable. (Recall that Figure 1 actually depicts both types of bisyllabic words: the trochaic word /ɑˈlə/ ‘all’ and the iambic word /əlkaːr / ‘each other’). That infants were sensitive to this type of prosodic regularities in their language was revealed by studies who found that infants listened longer to trochaic than to iambic bisyllables, if in their native language the trochee is the predominant stress pattern (German-learning six-month-olds: Höhle et al., 2009; English-learning nine-month-olds: Jusczyk, Cutler, & Redanz, 1993) but show the opposite preference if the iambic pattern is predominant in their language (Hebrew-learning nine-month-olds: Segal & Kishon-Rabin, 2012), and show no preference if their language has no lexical stress (French-learning six-month-olds: Höhle et al., 2009).

In summary, present research suggests that lexical stress is subject to perceptual attunement in infancy, manifested in a decrease of sensitivity to lexical stress contrasts in infants learning a language without lexical stress, and an acquisition of language-specific knowledge of lexical stress. Future research should expand the study of lexical stress perception to infants learning languages with fixed lexical stress.

Lexical tone. Not all languages have lexical tone (e.g. English, German do without), but if they do, its syllables carry contrastive pitch contour characteristics (e.g., rising, falling, flat) that point to different word meanings (e.g., in Mandarin, /ma/ can mean ‘mother’, ‘hemp’, ‘horse’ or ‘scold’ depending on the tone) or different grammatical forms (e.g., in Roermond Limburgian, a Dutch dialect, the word /kni:n/ meaning ‘rabbit’ can be either singular or plural depending on the tone). First evidence suggested that lexical tone acquisition is also subject to perceptual tuning, just like phoneme and prosody acquisition. Mattock and Burnham’s (2006) seminal study demonstrated a decrease in sensitivity to non-native tone contrasts (i.e., contrastive in Thai but not Chinese) by English-learning but not Chinese-learning infants at nine months, while at six months, both groups could discriminate the tone contrasts (for similar findings, see Yeung, Chen, & Werker, 2013).

Newer research on the development of lexical tone sensitivity, however, suggests a more complex picture. It turns out that some tone contrasts are relatively salient and can be perceived even by infants for whom the contrast is non-native (e.g. Liu & Kager, 2014; Ramachers, Brouwer, & Fikkert, 2018), while, on the other hand, the discrimination of less salient contrasts can remain effortful even for older (up to 24-month-old) infants for whom the contrast is native (Cheng & Lee, 2018; Fan, Li, & Chen, 2018). On top of this, there are cases in which infants’ perception of non-native contrasts can even *increase* with infants’ growing age (Chen & Kager, 2016; Chen, Stevens, & Kager, 2017), which possibly stems from infants’ need to be sensitive to tonal contrasts at other levels than the segmental level (i.e., to detect intonation at the phrase or sentence level). Future research will need to disentangle the contribution of acoustic salience, language experience, and the relevance of tone at the sentence level in infants’ perception of lexical tone.

Lexical pitch. Not all languages have lexical pitch accent (e.g., French, English go without), but in languages that do have it (e.g., Japanese), the presence or absence of a pitch accent (marked by a steep pitch movement, e.g., a pitch fall in Japanese) and its location can change the meaning of a word (e.g., in Japanese, /HAshi/ with falling contour and pitch accent on /HA/ means ‘chopsticks’, whereas /haSHI/ with rising contour and no pitch accent means ‘bridge’). Currently, hardly anything is known about the acquisition of lexical pitch accent. We know that newborns discriminate between lexical pitch accents (Nazzi, Floccia, & Bertoncini, 1998). One further study (Sato, Sogabe, & Mazuka, 2010) has investigated the perception of lexical pitch accent in infants acquiring Japanese, with the expected result that both four- and 10-month-olds discriminate between pitch accents. Evidently, more research is needed that includes infants acquiring languages without lexical pitch accent.

15.5 How do children recognize words in speech?

Most of the words that infants hear occur within continuous speech (e.g., van de Weijer, 1998), while - as we have seen in Figure 1- continuous speech does not provide clear silences between words to signal word boundaries. Although the evidence that input predominantly comprises multiword utterances comes from Germanic languages, the general assumption is that infants of any language are confronted with the *speech segmentation problem*: in order to learn words from the input, infants are required to segment the speech stream into word-like units. Yet languages differ widely in what they consider word-like units. In some languages like Mandarin, all words are morphologically simple, while in agglutinative languages like Turkish, words are composed more complexly (i.e., structured around a word stem plus multiple possible suffixes). Therefore, infants should have formed a notion about typical word patterns in their native language before they can apply this knowledge to start extracting words from running speech. This presents us with a riddle, given that most of these words are presented in multi-word utterances: how then do infants learn what words are?

There are two likely scenarios. First, it is possible that infants initially zoom in on those parts in the input that robustly highlight the typical shapes of words: that is, isolated words, and, to a lesser extent, words at the edges of utterances. This is the top-down view. Although words in isolation surely do not occur frequently enough to be the sole explanation for vocabulary construction, they might occur sufficiently to build a rudimentary idea of how words pattern (Lew-Williams, Pelucchi, & Saffran, 2011). Words at phrase boundaries also prove informative, since here silence demarcates at least one of the word boundaries (Seidl & Johnson, 2006). Evidence that isolated words signal structure comes from experiments that show that the type of regularities within a set of isolated words of an artificial language subsequently shapes how infants segment a stream of words from that language (Thiessen & Erickson, 2012; Thiessen, Kronstein, & Hufnagle, 2013). Recognising how words pattern could also be acquired through sensitivity to the distributional structure of the speech input (see also Cannistraci and colleagues, this volume). This bottom-up view assumes that infants can track which combinations of elements (e.g. sounds, syllables) often occur together, which indicates that these combinations are likely to be part of the same word. For instance, while the syllable 'fol' is often followed by the syllable 'low' as in 'follow', or by 'ly' as in 'folly', a syllable like /is/ can occur with so many possible syllables that it shows it is not part of a word but stands alone. This view is first put forward by Saffran, Aslin and Newport (1996) as a way to find multi-syllabic words in speech. This seminal study demonstrated that when infants were presented with an artificial language composed of four trisyllabic words, they grouped syllables together based on which combinations always appeared successively. Of course, this requires sensitivity to the unit of syllables, which infants appear to be by at least two months (Bertoncini & Mehler, 1981).

