The what and when of universal perception: A review of early speech sound acquisition

Kateřina Chládková[a,b] and Nikola Paillereau[a,b]

[a]Institute of Phonetics, Faculty of Arts, Charles University, Prague, Czechia
[b]Institute of Psychology, Czech Academy of Sciences, Prague, Czechia

This is a pre-print version of the article, accepted for publication in *Language Learning.*

Abstract

The young universal listener is an established concept in psycholinguistics. However, it is unclear what abilities universal perception entails and at what age it exists. This paper aims to motivate re-thinking about what it means to be a universal listener. Early and recent studies on infant speech acquisition are reviewed, considered in the light of cross-language variation and adults' performance, and finally linked to the current understanding of foetal hearing and learning. It turns out that language-universal perception is best described as an auditory-based perception rather than an ability to perceptually categorize the sounds of any possible language. Interestingly, at birth infants might no longer listen in a language-universal mode since they begin to learn from the ambient speech signal at least several weeks before birth. Future studies need to answer the remaining questions concerning the point in perinatal development at which speech perception begins to take on language-specific traits and for which sounds.

Keywords: speech sound acquisition, early language development, universal listener, first language acquisition, foetal and newborn speech perception

Acknowledgments:

This work was supported by Charles University, grant Primus/17/HUM/19, and by Czech Science Foundation, grant 18-01799S. We are very grateful to Šárka Šimáčková, Václav Jonáš Podlipský, Simon Gill, the journal editor Emma Marsden, and four anonymous reviewers whose careful reading and detailed comments helped to improve the paper.

Address for correspondence:

Kateřina Chládková, Faculty of Arts, Charles University, Nám. Jana Palacha 2, 116 38, Praha, Czechia, +420 221 619 250, katerina.chladkova@ff.cuni.cz

1 Introduction

Humans have the remarkable ability to acquire language and learn to speak, and they do both of these seemingly effortlessly and rather fast. It takes a child about five years to produce the sounds of her language in an adult-like fashion (McLeod & Crowe, 2018), but she attains the perceptual and comprehension abilities specific to her native language much earlier than this. Before her seventh month, when she typically starts to babble (Molemans, van den Berg, Van Severen, & Gillis, 2012), an infant has already acquired substantial knowledge about the speech sounds that make up her native language. The view holds that native language categories are formed between four and six months of age for vowels, and slightly later, i.e. between the tenth and 12th months, for consonants (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Tsuji & Cristia, 2014; Werker & Tees, 1984). Evidence for the development of language-specific speech sound categories comes from studies that tested infants' processing of native and non-native speech sound contrasts and often found that younger infants can perceptually discriminate both native and non-native contrasts, while older infants, similarly to adults, discriminate the native contrasts robustly and show reduced or no perceptual sensitivity to the non-native ones.

It is relatively clear what to expect of an infant once she has become an adult-like listener who is perceiving and comprehending speech in ways specific of her native language. Yet, what one can expect of an infant at the very outset of her route towards being such a language-specific listener turns out to be much less obvious. This article aims to review what we know – and identify what is still to be found out – about that initial stage of speech acquisition.

Over the past 50 years, developmental speech researchers (represented by the works cited in this review and many others that fall outside the scope of the present study) have greatly advanced the field with their discoveries. Infants who are just starting to acquire their speech sound categories have been shown to discriminate many segmental speech sound contrasts, including those that are not functional in their mother tongue (Best, 1994; Jusczyk, 1995; Maurer & Werker, 2013; Werker & Tees, 1992). These maximal perceptual abilities, which occur at birth or shortly afterwards, seem to be what makes an infant a universal listener with universal perceptual abilities. Both terms most probably originated from Aslin and Pisoni's (1980) universal theory, and have been in use ever since (Kuhl, 2004; Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Lim, Fiez, & Holt, 2019; Nakisa & Plunkett, 1998; Werker, 1995, to name a few).

Theoretical frameworks of language acquisition address in detail how speech and language abilities develop from the initial, universal stage onwards but are often less specific about the very first stage itself. The initial capabilities, however, determine whether, when, and how any further processes will unfold; understanding them is thus vital for our knowledge of both first and second language acquisition.

A closer look at the speech acquisition literature reveals that authors may differ in how they view, or (implicitly) define, the concept of universality in speech perception, sometimes referred to as language-generality. Additionally, the literature does not pinpoint the age at which any such universal abilities are attested. What is more, the results of some studies indicate that even the youngest infants might exhibit non-universal perceptual patterns, that is, patterns that are specific to the ambient language. It therefore seems worthwhile to re-think and re-examine the speech perception abilities that humans have at the start of their linguistic development. We aim to do so in three steps.

The first goal of this paper is to review how one may interpret universality in speech perception. In Section 2, we exemplify the two interpretations of universality that can be found in the literature: one as language-general categorical perception and the other as a (language-

)general auditory sensitivity. We discuss the definition of categorical perception, cross-linguistic variability, and infants' as well as adults' actual performance in speech perception experiments. It turns out that universality, if considered as an ability to organize sounds into some language-general perceptual categories, is a rather implausible concept (and may not be measurable). Instead, the perceptual patterns found in the youngest infants seem to be better explained by language-general – or universal – auditory sensitivities.

The second question we investigate is whether there is empirical support for language-general perception in newborn infants. To that end, Section 3 reviews studies that focus on the youngest infant ages, showing that quite limited data is available about the speech sound discrimination capabilities that an infant has at birth. The section further discusses early and recent studies whose results suggest that newborns and very young infants may already show signs of language-specific perception of speech sounds. An overview of the methods and outcomes of previous studies shows that young infants do not always reveal their true perceptual abilities in behavioural paradigms, suggesting an advantage of using neural measures.

Our third goal is to identify the source of the language-specific perception of speech sounds at birth or shortly afterwards. Section 4 draws a link between the well-described effects of prenatal exposure on the perception of speech rhythm and melody and the hitherto underinvestigated but plausible prenatal learning of speech segments. Vowels especially appear to be ideal candidates for prenatal learning, because of their loudness and perceptual distinctiveness. The mechanisms and timeline of the early formation of speech categories are discussed.

Section 5 outlines developmental speech research areas that lie outside the scope of the present review. Section 6 then proposes several possible directions for future research. Table 1 and Figure 1 summarize the reviewed findings on speech sound discrimination in foetuses and infants younger than six months of age.

Our conceptual review leads to the conclusion that any universal sensitivities that language learners have at the outset of learning are most probably auditory based. The general auditory sensitivities are modulated by the psychophysical properties of the sound, such as the differing saliency across various stimulus dimensions, as well as by the physical constraints of the human body and the surrounding conditions, such as the type and range of sounds that a foetus is able to hear *in utero*. It is further argued that infants who are several hours, days, or weeks old might already be too experienced with (some of) the sounds of the ambient speech to be language-general, i.e. universal listeners. Instead, universal perception seems to describe more accurately the listening abilities that an individual has before he or she is even born.

2 What is universal perception?

2.1 Two possible interpretations: auditory or categorical language-universality

The first question we address deals with the definition of a universal listener, that is, the abilities that the term universality refers to. Finding an answer to this seemingly straightforward question is no trivial matter.