Besides sensitivity to statistical regularities, infants also appear to be born with certain universal linguistic biases that further helps them discern which combinations are permissible in their native language (Berent, 2013). As said in the beginning, languages differ widely in their phonotactics. Most languages allow syllables to begin with consonant clusters, although certain clusters are more preferred to others (e.g., /bl/ to /lb/; the sonority sequence principle, Clements, 1990). Other languages are stricter: Korean lacks any type of onset consonant clusters. Yet despite the absence of consonant combinations, Korean adult speakers show the same kind of biases when judging non-words with consonant clusters: they make less errors for consonant combinations that frequently occur in other languages than for those that are infrequent (Berent, Lennertz, Jun, Moreno, & Smolensky, 2008). Strikingly, even newborns show this phonological bias for onset clusters (Gómez et al., 2014).

At the level of prosody, the iambic-trochaic law is another example of a universal prosodic grouping bias (e.g., Hayes, 1995; Nespor et al., 2008). Recall that lexical stress is usually defined in three aspects: pitch, intensity and duration. This law predicts uniformity in when and why humans group pairs of weak and strong syllables into a certain pattern, either as weak-strong (iambic) or as strong-weak (trochaic). When the more dominant cue is duration, this law predicts an iambic bias, whereas dominance in pitch or intensity leads to a trochaic bias. Infants also display this grouping bias (Bion, Benavides-Varela, & Nespor, 2011), regardless of whether lexical stress in their native language is fixed or variable (Abboub, Boll-Avetisyan, Bhatara, Höhle, & Nazzi, 2016). Again, this bias appears already present at birth (Abboub, Nazzi, & Gervain, 2016).

By examining parallels in the phonotactic rules of various natural languages, recent research is now revealing the existence of other phonotactic biases, and examining whether these are universal or emerge over development. Take for instance the labial-coronal bias (e.g. MacNeilage, Davis, Kinney, & Matyear, 1999): most languages (but not Japanese) prefer syllables with labials in the onset and coronals in coda-position rather than the reverse (e.g., /bat/ vs. /tab/). This bias appears universal as even Japanese infants at first display the same preference as French infants; only by 13 months is this preference reversed, and starts to mirror the phonotactic patterns of their native language (Gonzalez-Gomez, Hayashi, Tsuji, Mazuka, & Nazzi, 2014). Clearly, infants come into the world equipped with certain prosodic and phonotactic biases that facilitates the building of word templates. Once infants have stored some possible words, the next step would then be to generalise across multiple possible words to learn about the permissible sound patterns in this language (that is, to acquire language-specific phonotactic and prosodic regularities).

Computational modelling studies offer additional support for the claim that infants can acquire knowledge on typical word patterns from bottom-up information. Computational models allow one to test which words computational learners find permissible, while directly controlling the type of input they are fed and the underlying learning mechanisms they have. By simply tracking the co-occurrences between syllables coupled with a sensitivity to word stress, computational learners that receive input from real speech corpora (e.g., Dutch, American-English) can indeed acquire phonotactic and prosodic regularities of these languages (Brent & Cartwright, 1996; Swingley 2005). These computational results mirror experimental findings that show that American-English infants initially expect stressed syllables always to signal word onsets (e.g., Jusczyk, Houston, & Newsome, 1999) and can use phonotactic probabilities for speech segmentation (e.g. Mattys & Jusczyk 2001a). Of course, both views (top down and bottom up) do not need to exclude each other. It is probable that infants employ both approaches to advance their lexical acquisition.

This sets the stage for the next question: how do infants extract words from speech? The speech stream is usually full of cues to signal word onset, from top-down cues, i.e. contextual cues and the lexicon up to more bottom-up cues like word stress and phonotactics. Although most cues often work in concord with each other, there are plenty of ambiguous cases in which cues are conflicting. For example, the segment /s/ in the Dutch segment sequence /eːnspeːr/ can either belong to the preceding word /eːns/, followed by /peːr/ (meaning ‘once pear’) or to the upcoming word /speːr/ (with /eːn/ /speːr/ meaning ‘one spear’).

There is not one cue that appears to be foolproof in signaling word boundaries (Cutler, 2012). Research on adults has shown that cues are valued hierarchically, with adults favoring top-down cues over bottom-up, supra-segmental cues (Mattys, White, & Melhorn, 2005). But does this hierarchy also hold for beginning infant learners? It is likely that this is not the case, but that the hierarchy of cues develops over time. First, top-down lexical cues are expected to play a minor role for a long time; after all, infants’ lexicons are still too small to make lexical cues very viable, even though there is evidence that infants as young as six months can use offsets of frequent words such as ‘mommy’ to learn new words (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005). Instead, it is more likely that infants begin mastering their speech segmentation skill by relying predominantly on bottom-up cues. Yet even at the level of bottom-up cues, there are multiple cues that shift in dominance over time. To illustrate, while one of the first dominant cues that infants can use is the language-general cue of the transitional probability between syllables (Kudo, Nonaka, Mizuno, Mizuno, & Okanoya, 2011; Thiessen & Saffran 2003; cf. Chapter 8, this volume), it appears that by eight months, infants increasingly rely on more language-specific cues such as word stress (e.g., Johnson & Jusczyk, 2001). By nine months, infants further show sensitivity to other segmental cues for speech segmentation in their native language: phonotactics (e.g., Mattys & Jusczyk, 2001a; Gonzalez-Gomez & Nazzi, 2013), coarticulations and allophones (e.g., Johnson & Jusczyk, 2001; Mattys & Jusczyk, 2001b).

Although the studies summarised above allows us to roughly sketch the order and ranking of speech segmentation cues, note that there are still crucial pieces missing. First, most studies focus on one language (American-English); we do not know whether this shift from language-general to language-specific cues is also present in languages that differ widely from English (for similar discussions, see Johnson, 2012; Junge, 2017). Second, most studies present only positive evidence that infants can use a certain cue from a certain age, but do not test infants of younger ages (but see Jusczyk & Aslin, 1995; Jusczyk et al., 1999). Thus we do not know whether a language-specific cue such as phonotactics can only be implemented once one has acquired sensitivity to word stress.

When infants segment novel words from speech, they appear sophisticated because they can immediately treat these words as possible word-candidates for novel meanings (Graf Estes, Evans, Alibali, & Saffran, 2007; Saffran, 2001). Nevertheless, infant word recognition skill remains rather fragile: infants often fail to recognize words across different speakers (Houston & Jusczyk, 2000; but see van Heugten & Johnson, 2012), across different dialects (e.g., Best, Tyler, Gooding, Orlando, & Quann, 2009; Schmale, Cristià, Seidl, & Johnson, 2010), or across different contexts (Barker & Newman, 2004; Singh, Morgan, & White, 2004). Infants also need to hear the words multiple times before they start recognising them embedded in continuous speech; for most infants, one exposure appears insufficient (Junge, Kooijman, Hagoort, & Cutler, 2012; Junge, Cutler, & Hagoort, 2014). In time, infants will gradually learn to cope with these indexical types of speech variation, presumably through increased exposure to different types of speech. Indeed, studies demonstrate that infant word recognition skill becomes more robust when infants listen to more varied speech (e.g., different speakers: Rost & McMurray, 2009; or different types of affect: Singh, 2008). All in all, through additional exposure to speech infants become more proficient in their ability to recognize words: they learn when (and when not) variation in segmental or suprasegmental levels of speech highlights changes in word meaning. This suggests that mere exposure to speech is sufficient to bootstrap language from speech, but is this all infants require? The next paragraph discusses the interplay between speech input and brain maturation.