According to the Universal Theory, newborns should possess maximal perceptual abilities that allow them to discriminate virtually "all the possible phonetic contrasts that may be used phonologically in any natural language" (Aslin & Pisoni, 1980, p. 79). Under a strict interpretation of the Universal Theory, then, the infant's initially maximal perceptual sensitivity will decline for speech sound contrasts not present in the ambient language and will remain unaffected for native contrast (note, however, that the Universal Theory is just one of three possible paths for speech sound development that Aslin and Pisoni, 1980 formulate). In line with that proposal, numerous studies have indeed demonstrated that young infants (usually several months old) are able to discriminate phonetic differences to which they have not been exposed before and that this ability diminishes with age. Initial good discrimination and/or its developmental decline for non-native phonetic distinctions have been reported both for consonants (Burns, Yoshida, Hill, & Werker, 2007; Kuhl et al., 2006; Lasky, Syrdal-Lasky, & Klein, 1975; Streeter, 1976; Trehub, 1976; Tsushima et al., 1994; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984; Werker & Lalonde, 1988) and for vowels (Bosch & Sebastián-Gallés, 2003; Polka & Werker, 1994; Trehub, 1976).

Since all languages use vocalic and consonantal categories, most developmental psycholinguistic studies available, and those included in the present review, have focused on how infants acquire vowels and consonants. Segmental phonology in some languages can also entail tone or length contrasts, realized through changes in the fundamental frequency and duration of individual speech sound segments, respectively. The acquisition of tone and length might unfold slightly differently compared to the other vocalic and consonantal contrasts, but as it is currently still understudied, the present review does not discuss the universality of the perception of tone and length; see Section 5 for more detail.

Let us now consider what it means to be a good perceiver of phonetic differences (which must be sufficiently audible to be discriminable, and thus can potentially constitute a phonological contrast in some language). One might understand the universal listener as a good auditory perceiver who discriminates well across clearly audible speech sound differences and discriminates less well across poorly audible differences. This interpretation appears to prevail in the current understanding of universality in speech perception (see Junge, Boll-Avetisyan, & Benders, 2019). In that view, then, universality is interpretable as an auditory-driven, non-phonological discrimination of speech sounds.

A different understanding of universality might be formulated in terms of language-general categorization of speech stimuli. Some studies, especially earlier ones, proposed that the youngest infant is able to discriminate any possible phonological (i.e. categorical) contrast, irrespective of her or his native language (Aslin, Pisoni, Hennessy, & Perey, 1981; Eimas & Corbit, 1973; Werker, 1995). In this latter view, the young infant does not discriminate speech sound distinctions only on the basis of their auditory distance but is particularly sensitive to those distinctions that constitute a phonological contrast in any of the world's languages. Such language-general perceptual categorization is then supposedly driven by cross-linguistically favoured and natural category boundaries (Kuhl, 2004).

These two interpretations of the universal listener agree in the assumption that at some early age in the infant's life, he or she perceives speech sounds in the same way as any other similarly young infant in the world, irrespective of their native languages. However, they diverge on what underlies such language universality – whether it is auditory-based or category-based. Below we review studies that tested speech sound discrimination in infants from various language backgrounds (2.2) and discuss why it is important to distinguish the two interpretations and show that the category-based view is somewhat counterintuitive (2.3). Consideration of the stimuli (2.4) and tasks (2.5) further reveals that infants' (and adults') perceptual categorization of speech sounds may not be straightforwardly measurable with the methods employed to date. The issues raised in the respective subsections lead to the conclusion that the auditory-based view of universality, but not the category-based view, is compatible with the previously reported findings.

2.2 Evidence for early auditory- versus category-based perception

Studies focusing on infants' discrimination of vowels often employed stimuli that were separated by a large acoustic distance. Testing the ability to discriminate across [a], [i], and [u], previous research found that infants distinguish those vowels from birth (Kuhl, 1979; Kuhl & Miller, 1982; Kujala et al., 2004; Marean, Werner, & Kuhl, 1992; Sebastián-Gallés & Bosch, 2009; Trehub, 1973) if not earlier (Lecanuet et al., 1987; Shahidullah & Hepper, 1994). These findings, however, do not allow us to deduce whether the infants' ability to discriminate between [a], [i], and [u] is auditory- or category-based as the differences amongst those vowels are both auditorily large (such that even non-human animals discriminate them; Baru, 1975; Burdick & Miller, 1975; Dewson, 1964), and categorically relevant in all of the world's languages (all languages contrast at least two vowels with qualities approximating those of [a] versus [i] or [u]; Maddieson, 1986).

Studies that tested infants on their discrimination of acoustically smaller non-native contrasts can reveal a little more about the true nature of the early perceptual abilities. A large body of previous work has shown that young infants can discriminate relatively small acoustic differences between stimuli even if those do not represent a phonemic contrast in their native language. Two-month-old English-learning infants discriminate the French oral vs. nasal vowel difference /a/-/ã/ cued by a shift in the vowels' F1 and F2 (Trehub, 1976). Four-month-old Spanish-learning infants discriminate the non-native Catalan /e/-/ɛ/ distinction that is primarily cued by the vowels' first formant (Bosch & Sebastián-Gallés, 2003). At six to eight months (and also at ten-12 months), English-learning infants discriminate the German /u/-/y/ contrast, and, vice versa, German-learning infants discriminate the English /ɛ/-/æ/ contrast (Polka & Bohn, 1996). Very young infants also discriminate non-native consonants. The Czech contrast /ʒ/-/r̝/, realized mainly through the intensity of the fricative formant (stridency), is discriminated by two-month-old English-learning infants (Trehub, 1976). At the same age, Kikuyu-learning infants discriminate the non-native English-like /b/-/p/ contrast (Streeter, 1976). At six months, English-learning infants discriminate two Hindi contrasts: the retroflex vs. dental /ʈa/-/ta/ and the voiceless aspirated vs. voiced breathy /tʰ/-/dʰ/ (Werker et al., 1981). Four-to-six-month-old English-learning infants can discriminate the non-native Filipino coronal vs. dorsal /n/-/ŋ/ consonantal contrast, as well as the Tamil nasal /n̪/-/ɳ/ and lateral /l/-/ɭ/ dental vs. retroflex contrasts, which are realized primarily through F2-F3 distance (Sundara et al., 2018).

Infants can thus discriminate relatively subtle auditory differences in speech sounds even if these do not constitute a contrast in their native language. However, this fact is, again, compatible with both the auditory and the category-based view of young infants' perception. Below, we investigate whether we can find evidence that favours one and disproves the other.

There is one early study that reported the initial ability to perceive some non-native speech sound contrasts *categorically*, that is, to discriminate auditory differences more robustly if they represent a between-category, i.e. phonologically relevant or phonemic change than if they represent a within-category, i.e. auditory only difference in a *non-native* language. Werker and Lalonde (1988) tested American English-learning infants on their discrimination of differences in the place of articulation in plosives that either crossed a (Hindi or English) category boundary or fell within a single (Hindi) category. The authors found that six-to-eight-month-olds discriminated a change between [b̥] and [d̥], corresponding to a native English /b/-/d/ categorical contrast, as well as a change between [d̥] and [ɖ̥], corresponding to a non-native Hindi /d̪/-/ɖ/ (dental-retroflex) categorical contrast, but were less sensitive to a comparable acoustic change within the Hindi /ɖ/-category. Older, 11-13-month-old infants (and American

English adults) did not have great sensitivity to the non-native categorical contrast. The six-to-eight-month-olds' better discrimination of the non-native between-category difference than the non-native within-category one was taken to reflect early-stage universal categorical effects. Note that in the 1988 study (Werker & Lalonde, 1988), infants were tested on three successive days, always receiving the native contrast on the first day, the non-native between-category on the second day, and the non-native within-category on the last day. Very speculatively, the fixed order (which the authors discuss and give reasons for) might have confounded the category-status factor and might also have affected younger and older infants differently. This potential confound should be borne in mind when searching for unequivocal evidence of either the auditory or the categorical nature of infants' initial speech perception abilities.

2.3 The importance of defining the early perceptual ability accurately

The two-fold interpretation of infants' early language-general abilities that we brought up here has, to the best of our knowledge, not been explicitly addressed in the literature to date. For that reason, the design of most previous studies does not allow the auditory-driven early perception to be disentangled from its category-driven counterpart. The knowledge of the development of speech perception, and theories thereof, have made great advances in recent decades and perhaps now is the time for a careful examination of the perceptual mechanisms that an infant has at its disposal in the very initial stage.

Uncovering the nature of infants' earliest perception seems particularly intriguing because the two interpretations we have summarized above lead to different predictions as to what and how the developing infant is going to learn. If she starts with a language-general *auditory* perception, the infant will have to learn to categorize the ambient speech sound world from scratch. In contrast, if she starts with a language-general *categorical* perception, the infant's (perhaps slightly less demanding) task will be to maintain (and possibly reinforce) the categorical contrasts that occur in her environment and discard those that do not. The various learning scenarios that the infant needs to go through could be related back to some of the existing theories of the development of speech perception (for instance, Aslin and Pisoni's 1980 perceptual learning and attunement theory, respectively) and provide ground for new, hypothesis-driven experiments.

Although the evidence to date does not unequivocally speak for either auditory-based or category-based early perception, in the remainder of this section we argue that the concept of the language-general categorical listener is in itself somewhat counterintuitive.

By definition, categorical effects in speech perception are found when stimuli belonging to the same phonemic category are less likely to be discriminated than stimuli – separated by the same amount of psychoacoustic distance – belonging to two different categories (Pisoni, 1973; Repp, 1984). Such perceptual categorization of speech does not require the categories to be labelled but arises from a listener's general cognitive ability to discern categorical (probability) structures in the input (Holt & Lotto, 2010). Some developmental literature argues that the young universal listener can discriminate any possible phonologically relevant contrast, i.e. a speech sound difference that spans a category boundary in at least some language of the world, and does not discriminate (so well) an equal-sized difference between speech sounds that do not represent a phonologically relevant contrast in any language (Kuhl, 2004; Werker, 1995; see also Gervain & Werker, 2008, p. 1158).

Assuming that a young infant without previous exposure to speech perceives any speech sound contrast from any possible language categorically, one would hardly be able to *demonstrate* any universal categorical perception effects. This is because a given speech sound difference that constitutes a within-category distinction in one language may (and often does) represent a between-category distinction in another language: for instance, the VOT distinction of -85 ms versus +5 ms represents a within-category difference in English but a between-category difference in languages such as Dutch, Spanish, and Hungarian (see Lisker and Abramson, 1964), or the spectral (F1) distinction between an [e]-like and an [ɛ]-like sound represents a within-category variation in Spanish but a between-category difference in Portuguese (Chládková, Escudero, & Boersma, 2011; Escudero, Boersma, Rauber, & Bion, 2009). Thus, an infant whose perception is guided by all of the world's languages' categories must nearly always perceive and discriminate a given stimulus pair as a between-category difference (from whichever language) and will thus have only rare opportunities to show reduced discrimination of within-category changes. Along those lines, revealing reduced discrimination of a within-category difference (such as the -85 ms vs. -5 ms VOT in English) could possibly even be understood as evidence *against* the universal *categorical* listener.

2.4 Psychoacoustic saliency confounds in tests of categorical perception

One might counter that infants repeatedly show reduced discrimination of plosive voicing differences between long-lag and extra-long-lag VOT compared to differences between short-lag and long-lag VOT, and that this effect seems driven by some universal boundaries. However, a caveat to that objection is that in many previous studies, the within-category contrasts that were tested were less distinct than the between-category contrasts on *psychoacoustical* scales, which may have caused the relatively weak discrimination for within-category changes.

To illustrate the psychoacoustic saliency confound, let us now turn to the seminal study by Eimas, Siqueland, Jusczyk, and Vigorito (1971, described in detail in Section 3.2), in which English-exposed one- and four-month-old infants discriminated a +20 vs. +40 ms VOT difference better than they discriminated +40 vs. +60 ms or -20 vs. 0 ms. We propose here that not only category membership but also *psychoacoustic distance* could explain the better discrimination of +20 vs. +40 ms VOT vis-à-vis the poorer discrimination of +40 vs. +60 ms VOT. Since differences in the temporal dimension are most probably processed in relative rather than in absolute terms (Abel, 1972; Smits, Sereno, & Jongman, 2006), it becomes self-evident that the infants discriminated the psychoacoustically greater difference (a 20-ms noise vs. a 40-ms noise representing a ratio of 1:2) better than they discriminated the psychoacoustically smaller difference (a 40-ms noise vs. a 60-ms noise, representing a ratio of 1:1.5; for a similar psychoacoustic distance confound but this time coinciding with a non-native boundary see Burns et al., 2007).

The perceptual saliency of individual acoustic *dimensions* could then explain why the infants discriminated +20 vs. +40 ms better than -20 vs. 0 ms. In Eimas et al.'s stimuli the former difference was realized as the duration of a noise source signal (containing a second and third formant) following a plosive burst, while the latter was realized as the duration of a voicing band during plosive closure (comparable to humming). The difference between silence and a soft 20-ms-long hum might be much less audible than the difference between a short and a twice-as-long aspiration noise, which could explain why the -20 vs. 0 ms VOT distinction was less well discriminated than the +20 vs. +40 ms VOT distinction. An explanation in terms of acoustic salience has been recently formulated also by Narayan (2019), who argues that infants discriminate aspiration-based contrasts better than prevoicing contrasts because the former ones are cued by more robust acoustic cues. It would be an interesting direction for future research to test whether the youngest infants' perceptual patterns reported in the literature are indeed attributable to *relative* VOT differences and to psychoacoustic *saliency* of prevoicing versus aspiration, rather than to the category membership of the stimuli.

Naturally, the perceptual saliency of stimulus differences matters: infants do not always discriminate speech sound differences, even those representing a phonemic distinction in their own native language, suggesting that some contrasts, both native and non-native, are difficult to perceive without extensive exposure. American English-learning infants initially discriminate native /ɹ/-/l/ relatively poorly (at six to eight months) and only improve with age (at ten to 12 months; Kuhl et al., 2006). Comparably, the English /f/-/θ/ contrast may be difficult to distinguish perceptually even at 12 months of age (Eilers, Wilson, & Moore, 1977), although some studies report good discrimination in two- and six-month-old infants (Holmberg, Morgan, & Kuhl, 1977; Levitt, Jusczyk, Murray, & Carden, 1988). Finnish newborns have a strong neural discrimination response (mismatch response) to the [i]-[y] distinction realized through the vowel's second formant as a 2400-Hz vs. 1700-Hz difference and a small, unreliable response to the [i]-[ɨ] distinction of 2400 Hz vs. 2100 Hz (Cheour-Luhtanen et al., 1995). Three-week-to-six-month-old Swedish infants have difficulty discriminating the vowel pair [ɑ]-[a], although they can discriminate another vowel pair, [ɑ]-[ɔ], both resembling native-like phoneme contrasts (Lacerda, 1992).