15.6 Speech input and brain maturation.

Throughout this chapter we have seen several examples of how speech perception within the first year of life becomes more and more language-specific, that is, shaped by the listener's native language. Clearly speech perception changes and improves as a function of time. Although it will take years before speech perception is considered mature, it appears that these changes are most pronounced in the second semester of the first year. This led researchers to conclude that the first year of life forms a critical (or sensitive) period for speech perception (Kuhl et al., 2005; Werker & Hensch, 2015).

Critical periods (CP) are characterised as a window in time in which experience has a profound effect on development. For speech perception the critical period seems to fall in the first year. One key distinction of critical periods is that they appear plastic: their timing (that is, when they begin or end) can shift depending on changes in brain maturation or in exposure (Werker & Hensch, 2015). But what triggers the onset of this period for speech perception? Is it brain maturation, or accumulated exposure to input that drives these changes, or a combination of both? Some insights are provided when examining infants with various types of atypical development.

Take the case of healthy preterm infants (born more than 4-6 weeks premature), who - compared to full-term infants – listen to comparable amounts of speech postnatally while their brains are still immature. If the critical period hinges solely on brain maturation, one would expect that the critical period shifts as a function of maturational (corrected) age rather than by postnatal age. Here, studies on preterms show mixed results, depending on which aspect of language-specific speech perception is examined (Gonzalez-Gomez, 2017). Preterms appear to be delayed in their perceptual attunement when it concerns their sensitivity to language-specific prosody, e.g. to rhythm required for language discrimination (Herold, Höhle, Walch, Weber, & Obladen, 2008; Peña, Pittaluga, & Mehler, 2010; but see Bosch, 2011) or to lexical pitch (Gonzalez-Gomez, 2017). Next, delays have also been reported for their phonological development (Peña, Werker, & Dehaene-Lambertz, 2012, but see Gonzalez-Gomez, 2017) and for word segmentation abilities (Bosch, 2011). However, preterms resemble chronologically age-matched controls in when their sensitivity to phonotactics appears (Gonzalez-Gomez & Nazzi, 2012; Gonzalez-Gomez, 2017). All in all, it seems that brain maturation greatly shapes the timing of the CP: if the brains are immature, the timing of the CP will also be delayed. But how relevant is the amount of input?

To test how (lack of) exposure contributes to the timing of the CP, we can turn to children who are naturally deprived from any speech, because they are born deaf from hearing parents. How does the early deprivation of speech affect their spoken language development? A growing literature suggests that the timing is critical of when exposure to speech starts, with the later children have their hearing reinstated (i.e. after their first birthday), the less likely they are to catch up with their hearing peers, with delays particularly manifested in the domains of vocabulary and syntax (Caselli, Rinaldi, Varuzza, Giuliani, & Burdo, 2012). Similar delays in syntax are also present in another group of children exposed to their (new) native language relatively late: children who were internationally adopted at a mean age 16 months (Gauthier & Genesee, 2011; although initial knowledge of their first language phonology might unconsciously be preserved; Choi, Boersma, & Cutler, 2017). Some researchers suggest that these higher-level language delays can be ultimately linked back to the suboptimal critical period for speech perception, with negative cascading effects for subsequent stages (Werker & Hensch, 2015). Conversely, infants who are delayed in their speech exposure by less than 12 months reach similar milestones at the same speed (calculated from the moment they can hear) and in the same order as typically-developing children (e.g., Dettman et al., 2016; Nicholas & Geers, 2007; cf. Levine, Strother-Garcia, Golinkoff, & Hirsh-Pasek, 2016). For example, research reveals that within six months of speech exposure, young infants with cochlear implants show similar phonological discrimination skills and rhythm biases as typically-developing children (e.g., Vavatzanidis, Mürbe, Friederici, & Hahne, 2015, 2016).

Another way to test the contribution of language exposure is by comparing speech perception skills in infants raised in monolingual vs. those in bilingual households, as bilinguals receive less amount of language input than their monolingual counterparts do, and are therefore hypothesized to commit to their native language at a slower pace. Although it falls outside the scope of this chapter to describe in length the differences between monolinguals and bilinguals (cf. Chapter 14, this volume), the general consensus is that bilinguals go through similar milestones, albeit sometimes slightly delayed, compared to their monolingual peers (cf. Werker, Byers-Heinlein, & Fennell, 2009). The pace of language acquisition in the dominant language is further mediated by the relative amount of exposure in that language: the more dominant a language is, the more the acquisition of this language resembles the pace of acquisition in monolinguals (Hoff et al., 2012).

Together, such findings underscore that brain maturation as well as exposure to speech are fundamental for acquiring the skills to decode native speech.

15.7 Concluding remarks and future directions

This chapter highlights that perceiving structure in speech is crucial for language development. The foundations of speech perception appear firmly in place by the end of the first year in life. In this year, children become gradually sensitive to different aspects of phonological development, with the influence of their native language starting to emerge over time. It is likely that this phonological and prosodic knowledge further aids future language development, such as vocabulary acquisition or even the acquisition of language-specific syntax. This latter hypothesis is known as the *prosodic bootstrapping hypothesis*, first proposed by Gleitman and Wanner (1982).

There are indeed encouraging signs that infant speech perception skill is fundamental for subsequent language development. Research with children with a genetic risk of language impairments, either because developmental language delays or dyslexia runs in their families, have also been shown to have atypical speech perception skills (e.g., van Alphen et al., 2004). Such atypicalities in speech perception are even noticeable in infancy, long before these children can be officially diagnosed as having language problems (e.g., Friedrich, Weber, & Friederici, 2004; Leppänen et al., 2002; see also Section IV, this volume). Moreover, a recent meta-analysis reveals that for typically-developing children, early speech perception connects to subsequent language growth at the individual level, for a variety of infant speech perception tasks that probes sensitivity to sounds, to words, or to prosodic knowledge (Cristia, Seidl, Junge, Soderstrom, & Hagoort, 2014). For instance, one study shows that early sensitivity to speech contrasts have different consequences depending on whether these contrasts are meaningful in the native language: whereas sensitivity to native contrasts yield positive correlations with lexical development, sensitivity to non-native contrasts in contrast appears to slow down lexical acquisition (Kuhl et al., 2005). Several studies on speech segmentation also reveal positive links with vocabulary growth (Newman, Ratner, Jusczyk, Jusczyk, & Dow, 2006; cf. Junge & Cutler, 2014).