In this subsection we aimed to show that infants' failure to discriminate a specific speech sound difference may not automatically be attributable to reduced within-category sensitivity, but might be due to the generally lower saliency of the stimulus difference in question. We subsequently reviewed some evidence showing that infants' early speech perception is not unlimited in that for some speech sound differences the infant needs exposure to be able to discriminate them. Another limitation of some developmental studies is the assumption that adult-like perceptual categorization of speech sounds should elicit greater between-category than within-category discrimination. Below we show that such an assumption might not hold across testing paradigms.

2.5 Within-category sensitivity across the life span

Some work on speech perception in human adults, infants, and virtual listeners suggests that testing categorical speech perception in the traditional sense tells us little about whether or not a listener employs her knowledge of the native language phoneme categories and phoneme boundaries. In adults, within-category variation is not ignored during the speech comprehension process but instead makes a substantial contribution to, for instance, word recognition patterns (McMurray, Tanenhaus, & Aslin, 2002). Recent modelling work indicates that particularly in discrimination tasks, adult users of a language may be primarily sensitive to phonetic *detail*, i.e. within-category variation, such that the degree of categorical effects in their discrimination outcomes is dependent on how much of the recovered phonetic variation they attribute to meaningful information and how much they attribute to random noise (Kronrod, Coppess, & Feldman, 2016).

Sensitivity to within-category phonetic variation has also been evidenced in infants (McMurray & Aslin, 2005) and it has been argued that it strengthens gradually throughout their development. Reanalysing the findings of infant studies on the perception of plosive voicing, Galle and McMurray (2014) pointed out that infants do discriminate within-category variation and that this ability increases with age (alongside the strengthening of between-category sensitivity). The authors propose a parallel channels model according to which listeners process speech through a category-encoding and a sensory-sensitivity channel, whose contributions are differentially reflected across various testing paradigms. Galle and McMurray argue that infants' exposure to speech sound variants over time leads to perceptual enhancement of both channels, thus boosting discrimination abilities for between-category as well as within-category differences.

In some experimental paradigms, infants directly exhibit high levels of sensitivity to indexical within-category variation. Mulak, Bonn, Chládková, Aslin, and Escudero (2017) found that Dutch and Australian English-learning 12-month-olds discriminated within-category (speaker identity) and between-category (phonemic) changes in vowels to a comparable extent. It is highly unlikely that such enhanced within-category discrimination in 12-month-old infants was due to the fact that they had not yet acquired the vowel categories of their native language. Instead, it appears that in that particular experiment, the infants' speech processing faculty aimed comparably at deciphering non-categorical variation, as well as at uncovering the linguistic content. Such a phonetic listening strategy might have been preferred with isolated vowel stimuli that hardly carry any lexical content in real-life situations (Mulak et al., 2017). Although using slightly more naturalistic stimuli such as CV or CVC (as most other studies do) may be more likely to prompt linguistic processing in infants, comparing discrimination scores for within- and between-category changes in nonsensical monosyllables might not unequivocally show the extent to which an infant has acquired (and has activated) her native language phonological categories.

To sum up, the typically employed tests of categorical perception that compare the discrimination of within-category and between-category stimulus differences do not seem to be the optimal way of assessing the categorical nature of infants' perception. Instead, we may be able to gain a better understanding of an infant's perceptual abilities with experimental protocols that assess perceptual sensitivity to various stimulus pairs from a particular (phonologically relevant) phonetic continuum and compare it across infants' ages and, crucially, across different languages (in the same way in which the universality or reproducibility of other phenomena in infant development is being investigated in the ManyBabies consortium; Frank et al., 2017).

3 The age of the universal listener

3.1 Lack of newborn data

The previous section raised a crucial question about the nature of an infant's initial perceptual abilities – whether they are auditory-based or defined by categorical perception. In the absence of unequivocal evidence for either interpretation, we presented an argument for why the former seems more plausible. Besides understanding the nature of infants' early perception abilities, another – and perhaps even more rudimentary – question in the debate is whether infants are universal (i.e. language-general) listeners at any age.

Let us suppose that there is indeed a stage at which infants perform more or less similarly across all languages. At this stage infants would be universal listeners, albeit not necessarily because of language-general categorical perception. Let us consider what evidence there is for such a developmental stage. The ages reported for the order of acquisition of vowels and consonants (see e.g. Kuhl, 2004; Werker & Tees, 1992) suggest that if language-general perception exists, it would operate sometime before the fourth month for vocalic sounds and perhaps a little longer for consonantal ones. In line with that reasoning, the (earlier) literature suggested that infants possess universal perceptual abilities at birth or linked universal perception to the initial four- or six-month period of the infant's life (Best, 1994; Best, McRoberts, & Sithole, 1988; Dehaene-Lambertz & Dehaene, 1994; Gervain & Mehler, 2010; Jusczyk, 1995; Kuhl, 1979; Maurer & Werker, 2013; Polka, 1991; Werker & Tees, 1992).

However, a review of the *actual* infant ages that were predominantly tested in earlier studies shows that the perceptual abilities of newborns and infants of a few months old had not been sufficiently documented. In fact, only a handful of behavioural studies tested the cross-linguistic perception of non-native segmental contrasts in infants aged four months or younger (Eimas et al., 1971; Bosch & Sebastián-Gallés, 2003; Moon, Lagercrantz, & Kuhl, 2013; Polka & Werker, 1994; Streeter, 1976; Trehub, 1976; for studies on young infants' perception of native sounds see Table 1). In spite of the rather limited amount of data, the recurring assumption has been that an infant is a universal listener at and shortly after birth.

Considering that an infant's speech sound perception gradually becomes attuned to her native language within the first few months of life (Kuhl et al., 1992; Yeung & Werker, 2009), and that even a several-minute exposure to tokens of a novel speech sound contrast can alter an eight-month-old's, as well as a two-month-old's, discrimination abilities (Maye, Werker, & Gerken, 2002; Wanrooij, Boersma, & Van Zuijen, 2014), an infant who has already been getting language input for a couple of months is far too old to exhibit the perceptual abilities she had at birth, when she might supposedly have been a universal listener.

The recent literature does test younger infants, including newborns, more often, using neurophysiological measurements. This line of work is reviewed in Section 4. Rather curiously, some of the pioneering literature that focused on very young infants (specifically a comparison of Eimas et al., 1971 and Streeter, 1976) but did not pursue the question of language-generality versus language-specificity, indicates that at one or two months, infants may already no longer be universal, language-general listeners.

3.2 Eimas et al., 1971: no evidence for language-general categorization

Going through the literature on infants' early universal perception, we noticed that the classic work of Eimas et al. (1971) has repeatedly been somewhat misleadingly cited as evidence for infants' initial *language-general* categorical perception. For instance, in the opening lines of their article, Dehaene-Lambertz and Dehaene (1994) attribute to Eimas (1971) the finding that "[y]oung infants can discriminate phonemes even if they are not used in their native language, […]" (p. 292). More recently, Liu and Kager (2016) write that "Infants are born with the ability to discriminate a wide range of native and non-native contrasts regardless of their language background (Eimas, Siqueland, Jusczyk, & Vigorito, 1971)." (p. 336; for a similar attribution see e.g. Gervain & Mehler, 2010; Mazuka, Hasegawa, & Tsuji, 2013; Palmer, Fais, Golinkoff, & Werker, 2012.) In contrast to the above studies, our reading of Eimas et al. (1971) is as follows: what their findings did show is that young infants exhibit categorical perception but what they did not demonstrate – or even assess – is the universality, i.e. language-generality, of that phenomenon.

At this point it seems worthwhile to revisit Eimas et al.'s (1971) experiment, which laid the ground for infant speech perception research, and summarize its design and findings, as reported in the original article. Eimas et al. (1971) tested English-exposed infants aged one and four months in a high-amplitude sucking paradigm on their ability to discriminate between phonologically voiced and voiceless plosives, which are realized in English with a short positive voice onset time (VOT) and with a long positive VOT, respectively. With synthetic CV stimuli, the authors assessed whether the American English-learning infants discriminated between short-VOT and long-VOT stimuli, i.e. between +20 msec and +40 msec, representing English /b/ and /p/, respectively, better than they discriminated between stimuli from the same adult phonemic category, e.g. between -20 msec and 0 msec and between +40 and +60 msec (which we can approximate to a Spanish-like and a Korean-like phonemic distinction, respectively). As a control, the authors also included a no-change condition, testing discrimination of identical stimuli. An increase in the infants' sucking rate following habituation was taken as an index of their ability to discriminate.

For both the one- and the four-month-olds, Eimas et al. detected a *significant* increment in the sucking rate for the *between-category change* and did not find any significant changes in the sucking rate for the within-category change or for the control condition. The authors noted that for the within-category change, the four-month-olds had a non-significant decrement in their sucking rate and the one-month-olds had a non-significant increment, but did not draw any conclusions from those non-significant tendencies. Directly comparing the magnitude of the recovery of the sucking rate between conditions, Eimas et al. found that the between-category change yielded a larger recovery than either of the other two conditions.

The authors concluded that infants perceive a phonetic difference between two consonants before being able to produce them and that they discriminate the contrast according to the American English adult-like phonemic boundaries, indicating that they process the VOT dimension in a categorical manner. Two years after the groundbreaking 1971 study, Eimas and Corbit (1973) speculated that the "categorical nature of the perception of the VOT continuum appears to be universal" (p. 101), supporting their claim by the categorical effects found in adults' perception of the VOT continuum across languages (Lisker & Abramson, 1964), and by their own earlier findings with English-raised one- and four-month-old infants.

On the basis of Eimas et al.'s (1971) data, little can be said about the infants' ability to discriminate non-native, i.e. within-category, differences since that condition gave a null effect. What the data does show is that the infants' ability to perceive VOT differences corresponding to non-native contrasts is either smaller than their ability to do so for the native contrasts or may not even exist at all. Hence, Eimas et al. (1971) should not be taken as evidence for young infants' universal, i.e. language-general, perceptual abilities. Instead, the work of Eimas et al. (1971) should be considered together with the findings of other studies. For instance, Streeter (1976) tested discrimination of VOT differences in two-month-old infants learning Kikuyu, which contrasts pre-voiced, i.e. negative-VOT, and plain unaspirated coronal and dorsal plosives. Streeter found that, unlike Eimas et al.'s (1971) English-learning infants, the Kikuyu infants robustly discriminated a -30-ms negative VOT from a 0-ms VOT in bilabial stops (while still also discriminating the English-like short-lag vs. long-lag 10 ms vs. 40 ms distinction). Lasky et al. (1975) tested discrimination of voicing differences in four-to-six-and-half-month-old infants acquiring Spanish, which is similar to Kikuyu in that it contrasts pre-voiced and plain unaspirated stops. The infants could discriminate relatively large differences in short-lag vs. long-lag VOT, namely +20 vs. +60 ms (similarly to the infants in Eimas et al.'s and Streeter's studies), as well as comparable differences in long vs. short prevoicing, namely -60 vs. -20 ms (somewhat similar to what Streeter found for -30 vs. 0 ms VOT), but failed on a -20 vs. +20 difference (which was most probably the least salient one, although according to the authors it represented the adult phoneme contrast).

To sum up, for some phonetic dimensions, such as the short-lag vs. long-lag VOT of plosives, one may find partially language-general good discrimination of some stimuli, perhaps as a result of the generally high psychoacoustic saliency of post-burst aspiration noise that differentiates short- and long-lag positive VOT (and is also perceived by animals, which do not acquire human language; Kuhl & Miller, 1978; Kuhl & Padden, 1982). For other, potentially less salient phonetic dimensions such as prevoicing, it may be as early as at two months of age that humans show language-specific perception as a result of exposure to categorical variation in pre-burst humming when growing up in a prevoicing-language environment.

3.3 What is measurable in the youngest infants

In order to measure speech sound discrimination abilities accurately, particularly in the youngest infants, special care must be taken to select a testing paradigm appropriate for their young age. If suboptimal methods are chosen, young infants may appear not to be able to discriminate speech stimuli, which may lead to a conclusion of initial poor perceptual sensitivity and/or its independence of the infants' native language.

For instance, developmental studies employ testing paradigms that either may or may not be contingent on infants' looking behaviour. In a paradigm that is contingent on infant looking, the presentation of the auditory stimulus during a trial stops when the infant stops being engaged with the task and looks away from an attention-catching visual display. In a non-contingent paradigm, the duration of the auditory stimulation is not dependent on the infant's looking behaviour or attention. Using a visual fixation paradigm that was *not* contingent on infants' looking behaviour, some studies failed to find discrimination effects in young infants. Narayan, Werker, and Beddor (2010) did not find discrimination of the native /n/-/ŋ/ in Filipino-learning six-to-eight-month-olds. Contrary to that, the experiments carried out by Sundara et al. (2018), which used more age-appropriate tasks contingent on the infant's looking behaviour, demonstrate that very young infants can discriminate such (or even more subtle) speech sound differences, even if they do not occur in the infants' native language.

These conflicting findings indicate that some behavioural methods may not accurately reveal young infants' speech processing capabilities, as they are dependent on the general cognitive and motor abilities of a particular developmental period. In order to pinpoint the age at which infants listen in a language-universal way and avoid any age-dependent infant biases or motor abilities, it might be safest to use behaviour-independent measures, amongst which are the cortical event-related response or the distribution of brain blood oxygenation levels. In the remainder of this section, we review studies that tested infants' discrimination of speech sounds at the level of neural processing.

Studies that tap into infants' neural processing of speech mostly detect perceptual discrimination of virtually any type of contrast in the youngest group of infants that they test. Dehaene-Lambertz and Dehaene (1994) showed that the brains of two-to-three-month-old American English infants promptly detect a change between [ba] and [ga]. Peña, Werker, and Dehaene-Lambertz (2012) demonstrated that both a change from [b̥] to [d̥] and a change from [d̥] to [ɖ̥], roughly corresponding to native Spanish /p/-/t/ and non-native Hindi /d/-/ɖ/ consonantal place contrasts, elicited a strong mismatch response in nine-month-old Spanish infants (while it was only the native contrast that elicited a reliable response in 12-month-olds). The event-related potentials to native and non-native VOT contrasts reported in Rivera-Gaxiola, Silva-Pereyra, and Kuhl (2005) indicate that at seven as well as at 11 months, American English infants show neural discrimination of both a change from [t] to [tʰ] and a change from [d] to [t], roughly corresponding to a native English and a non-native (e.g. Spanish-like) /d/-/t/ voicing contrast, respectively.