Interestingly, not all advances in speech perception in the first year prove equally predictive; language discrimination, for instance, does not predict subsequent language performances (Newman et al., 2006). It remains unclear why some aspects of early speech perception skill prove more far-reaching than others. Is it simply that the critical period of speech perception actually marks a series of critical windows on aspects of phonological development (Werker & Hensch, 2015), and that once the CP for language discrimination has been closed, sensitivity to this ceases to have an effect? Or could it be that the unit of a non-native language is less informative for a child than the units of sounds, words or prosodic phrases within one's language? Moreover, note that most prospective studies focus on vocabulary size (but see e.g. Höhle, Pauen, Hesse, & Weissenborn, 2014), but it remains unclear whether these links stretch to the acquisition of syntax, as suggested by the prosodic bootstrapping hypothesis. Future studies should examine whether individual differences in speech perception skill in infancy remain stable across childhood and predictive of language competence that goes beyond the lexicon.

Finally, while we focus predominantly on the emergence of speech perception skills in the first year, it is important to realize that the fine-grained aspects of speech perception are still not adult-like when children are 10 to 12 years of age. Compared to adults (cf., Diehl, Lotto, & Holt, 2004), children for instance discriminate less well between the small acoustic differences that cue phonemic contrasts (e.g., Elliott, Longinotti, Meyer, Raz, & Zucker, 1981; Hazan & Berrett, 2000). It is possible that further sophistication of speech perception requires full maturation of the auditory pathways, (Elliott et al., 1981; cf. Nittrouer, 1996), which continues until children are about 10 years of age (Ponton, Eggermont, Coupland & Winkelaar, 1992). Alternatively, children might require extended exposure to the ambient language to show adult-like perception (Nittrouer, Manning & Meyer, 1993). A full description of children's sophistication in their speech perception skill is beyond the scope of the present chapter. However, while fine-tuning of speech perception spans the entire childhood period, we believe it is important to stress that the early and remarkably fast attunement to the native language starts in infancy.

References

Abboub, N., Boll-Avetisyan, N., Bhatara, A., Höhle, B., & Nazzi, T. (2016). An exploration of rhythmic grouping of speech sequences by French-and German-learning infants. *Frontiers in human neuroscience, 10*, 292. DOI: 10.3389/fnhum.2016.00292

Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. *Brain and language, 162,* 46-59. DOI: 10.1016/j.bandl.2016.08.002

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech, 46,* 155–182. DOI: 10.1177/00238309030460020601

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception and recognition. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology, Vol. 2: Perception* (pp. 67–96). New York: Academic Press.

Baran, J. A., Laufer, M. Z., & Daniloff, R. (1977). Phonological contrastivity in conversation: a comparative study of voice onset time. *Journal of Phonetics, 5*, 339–350.

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*, B45-B53. DOI: 10.1016/j.cognition.2004.06.001

Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development, 36,* 847–862. DOI: 10.1016/j.infbeh.2013.09.001

Berent, I. (2013). The phonological mind. *Trends in cognitive sciences, 17,* 319-327. DOI: 10.1016/j.tics.2013.05.004

Berent, I., Lennertz, T., Jun, J., Moreno, M. A., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences*, *105*, 5321-5325. DOI: 10.1073/pnas.0801469105

Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior and Development*, *4*, 247-260. DOI: 10.1016/S0163-6383(81)80027-6

Best, C.T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants. *Journal of Experimental Psychology: Human Perception and Performance, 14(3),* 345–360. DOI: 10.1037/0096-1523.14.3.345

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native-and Jamaican-accented words. *Psychological Science, 20*, 539-542. DOI: 10.1111/j.1467-9280.2009.02327.x

Bijeljac-Babic, R., Serres, J., Höhle, B., & Nazzi, T. (2012). Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PLOS ONE, 7*, 1-8. DOI: 10.1371/journal.pone.0030843

Bion, R. A., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech, 54,* 123-140. DOI: 10.1177/0023830910388018

Bortfeld, H., Morgan, J.L., Golinkoff, R.M. & Rathbun, K. (2005). Mommy and me familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*, 298-304. DOI: 10.1111/j.0956-7976.2005.01531.x

Bosch, L. (2011). Precursors to language in preterm infants: speech perception abilities in the first year of life. *Progress in brain research, 189,* 239-257. DOI: 10.1016/B978-0-444-53884-0.00028-2

Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition, 65*, 33-69. DOI: 10.1016/S0010-0277(97)00040-1

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61,* 93-125. DOI: 10.1016/S0010-0277(96)00719-6

Caselli, M. C., Rinaldi, P., Varuzza, C., Giuliani, A., & Burdo, S. (2012). Cochlear implant in the second year of life: Lexical and grammatical outcomes. *Journal of Speech, Language, and Hearing Research*, *55*, 382-394. DOI:10.1044/1092-4388(2011/10-0248)

Chen, A., & Kager, R. (2016). Discrimination of lexical tones in the first year of life. *Infant and Child Development, 25*, 426-439. DOI: 10.1002/icd.1944

Chen, A., Stevens, C. J. & Kager, R. (2017). Pitch perception in the first year of life: a comparison of lexical tones and musical pitch. *Frontiers in Psychology 8:*297. DOI: 10.3389/fpsyg.2017.00297

Cheng, Y. Y., & Lee C. Y. (2018). The development of mismatch responses to mandarin lexical tone in 12-to 24-Month-Old Infants. *Frontiers in Psychology 9:*448. DOI: 10.3389/fpsyg.2018.00448

Choi, J., Broersma, M., & Cutler, A. (2017). Early phonology revealed by international adoptees' birth language retention. *Proceedings of the National Academy of Sciences*, *114*, 7307-7312. DOI: 10.1073/pnas.1706405114

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M.E. Beckman (Eds), *Papers in laboratory phonology I: Between the grammar and the physics of speech* (pp 283 – 333). New York: Cambridge University Press.