An ingeniously designed electrophysiological experiment by Cheour et al. (1998) compared auditory and linguistic vowel perception between six- and 12-month-old infants. At six months, Finnish-learning infants showed discrimination, in terms of the neural mismatch response, of the native Finnish /e/ versus /ø/, as well as of the non-native Estonian /e/ versus /ɤ/, in which the latter pair entails a larger acoustic dissimilarity than the former. The acoustically-based discrimination came to be modulated by the native phonology in the infants' 12th month, when the mismatch response strengthened for the acoustically small but linguistically relevant /e/-/ø/ contrast, and weakened for the acoustically large but non-native

/e/-/ɤ/ contrast. Cheour et al.'s (1998) findings demonstrate that an initial acoustically-moderated sensitivity becomes modulated by the native phonology by 12 months of age, when infants display enhanced perceptual discrimination of native phonemic differences, as well as attenuated discrimination of the non-native ones.

On a related note, early auditory-based discrimination was reported in a recent neurophysiological study by McCarthy, Skoruppa, & Iverson (2019). McCarthy et al. traced the development of British English infants' neural sensitivity to acoustic changes in vowels across multiple native phoneme contrasts, assessing the acoustic change complex (ACC) response recorded with an EEG. The authors showed that the infants' neural discrimination of vowels mimicked the magnitude of the first-formant distances between the stimuli relatively well at a younger age (four months) but by several months later (ten to 11 months) it became less auditory-based and less predictable by the first formant.

The experiments reviewed above show that when neurocognitive or age-appropriate behavioural methods are employed, the youngest infants tested in each study display some perceptual sensitivity to even small speech sound differences, both in vowels and consonants. The young infants' ability to perceptually discriminate stimulus differences seems to occur irrespective of whether these coincide with a native or a non-native phonemic contrast. One could thus conclude from those studies that young infants' discrimination of speech sounds is not specific to their parents' language and may be attributable entirely to general auditory capacities. However, there is emerging evidence which suggests otherwise: the next section highlights several recent experiments with newborn infants whose results suggest that for some types of speech sounds humans might already have begun forming phonological categories (at some undefined level of abstractness) before birth. Consequently, at birth they already perceive those speech sound contrasts in ways specific to their native language environment, i.e. language-specifically. Such recent evidence for early language-specific listening at birth confirms and extends our observation of potential language-specific effects in some of the early studies with very young infants (namely, the comparison of Eimas et al., 1971 and Streeter, 1976 in Section 3.2).

4 Language-specific perception at birth

4.1 Prenatal learning: prosody and segments

Starting in the 1980s, behavioural research has shown that newborn infants recognize and prefer listening to their mother's voice and to their native language over other, unfamiliar languages, and recognize rhymes that they heard during the last weeks of pregnancy (Byers-Heinlein, Burns, & Werker, 2010; DeCasper & Fifer, 1980; DeCasper & Spence, 1986; Mehler et al., 1988; Moon, Cooper, & Fifer, 1993). More recent neurophysiological research confirms the language specialization in the neonate brain. Not only do newborn infants have developed cortical areas that specialize in processing speech in general, as opposed to non-speech (May, Gervain, Carreiras, & Werker, 2017), but they also display language-specific neural attunement, differentially processing their native language as opposed to non-native-language (or otherwise unfamiliar) speech (Sato et al., 2012). Studies with foetuses confirm that the developing individual acquires such linguistic specialization in the prenatal period through exposure to ambient speech signals: at least one month before term foetuses show a preference for their mother's voice and for their native language (Kisilevsky et al., 2009).

In sum, on the basis of studies on the early processing of speech prosody (rhythmic and melodic features of spoken utterances), one could conclude that at birth humans show signs of language-specific listening, as far as entire utterances carrying the global (i.e. suprasegmental) aspects of native language speech are concerned. Much less is still known about language-specificity when it comes to the basic building blocks of speech, i.e. individual speech sound segments. There have been behavioural and neurophysiological studies that investigated segmental speech processing in newborns (Aldridge, Stillman, & Bower, 2001; Cheour-Luhtanen et al., 1995; Kujala et al., 2004) but only a handful of them speak directly to language-specificity in the infants' perception (Moon et al., 2013).

Aldridge et al. (2001) noted the lack of newborn speech perception data and tested infants' vowel perception within two days after birth. Using the preferential sucking paradigm, the authors found that a mixed group of Spanish- and English-exposed newborns perceptually divided up the vowel space into areas somewhat resembling adult-like vowel categories. Moreover, the category-like perceptual discontinuities appeared strongest for [i], followed by [u], then [y], and smallest for [ɯ]. Those discontinuities seemed to be language-conditioned: strongest for [i]-stimuli that occur as vowel phonemes in both of the ambient languages (namely, English and Spanish spoken in Texas) and weakest for [ɯ]-stimuli that do not occur in either English or Spanish. However, the hypothetical language-specificity cannot be unequivocally resolved as a result of a lack of comparison to infants from a language background that does use e.g. [ɯ]-like sounds phonemically. Aldridge et al.'s study indicated category-like discontinuities in vowel perception immediately after birth, leaving the question of universality vs. language-specificity unanswered.

A leap towards uncovering the language-specificity in the speech sound perception of newborns was taken by Moon et al. (2013). They exposed American English and Swedish newborns to tokens of American English /i/ and Swedish /y/ vowels and found that infants from either language sucked on a dummy more when hearing variants of the non-native vowel. The authors explained these findings in terms of newborns' greater perceptual sensitivity to acoustic variation within the non-native vowel category and lesser sensitivity to variation within a native category. To what extent that result alone reflects language-specific *categorical* perception at birth (which the authors claimed) is unclear as the infants' higher sucking reaction to the non-native variant could also have been an effect of a preference for novelty. An objection could be raised concerning the "non-nativeness" of the /i/-stimuli to the Swedish infants, or of the /y/-stimuli to the American English infants. As for the latter, Kuhl et al. (1992) noted that the /y/-stimuli were judged as non-native by American English adults (in or before 1992). However, given the documented /u/-fronting across many varieties of English, including American dialects (see Clopper, Pisoni, & de Jong, 2005), it is somewhat questionable whether the American English participants in the 2013 study would still perceive (all) the [y] sounds as non-native. Moon et al.'s (2013) findings nevertheless indicate that at birth infants may perceive vowels in language-specific ways, an effect that is most probably due to prenatal experience with the segmental level of native-language speech. It now remains to be shown whether infants not only recognize which speech sounds they heard (or did not hear) in the womb but whether they start to create some form of mental representations or categorical warping for the speech sounds that they encounter.

The electrophysiological experiment by Partanen et al. (2013) aimed to directly address controlled prenatal learning of the segmental information in speech. Stimulating foetuses from the 29th week of gestational age with the trisyllabic word *tatata*, occasionally containing a change in the second syllable in terms of vowel quality or of pitch, modulated the individuals' ability to pre-attentively detect some of these changes after birth. Specifically, when compared to a control group, the prenatally trained newborns had a more pronounced neural mismatch response to second-syllable changes in pitch, to which they had been exposed in training, as well as to duration, to which they had not been exposed. The improved processing of pitch changes demonstrates that prenatal learning took place. To explain the improved processing of the unexposed changes in duration the authors proposed generalization of learning (Partanen et al., 2013). Potentially, the newborns could have generalized from familiar pitch contrasts to novel durational contrasts, although it is not yet clear how. Alternatively, the improved processing of duration could also be explained by the foetuses having memorized just the frequent *tatata* stimulus – at birth, they were then able to discriminate any salient deviation from this memorized stimulus. The foetal learning findings of Partanen et al. (2013) extend those of Cheour et al. (2002), who showed with newborns that nocturnal exposure to sequences of isolated vowel stimuli with frequent [i] sounds and infrequent [y] and [ɨ] leads to an increased neural mismatch response to both types of deviant vowels at post-test (as well as to non-trained deviations in pitch). Whether the learning effects in those two studies were driven by a generalization mechanism across various types of speech stimuli (as Partanen et al. conclude) or by a possibly simpler mechanism of mental encoding of the (single) frequent *tatata* stimulus, or the frequent [i]-vowel in Cheour et al., cannot be resolved. Importantly, however, Partanen et al.'s (2013) results demonstrate that some form of *prenatal* learning pertaining to individual speech sound segments took place and modulated the infants' speech processing after birth.

It would be premature to conclude that newborn infants show native-language specialization in speech sound perception on the basis of a single study that indicated language-specific listening as early as at birth (Moon et al., 2013). However, if we also take into account the studies presenting evidence for prenatal and neonatal learning of segmental information (Partanen et al., 2013; Cheour et al., 2002), the scenario in which human near-term foetuses learn the speech sounds of their native language from naturalistic exposure no longer seems improbable. Below we briefly consider what learning mechanism and what kind of input the human individual may have available prenatally that permit speech-sound learning.

4.2 Prenatal speech-sound learning: mechanism and input

As for the exact learning mechanism that might be at work in the earliest stages of life, one could presume that the youngest learner employs bottom-up learning from the ambient stimulus distributions. The availability of such a statistical learning mechanism has been reported in infants as young as two months. Wanrooij et al. (2014) exposed Dutch-raised two-month-olds to bimodal and unimodal probability distributions between [æ]- and [ɛ]-like stimuli representing, respectively, a two-way non-native English contrast /æ/-/ɛ/, and a single non-native category halfway between those two English phonemes. They found that after only 12 minutes of exposure, a change between the English /æ/ and /ɛ/ elicited a larger mismatch response in the bimodally exposed infants than in the unimodally exposed (awake or active-sleeping) ones. Similar effects of distributional exposure have been reported for the [d]-[t] consonantal dimension for slightly older infants, namely, eight-month-old American English listeners in Maye et al. (2002). Note that the eight-month-old infants, or perhaps even the two-month-olds, might already have had some native language speech sound categories (more or less) in place and that these categories might in some way have modulated the distributional leaning process. Additionally, the effects of laboratory exposure on the infants' discrimination of the trained speech sounds were most probably short-term and ceased to exist as a result of further learning from natural language input which the infants experienced immediately after the lab session ended.

All in all, since some learning in Maye et al. and Wanrooij et al. occurred only after a brief, purely bottom-up exposure (and in the latter study in infants who were asleep), it is plausible it was primarily a bottom-up distributional learning mechanism that was at play in the two-month-old and eight-month-old infants. Moreover, in scenarios where both top-down and bottom-up information is available but unaligned, listeners learn more robustly from the low-level stimulus statistics (Emberson, Liu, & Zevin, 2013). Arguably, bottom-up learning from natural language auditory input could also be the mechanism which is active from the very earliest – prenatal – stages of exposure to speech sounds. Whether and to what extent the learning would apply for a particular speech sound category (or contrast) would depend on whether the relevant acoustic information reaches the foetus (and how much of it). That is, the intrauterine learnability of individual speech categories would be conditioned by their auditory discernibility.

As the speech signal, uttered *ex utero*, passes through a pregnant woman's abdomen and amniotic fluid to the foetal inner ear, it undergoes modulations, especially in the spectral domain. In general, frequencies between 100 Hz and 1000 Hz can reach the foetus virtually unmodified, while higher frequencies become progressively attenuated (Granier-Deferre, Ribeiro, Jacquet, & Bassereau, 2011; Lecanuet & Schaal, 1996). Being heard veridically *in utero,* speech sound contrasts realized in the low-frequency range are more likely to be heard and learnt prenatally than contrasts realized in the high-frequency range. Vowels, unlike most consonants, often contain primary distinguishing information at low frequencies (the first formant that cues the identity of many vowel categories ranges between 250 Hz and 1000 Hz in women's speech; e.g. Adank, Van Hout, & Smits, 2004), which, together with their greater loudness in general, makes them ideal candidates for prenatal learning (Granier-Deferre et al., 2011; Tsuji & Cristia, 2014).

Studies do indeed show that foetuses at a gestational age of 35 weeks and older can discriminate a vocalic [a]-[i] distinction (embedded in a mono- or disyllabic CVCV frame) but do not discriminate the consonantal [da]-[ta] distinction (Lecanuet et al., 1987; Shahidullah & Hepper, 1994; Weikum, Oberlander, Hensch, & Werker, 2012). At birth, vowel differences continue to be processed robustly, also when directly compared to consonantal differences: Benavides-Varela, Hochmann, Macagno, Nespor, and Mehler (2012) show that newborns have a cortical novelty response to a change between vowels in [lala] vs. [lili] but do not have such a response to a change in the consonantal makeup of [lala] vs. [tata] (although pre-term infants born at 29-32 weeks were able to discriminate the consonantal [ba]-[ga] distinction in a different study; Mahmoudzadeh et al., 2013). There is, of course, some variation within the class of vowels and within the class of consonants as to which frequency range cues a particular category: some vowels might be cued by frequencies above 2500 Hz (e.g. /i/-/y/ in French; Schwartz & Escudier, 1989) and some consonants by frequencies as low as 1000 Hz (e.g. /p/-/p^{j}/ or /l/-/ɫ/ in Irish or Russian; Chiosáin & Padgett, 2012; Recasens, 2012). The proposal that vowels are more likely than consonants to be acquired prenatally should be understood as a generalization and may not apply to every single speech sound category.

Apart from spectral properties, in some languages speech sound categories are also cued by pitch changes (lexical tone) or by durational information (phonemic length). For instance, languages such as Finnish, Japanese, Czech, or Hungarian contrast vowels by duration, with a short and a long segment of the same spectral quality representing two different phonemes. Unlike spectral information, temporal information is transmitted to the foetus in an unchanged form (Querleu, Renard, Versyp, Paris-Delrue, & Crèpin, 1988), which leads to a testable hypothesis that length contrasts would be more probably or robustly acquired prenatally than (higher-frequency) spectral contrasts (see Burnham, 1986 who, similarly, proposed a potential developmental precedence of duration-cued contrasts), provided that their frequency of occurrence allows distributional learning (which may not be the case for e.g. Japanese; Bion, Miyazawa, Kikuchi, & Mazuka, 2013). In that respect, Thiede et al.'s (2019) experiment with Finnish newborns suggests that typically developing newborns have more mature neural discrimination responses to a change in vowel duration [tata] vs. [tata:] and to a pitch change (at around 170 Hz) than to a change in vowel quality [tata] vs. [tato]. In our view, this result could be attributed either to the greater saliency of the durational and pitch differences in general or to an earlier, prenatal, onset of learning for the durational and pitch contrasts over the spectral contrast, or to both.