Conboy, B. T., Brooks, R., Meltzoff, A. N., & Kuhl, P. K. (2015). Social interaction in infants' learning of second-language phonetics: An exploration of brain–behavior relations. *Developmental Neuropsychology, 40*, 216-229. DOI: 10.1080/87565641.2015.1014487

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584-1595. DOI: 10.1111/j.1467-8624.1990.tb02885.x

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child development*, *85*, 1330-1345. DOI: 10.1111/cdev.12193

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words.* Cambridge, MA: Mit Press.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, *298*, 2013-2015. DOI: 10.1126/science.1077066

Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., & Dehaene, S. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language*, *114*, 53-65. DOI: 10.1016/j.bandl.2009.09.003

Dettman, S. J., Dowell, R. C., Choo, D., Arnott, W., Abrahams, Y., Davis, A., .. & Briggs, R. J. (2016). Long-term communication outcomes for children receiving cochlear implants younger than 12 months: A multicenter study. *Otology & Neurotology, 37*(2), e82-e95. DOI: 10.1097/MAO.0000000000000915

Diehl, R., Lotto, A., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology, 55,* 149-179. DOI: 10.1146/annurev.psych.55.090902.142028

Dijkstra, N., & Fikkert, P. (2011). Universal Constraints on the Discrimination of Place of Articulation? Asymmetries in the Discrimination of 'paan' and 'taan' by 6-Month-Old Dutch Infants. In N. Danis, K. Mesh, & H. Sungh (Eds.), *Proceedings of the 35th Annual*

Boston University Conference on Language Development (pp. 170–182). Somerville, MA: Cascadilla Press.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental Changes in Speech Discrimination in Infants. *Journal of Speech and Hearing Research, 20*, 766–780. DOI:10.1044/jshr.2004.766

Elliott, L. L., Longinotti, C., Meyer, D., Raz, I., & Zucker, K. (1981). Developmental differences in identifying and discriminating CV syllables. *The Journal of the Acoustical Society of America, 70,* 669–677. DOI: 10.1121/1.386929

Englund, K., & Behne, D. (2005). Infant directed speech in natural interaction - Norwegian vowel: Quantity and quality. *Journal of Psycholinguistic Research, 34,* 259–280. DOI: 10.1007/s10936-005-3640-7

Fan, S., Li, A. & Chen, A. (2018). Perception of lexical neutral tone among adults and infants. *Frontiers in Psychology 9:*322. DOI: 10.3389/fpsyg.2018.00322

Fant, G. (1960). *Acoustic Theory of Speech Production.* The Hague: Mouton.

Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review, 120*, 751-778. DOI: 10.1037/a0034245

Ferguson, C. A. (1977). Baby Talk as a Simplified Register. In C. Snow & C. A. Ferguson (Eds.), *Talking to Children* (pp. 219 -235). Cambridge, MA: Cambridge University Press.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8, 181–195*. DOI: 10.1016/S0163-6383(85)80005-9

Fish, M. S., García-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2017). Infant-directed speech in English and Spanish: Assessments of monolingual and bilingual caregiver VOT. *Journal of Phonetics, 63*, 19–34. DOI: 10.1016/j.wocn.2017.04.003

Friedrich, M., Weber, C., & Friederici, A. D. (2004). Electrophysiological evidence for delayed mismatch response in infants at-risk for specific language impairment. *Psychophysiology, 41*, 772-782. DOI: 10.1111/j.1469-8986.2004.00202.x

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. The *Journal of the Acoustical Society of America, 27,* 765–768. DOI: 10.1121/1.1908022

Gauthier, K., & Genesee, F. (2011). Language development in internationally adopted children: A special case of early second language learning. *Child Development*, *82*(3), 887-901. DOI: 10.1111/j.1467-8624.2011.01578.x

Gleitman, L. R., & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3-48). Cambridge, MA: Cambridge University Press.

Gómez, D. M., Berent, I., Benavides-Varela, S., Bion, R. A., Cattarossi, L., Nespor, M., & Mehler, J. (2014). Language universals at birth. *Proceedings of the National Academy of Sciences*, *111*, 5837-5841. DOI: 10.1073/pnas.1318261111

Gonzalez-Gomez, N. (2017). Early language development under difficult circumstances: Exploring maturational and environmental factors. Talk presented at the *Many Paths to Language Acquisition workshop.* Nijmegen: October 5, 2017.

Gonzalez-Gomez, N., Hayashi, A., Tsuji, S., Mazuka, R., & Nazzi, T. (2014). The role of the input on the development of the LC bias: A crosslinguistic comparison. *Cognition, 132,* 301-311. DOI: 10.1016/j.cognition.2014.04.004

Gonzalez-Gomez, N., & Nazzi, T. (2012). Phonotactic acquisition in healthy preterm infants. *Developmental Science, 15,* 885-894. DOI: 10.1111/j.1467-7687.2012.01186.x

Gonzalez-Gomez, N., & Nazzi, T. (2013). Effects of prior phonotactic knowledge on infant word segmentation: the case of nonadjacent dependencies. *Journal of Speech, Language, and Hearing Research, 56,* 840-849. DOI:10.1044/1092-4388(2012/12-0138)

Graf Estes, K., Evans, J.L., Alibali, M.W., & Saffran, J.R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*, 254-260. DOI: 10.1111/j.1467-9280.2007.01885.x

Hayes, B. (1995). Metrical stress theory: Principles and case studies. In *Proceedings of the eleventh annual meeting of the Berkeley linguistics II* (pp. 429-446). Chicago: University of Chicago Press.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *Journal of Phonetics, 28,* 377–396. DOI: 10.1006/jpho.2000.0121

Herold, B., Höhle, B., Walch, E., Weber, T., & Obladen, M. (2008). Impaired word stress pattern discrimination in very-low-birthweight infants during the first 6 months of life.

Developmental Medicine & Child Neurology, *50*, 678-683. DOI: 10.1111/j.1469-8749.2008.03055.x

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language, 39*, 1-27. DOI:10.1017/S0305000910000759

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development, 32*, 262-274. DOI: 10.1016/j.infbeh.2009.03.004

Höhle, B., Pauen, S., Hesse, V., & Weissenborn, J. (2014) Discrimination of rhythmic pattern at 4 months and language performance at 5 years: A longitudinal analysis of data from German-learning children. *Language Learning, 64,* 141-164. DOI: 10.1111/lang.12075

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

Johnson, E.K. (2012). Bootstrapping language: Are infant statisticians up to the job? In P. Rebuschat & J. Williams (Eds.). *Statistical learning and language acquisition* (pp.55-89). Berlin: Mouton de Gruyter.

Johnson, E.K., & Jusczyk, P.W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of memory and language*, *44*, 548-567. DOI: 10.1006/jmla.2000.2755

Junge, C. (2017). The proto-lexicon: Segmenting word-like units from the speech stream. In G. Westermann & N. Mani (Eds.) *Early Word learning* (pp. 20 -45). New York: Routledge.

Junge, C, & Cutler, A (2014). Early word recognition and later language skills. *Brain Sciences*, *4*(4), 532-559. DOI: 10.3390/brainsci4040532

Junge, C., Cutler, A., & Hagoort, P. (2014). Successful word recognition by 10-month-olds given continuous speech both at initial exposure and test. *Infancy*, *19*, 179-193. DOI: 10.1111/infa.12040

Junge, C., Kooijman, V., Hagoort, P., & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental science*, *15*, 463-473. DOI: 10.1111/j.1467-7687.2012.1144.x

Jusczyk, P.W., & Aslin, R. N. (1995). Infants′ detection of the sound patterns of words in fluent speech. *Cognitive psychology*, *29*(1), 1-23. DOI: 10.1006/cogp.1995.1010

Jusczyk, P., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*, 675-687. DOI: 10.1111/j.1467-8624.1993.tb02935.x

Jusczyk, P.W., Houston, D.M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive psychology*, *39*, 159-207. DOI: 10.1006/cogp.1999.0716

Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., & Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Developmental Science 14*, 1100–1106. DOI: 10.1111/j.1467-7687.2011.01056.x

Kuhl, P.K. (1991). Human adults and human infants show a ``Perceptual Magnet Effect'' for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*, 93–107.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V., Stolyarova, E., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*, 684-686. DOI: 10.1126/science.277.5326.684

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 (NLM-e). *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 363(1493), 979-1000. DOI: 10.1098/rstb.2007.2154

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the" critical period". *Language Learning and Development, 1*, 237-264. DOI: 10.1080/15475441.2005.9671948

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences, 100*, 9096-9101. DOI: 10.1073/pnas.1532872100

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, New Series, *Vol. 255*, No. 5044, 606–608. DOI: 10.1126/science.1736364

Lecanuet, J. P., Gautheron, B., Locatelli, A., Schaal, B., Jacquet, A. Y., & Busnel, M. C. (1998). What sounds reach fetuses: biological and nonbiological modeling of the transmission of pure tones. *Developmental Psychobiology, 33,* 203–19. DOI: 0.1002/(SICI)1098-2302(199811)33:3<203::AID-DEV2>3.0.CO;2-V

Leppänen, P. H., Richardson, U., Pihko, E., Eklund, K. M., Guttorm, T. K., Aro, M., & Lyytinen, H. (2002). Brain responses to changes in speech sound durations differ between infants with and without familial risk for dyslexia. *Developmental Neuropsychology, 22*, 407-422. DOI: 10.1207/S15326942dn2201_4

Levine, D., Strother-Garcia, K., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Language development in the first year of life: What deaf children might be missing before cochlear implantation. *Otology & Neurotology, 37,* e56-e62. DOI: 10.1097/MAO.0000000000000908

Lew-Williams, C, Pelucchi, B, & Saffran, JR (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science, 14*, 1323-1329. DOI: 10.1111/j.1467-7687.2011.01079.x

Liberman, A. M., Safford Harris, K., Hoffman, H. S., & Griffith, B. C. (1957). The Discrimination of Speech Sounds within and Across Phoneme Boundaries. *Journal of Experimental Psychology, 54,* 358–368. DOI: 10.1037/h0044417

Liu, H.M., Kuhl, P.K., & Tsao, F.M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*, F1-F10. DOI: 10.1111/1467-7687.00275

Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition, 133*, 385-394. DOI: 10.1016/j.cognition.2014.06.004

MacNeilage, P.F., Davis, B.L., Kinney, A., & Matyear C.L. (1999). Origin of serial-output complexity in speech. *Psychological Science, 10,* 459-460. DOI: 10.1111/1467-9280.00187

Malsheen, B. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In: G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.), Child phonology (Vol. 2) (pp. 173–184). New York: Academic Press.

Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. (2016). Utterances in infant-directed speech are shorter, not slower. *Cognition, 156,* 52-59. DOI: 10.1016/j.cognition.2016.07.015

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers Speak Less Clearly to Infants Than to Adults: A Comprehensive Test of the Hyperarticulation Hypothesis. *Psychological Science, 26,* 341–347. DOI: 10.1177/0956797614562453

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*, 241-265. DOI: 10.1207/s15327078in1003_3

Mattys, S.L., & Jusczyk, P.W. (2001a). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78(2),* 91-121. DOI: 10.1016/S0010-0277(00)00109-8

Mattys, S.L. & Jusczyk, P.W. (2001b). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance, 27(3)*, 644. DOI: 10.1037/0096-1523.27.3.644

Mattys, S.L., White, L., & Melhorn, J.F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477. DOI: 10.1037/0096-3445.134.4.477

May, L., Gervain, J., Carreiras, M., & Werker, J. F. (2017). The specificity of the neural response to speech at birth. *Developmental Science, 21*. DOI: 10.1111/desc.12564

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition, 129*, 362–378. DOI: 10.1016/j.cognition.2013.07.015

Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143-178. DOI: 10.1016/0010-0277(88)90035-2

Minagawa-Kawai, Y.,van der Lely, H., Ramus, F., Sato, Y., Mazuka, R., & Dupoux, E. (2011). Optical brain imaging reveals general auditory and language-specific processing in early infant development. *Cerebral Cortex, 21*, 254–261. DOI: 10.1093/cercor/bhq082

Moon, C. M., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica, International Journal of Paediatrics, 102,* 156–160. DOI: 10.1111/apa.12098

Narayan, C. (2013). Developmental perspectives on phonological typology and sound change. In A. C. L. Yu (Ed.), *Origins of sound Change: Approaches to phonologization* (pp. 9–40). Oxford: Oxford University Press.

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The Interaction Between Acoustic Salience and Language Experience in Developmental Speech Perception: Evidence from Nasal Place Discrimination. *Developmental Science, 13,* 407–420. DOI: 10.1111/j.1467-7687.2009.00898.x

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 756-766. DOI: 10.1037/0096-1523.24.3.756

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development, 21,* 779-784. DOI: 10.1016/S0163-6383(98)90044-3

Nelson, D. G. K., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16,* 55-68. DOI: 10.1017/S030500090001343X

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realization in VO and OV languages. *Lingue e linguaggio, 7(2),* 1-28. DOI: 10.1418/28093

Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Developmental psychology*, *42*(4), 643-655. DOI: 10.1037/0012-1649.42.4.643

Nicholas, J. G., & Geers, A. E. (2007). Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe to profound hearing loss. *Journal of Speech, Language, and Hearing Research, 50,* 1048-1062. DOI:10.1044/1092-4388(2007/073)

Nittrouer, S. (1996). Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds. *Journal of Speech and Hearing Research, 39*, 278–297. DOI:10.1044/jshr.3902.278

Nittrouer, S., Manning, C., & Meyer, G. (1993). The perceptual weighting of acoustic cues changes with linguistic experience. *Journal of Acoustical Society of America, 1865,* 68131. DOI: 10.1121/1.407649

Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: an optical topography study of language recognition at birth.

Proceedings of the National Academy of Sciences, *100*, 11702-11705. DOI: 10.1073/pnas.1934290100

Peña, M., Pittaluga, E., & Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences, 107,* 3823-3828. DOI: 10.1073/pnas.0914326107

Peña, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier speech exposure does not accelerate speech acquisition. *Journal of Neuroscience*, *32*, 11159-11163. DOI: 10.1523/JNEUROSCI.6516-11.2012

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the acoustical society of America, 24*, 175-184. DOI: 10.1121/1.1906875

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation.* Retrieved from https://dspace.mit.edu/handle/1721.1/16065

Polka, L., & Bohn, O.S. (2011). Natural Referent Vowel (NRV) framework: An Emerging View of Early Phonetic Development. *Journal of Phonetics, 39*, 467–478. DOI: 1016/j.wocn.2010.08.007

Polka, L., & Werker, J.F. (1994). Developmental Changes in Perception of Nonnative Vowel Contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 421–435.

Pons, F., Abareda-Catellot, B., & Sebastián-Galles, N. (2012). The Interplay Between Input and Initial Biases: Asymmetries in Vowel Perception During the First Year of Life. *Child Development, 83,* 965–976. DOI: 10.1111/j.1467-8624.2012.01740.x

Ponton, C. W., Eggermont, J. J., Coupland, S. G., & Winkelaar, R. (1992). Frequency-specific maturation of the eighth nerve and brain-stem auditory pathway: Evidence from derived auditory brain-stem responses (ABRs). *The Journal of the Acoustical Society of America, 91,* 1576–1586. DOI: 10.1121/1.402439

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crepin, G. (1988). Fetal hearing. *European Journal of Obstetrics and Gynecology and Reproductive Biology, 28*, 191–212. DOI: 10.1016/0028-2243(88)90030-5

Ramachers, S., Brouwer, S., & Fikkert, P. (2018). No perceptual reorganization for Limburgian tones? A cross-linguistic investigation with 6- to 12-month-old infants. *Journal of Child Language*, *45*, 290-318. DOI: 10.1017/S0305000917000228

Ramírez- Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science, 17,* 880-891. DOI: 10.1111/desc.12172

Ramus, F., Hauser, M.D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*, 349-351. DOI: 10.1126/science.288.5464.349

Repp, B.H. (1982). Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychological Bulletin, 92,* 81–110. DOI: 10.1037/003

Repp, B. (1984). Categorical Perception: Issues, Methods, Findings. In N. J. Lass (Ed.), *Speech and Language: Advances in Basic Research and Practice* (Vol. 10, pp. 243–335). San Diego, CA: Academic Press.

Rosner, B. S., & Pickering, J. B. (1994). *Vowel perception and production*. Oxford: Oxford University Press.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science, 12,* 339-349. DOI: 10.1111/j.1467-7687.2008.00786.x

Saffran, J.R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81,* 149-169. DOI: 10.1016/S0010-0277(01)00132-9

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926-1928. DOI: 10.1126/science.274.5294.1926

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology, 33*, 3-11. DOI: 10.1037//0012-1649.33.1.3

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Development of hemispheric specialization for lexical pitch-accent in Japanese infants. *Journal of Cognitive Neuroscience, 22*, 2503-2513. DOI: 10.1162/jocn.2009.21377

Schmale, R., Cristià, A., Seidl, A., & Johnson, E. K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy*, *15*, 650-662. DOI: 10.1111/j.1532-7078.2010.00032.x

Segal, O., & Kishon-Rabin, L. (2012). Evidence for language-specific influence on the preference of stress patterns in infants learning an iambic language (Hebrew). *Journal of*

Speech, Language, and Hearing Research, 55, 1329-1341. DOI: 10.1044/1092-4388(2012/11-0087)

Seidl, A, & Johnson, E.K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental science, 9,* 565-573. DOI: 10.1111/j.1467-7687.2006.00534.x

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition, 106(2),* 833-870. DOI: 10.1016/j.cognition.2007.05.002

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' Listening Preferences: Baby Talk or Happy Talk? *Infancy, 3,* 365–394. DOI: 10.1207/S15327078IN0303_5

Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language, 51,* 173-189. DOI: 10.1016/j.jml.2004.04.004

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., Alves Limissuri, R., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science, 12*, 914–919. DOI: 10.1111/j.1467-7687.2009.00835.x

Soderstrom, M. (2007). Beyond Babytalk: Re-evaluating the Nature and Content of Speech Input to Preverbal Infants. *Developmental Review, 27,* 501–532. DOI: 10.1016/j.dr.2007.06.002

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50,* 86-132. DOI: 10.1016/j.cogpsych.2004.06.001

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B, 364*, 3617-3632. DOI: 10.1098/rstb.2009.0107

Tang, P., Xu Rattanasone, N., Yuen, I. & Demuth, K. (2017). Phonetic enhancement of Mandarin vowels and tones: infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America, 142*, 493–503. DOI: 10.1121/1.4995998

Thiessen, E.D. & Erickson, L.C. (2012). Discovering words in fluent speech: the contribution of two kinds of statistical information. *Frontiers in Psychology, 3*, 590. DOI: 10.3389/fpsyg.2012.00590

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7,* 53-71. DOI: 10.1207/s15327078in0701_5

Thiessen, E.D., Kronstein, A.T. & Hufnagle, D.G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological bulletin, 139,* 792 - 814. DOI: 10.1037/a0030801

Thiessen, E.D. & Saffran, J.R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental psychology*, *39*, 706 -716. DOI: 10.1037/0012-1649.39.4.706

Trubetzkoy, N. S. (1939/1969). *Principles of phonology translated by C.A. Baltaxe*. Berkeley, CA: University of California.

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America, 120*, 2285-2294. DOI: 10.1121/1.2338290

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, 56, 179–191. DOI: 10.1002/dev.21179

Tsuji, S., Mazuka, R., Cristia, A., & Fikkert, P. (2015). Even at 4 months, a labial is a good enough coronal, but not vice versa. *Cognition, 134,* 252–256. DOI: 10.1016/j.cognition.2014.10.009

van Alphen, P., De Bree, E., Gerrits, E., De Jong, J., Wilsenach, C., & Wijnen, F. (2004). Early language development in children with a genetic risk of dyslexia. *Dyslexia, 10(4),* 265-288. DOI: 10.1002/dys.272

van Heugten, M., & Johnson, E. K. (2012). Infants exposed to fluent natural speech succeed at cross-gender word recognition. *Journal of Speech, Language, and Hearing Research, 55,* 554-560. DOI:10.1044/1092-4388(2011/10-0347)

van de Weijer, J. (1998). *Language input for word discovery* (PhD Dissertation). MPI Series in Psycholinguistics 9, Radboud University Nijmegen.

Vavatzanidis, N. K., Mürbe, D., Friederici, A., & Hahne, A. (2015). The basis for language acquisition: congenitally deaf infants discriminate vowel length in the first months after cochlear implantation. *Journal of Cognitive Neuroscience 27*, 2427 - 2441. DOI: 10.1162/jocn_a_00868

Vavatzanidis, N. K., Mürbe, D., Friederici, A. D., & Hahne, A. (2016). The perception of stress pattern in young cochlear implanted children: an EEG study. *Frontiers in Neuroscience, 10*, 68. DOI: 10.3389/fnins.2016.00068.

Vouloumanos, A., & Werker, J.F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*, 159-164. DOI: 10.1111/j.1467-7687.2007.00549.x

Werker, J. F., Byers-Heinlein, K., & Fennell, C. T. (2009). Bilingual beginnings to learning words. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364*, 3649-3663. DOI: 10.1098/rstb.2009.0105

Werker, J.F. & Curtin, S. (2005). PRIMIR: a developmental framework of infant speech processing. *Language Learning and Development, 1*, 197-234. DOI: 10.1080/15475441.2005.9684216

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: new directions. *Annual Review of Psychology*, *66*, 173-196. DOI: 10.1146/annurev-psych-010814-015104

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49-63. DOI: 10.1016/S0163-6383(84)80022-3

Wright, R. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. M. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology (pp. 34–57)*. Cambridge, UK: Cambridge University Press.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input reorganize phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language, 68*(2), 123-139. DOI: 10.1016/j.jml.2012.09.004

Glossary for Chapter 15

- *Affect:* Emotional valence, which can be expressed in the speech signal.
- *Bottom-up:* Knowledge of segmental or suprasegmental categories in speech are learnt from patterns in speech input; does not require lexical knowledge (i.e., understanding of the speaker's message). Compare *top-down acquisition*.
- *Discrimination:* to tell apart two items. If perception is categorical, discrimination can be taken to indicate that these items belong to two distinct categories.
- *Duration:* the length of a unit of speech, such as a speech sound (phoneme), syllable or word.
- *Formants:* Prominent frequency bands in the speech signal, most typically in vowels, that result from vocal tract resonance. The combination of these formants, in particular the frequency of the first and second formant, render vowel quality.
- *Fundamental frequency:* The longest repeating pattern in the speech signal that results from vocal fold vibration. Adult males typically have a lower fundamental frequency (85 to 155 Hz) compared to females (165 – 225 Hz). It is perceived as the pitch of the voice and as such a key characteristic of speaker identity, and also signals intonation patterns. Also known as F0 ("F-zero").
- *Iambic:* Stress pattern of a bisyllabic stress-bearing unit at the word level that starts with an unstressed (weak) syllable followed by a stressed (strong) syllable (e.g., bi-ZAR, pro-POSE).
- *Intensity:* A measure of energy in the signal. The physical intensity of a speech signal is its sound pressure level or amplitude. The perceived intensity is expressed in deciBel (dB). Intensity contributes to the perception of loudness, and can vary between segments and intonation patterns.
- *Intonation:* Changes in fundamental frequency (falling or rising), intensity and duration (long, short elements) conveying prosodic information relevant at the phrase- or sentence-level in speech.
- *Lexical pitch accent:* Pitch pattern of a word (e.g. rising, dipping, flat) that is contrastive.
- *Lexical stress:* Accentuation of syllables within words, which renders them more prominent than other syllables within the word. Also known as word stress.
- *Lexical tone:* Pitch pattern of a syllable (e.g. rising, dipping, flat) that is contrastive.
- *Perceptual attunement:* the process of tuning general perceptual abilities to be specific to the environment. Of speech: the process of learning to discriminate between segments that are phonemic on one's native language, and learning not to discriminate between segments that are not phonemically contrastive.
- *Perceptual-magnet effect:* Discrimination between tokens of the same phoneme category is worse around tokens that listeners consider good or 'prototypical' exemplars of the category, and better around tokens that listeners consider poor or peripheral exemplars of the category. As phoneme categories and their exact prototypes are language specific, so is the perceptual magnet effect.
- *Phoneme:* Speech sound that distinguishes one word from another within one language.
- *Phonetics:* The study of the physical realization of speech, including the production, acoustics, and perception. Phonetics is often contrasted with phonology, which is the study of the abstract representation of speech.
- *Phonotactics:* Constraints regarding which combinations of phonemes (and in which order) are permissible in a language and which ones cannot. Dutch for instance allows /kn/ consonant clusters ('/knop', "bud") whereas English does not.

- *Pitch:* A listener’s perception of fundamental frequency as 'high' and 'low' sounds. See *fundamental frequency*.
- *Prosody*: The study of rhythm, stress, and pitch in speech.
- *Register:* A subset of forms of a language that is typically restricted to usage in a specific context. The register called "infant-directed speech" consists of a high and varied pitch, among other characteristics, and is restricted to interactions with infants.
- *Rhythm:* Rhythmic aspects of speech relate to its temporal (duration) and other suprasegmental (intensity, pitch) acoustic regularities.
- *Segmental:* Phonetic and phonological characteristics that are restricted to a single segment, i.e., at the speech sound level.
- *Spectral information:* The frequency information in the speech signal beyond the fundamental frequency. Typically represented using a spectrum or spectrogram (See Figure 1 from Chapter 15). See also *fundamental frequency* and *formants*.
- *Speech segmentation problem:* The situation that infants face that their speech input comprises mainly multi-word utterances, with no pauses between words to reliably mark word boundaries (See Figure 1 from Chapter 15). As infants can hardly rely on lexical knowledge, this makes it difficult for infants to extract words from speech.
- *Suprasegmental:* Phonetic and phonological characteristics that are not restricted to a single segment (i.e., go beyond the level of speech sound). Suprasegmental phonetic characteristics include pitch, loudness, and duration. Suprasegmental phonological units include the syllable, foot, and phrases.
- *Top-down acquisition:* Knowledge of segmental or suprasegmental categories in speech are learnt from lexical knowledge (i.e., word forms or word-meaning combinations). Compare *bottom-up acquisition.*
- *Trochaic:* Stress pattern of a bisyllabic stress-bearing unit at the word level that starts with a stressed (strong) syllable followed by an unstressed (weak) syllable (e.g., KING-dom, FE-male).
- *Vocal tract:* The sequence of cavities from the vocal folds (in the larynx) to the lips, including the pharyngeal cavity, oral cavity, and nasal cavity. The shape and size of the vocal tract differs between speakers as a result of anatomical differences, as well as within speakers as a result of (speech) articulation. See also *formant frequencies*.