It appears that at least for some speech sounds, the formation of categories could already begin in the prenatal period as soon as the central auditory system is sufficiently developed (i.e. from gestational week 28; Graven & Browne, 2008). There are various paths that the prenatal development of speech perception could take: it could involve the desensitization of the initially almost-perfect perception of speech sounds that do not occur in the environment at all, attunement to speech sounds that occur frequently, or a combination of both (in line with the established theories of speech sound development by Aslin and Pisoni, 1980 and Aslin, Werker, and Morgan, 2002, as well as with recent proposals for the prenatal learning of speech suprasegmentals; Abboub, Nazzi, & Gervain, 2016). The speech sound properties that are most available to prenatal learning are those that propagate well to the foetal ear. In that respect, vocalic contrasts cued by formant frequencies up to 1000 Hz (which occur in all languages), such as the Dutch or English /ɪ/-/ɛ/, or vowel contrasts cued by duration or pitch contour (occurring phonemically in some languages), will most probably begin to be learned prenatally, unlike, for instance, consonantal contrasts cued by high-frequency frication noise such as the Greek or English /s/-/θ/. At birth, when the infant encounters all the auditory dimensions of speech, learning will also begin for the ambient sounds that she has not heard prenatally.

In this section we have summarized studies that support a prenatal onset of speech sound learning through exposure. Many exciting avenues are open for future research into the initial stages of speech sound acquisition. Neurophysiological experiments with newly-born full- and pre-term infants from at least two different language backgrounds could help uncover the earliest stages of speech sound learning. Studies could examine when input-triggered speech sound acquisition begins, whether it occurs prenatally at all, for which types of sounds, what mechanisms allow potential prenatal learning, and how the learning mechanisms and input unfold once the individual's environment changes after birth.

5 Areas not covered in the present review

Our review of young infants' discrimination abilities focused primarily on speech sound distinctions in terms of vowel and consonant quality, which are types of contrasts that occur in all of the world's languages and, as such, have also been most widely researched. A very interesting, but to date underinvestigated area is the acquisition of tone and phonemic length contrasts, which are less frequent cross-linguistically than vocalic and consonantal quality. A recent comprehensive review of the development of tone and length contrasts is presented in Narayan (2020). The limited research in the area of tone and vowel length contrasts indicates atypical developmental curves, delayed, speeded-up, and sometimes U-shaped perceptual reorganization, with a dip in discrimination of native tone or length contrasts halfway through the infant's development (Mattock, Molnar, Polka, & Burnham, 2008; Minagawa-Kawai, Mori, Naoi, & Kojima, 2007; Mugitani et al., 2009; Sato, Sogabe, & Mazuka, 2010a; Sato, Sogabe, & Mazuka, 2010b; Yeung, Chen, & Werker, 2013). Given that they are acoustically realized in the low-frequency range (namely, by fundamental frequency and duration), tone and length contrasts are likely to start being acquired prenatally (see Granier-Deferre et al., 2011; Partanen et al., 2013). However, since fundamental frequency and duration are major cues to prosodic patterns, in certain languages infants might come to notice a conflict between the segmental and suprasegmental functions, which could cause e.g. vowel length contrasts to develop differently compared to vowel quality contrasts.

Another area that was left out of the present review is theoretical frameworks that address first language speech acquisition, such as the Native Language Magnet model (Kuhl, 2004), the framework for Processing Rich Information from Multidimensional Interactive Representations (Werker & Curtin, 2005), or the Perceptual Assimilation Model (Best, 1994). The Natural Referent Vowel (NRV) framework (Polka & Bohn, 2011) is one of the theories that focuses on infants' initial perceptual abilities, arguing that speech perception in the first stages of development reflects language-general auditory asymmetries. The NRV seems to place the pre-linguistic, language-general, stage at about six to nine months, which – if prenatal experience does modulate infants' early perception – could already be too late for language-general asymmetries to be observable, at least for some speech sound contrasts in some languages. With a well-thought-out stimulus and population design, this supposition is testable.

Computational models of the development of speech perception are not covered in the present review, although many of their outcomes are quite compatible with the idea of an auditory-based, possibly prenatal, initial stage of acquisition, as well as with the proposed distributional learning mechanism (Guenther & Gjaja, 1996; Vallabha, McClelland, Pons, Werker, & Amano, 2007). Seebach, Intrator, Lieberman, and Cooper (1994) aimed to show that speech sound feature categories (such as the consonantal place of articulation) do not need to be innately specified and can be learned prenatally. Seebach et al. trained a neural network with intrauterine-like realizations of the English /pa/, /ta/, and /ka/, and showed that on the basis of the auditory input alone, the virtual foetus learned to distinguish three classes of plosives: labial [p]-like, coronal [t]-like, and dorsal [k]-like. A more in-depth review of modelling work could help provide further insights into the plausible initial stages, as well as learning mechanisms, associated with early linguistic development.

6 Suggestions for a future research agenda

As pointed out in Section 2.5, within-category discrimination – the attenuation of which was typically thought to reflect adult-like representations – does not necessarily have to decrease as the native language speech faculty matures (Galle & McMurray, 2014). Additionally, infant testing paradigms may not always unequivocally reflect whether and to what extent an infant perceptually categorizes speech stimuli according to the sounds' phonological roles in the infants' native language (see Mulak et al., 2017).

With respect to the question of when in their development infants become language-specific listeners, we advise studies to employ paradigms that assess infants' perceptual discrimination across a variety of speech sound distinctions in such a way that some of the distinctions represent a phonological contrast in the infants' ambient language and others do not. Crucially, the auditory differences within all types of speech sound pairs should be psychoacoustically comparable whenever possible (e.g. distances measured on relative instead of absolute scales). The comparisons of perceptual sensitivities to native and non-native contrasts should be performed either with behavioural paradigms that are appropriate for all the age groups involved and/or with neurophysiological methods. Critically, to be able to draw firm conclusions about language-specificity, one would ideally test infants from at least two different language backgrounds.

A hypothetical experiment could test the observation we made concerning Eimas et al. (1971) and Streeter (1976) that the perception of voicing contrasts starts to be language-specific quite early in development, sometime between the second and fourth months of age. English- and Kikuyu-like infants – perhaps German and Dutch, respectively – could be tested on their perception of voice-onset-time distinctions in plosives, e.g. on [b]-[p] and [p]-[p^h] stimulus pairs. One could test newborns and, for instance, five-month-olds. *In utero*, the distinctions [b]-[p] and [p]-[p^h] are probably not very audible, which means that at birth, a German and a Dutch infant should not differ from one another in how they discriminate them. If infants pick up the native-language voicing contrasts quickly – perhaps within a few months after birth, then at about five months of age, the hypothetical German infant should outperform the hypothetical Dutch infant in her discrimination of [p]-[p^h] and, conversely, the Dutch infant should outperform the German one on [b]-[p]. Such a result would be in line with the indirectly observed difference between the very young infants exposed to English (in Eimas et al.) and to Kikuyu (in Streeter). Considering the newborn age proposed here, a suitable method for assessing perceptual sensitivity could be the electrophysiological index of discrimination, the neural mismatch response.

To test whether speech sound learning begins before birth, the field should attempt to focus on the earliest age possible. Currently available techniques such as electroencephalography and functional near-infrared spectroscopy have, in recent years, come to be widespread lab equipment in the field and are very well suited to auditory experiments with newborn infants. An intriguing area that seemed to have boomed temporarily with the spread of obstetric ultrasonography is foetal perception. Unfortunately, since newborns and, especially, foetuses are a rather difficult population to access, cross-linguistic studies seem somewhat hard to envision. To compensate, one could, for foetuses and newborns within a single language group, compare the development of native, non-native, and comparable non-speech perception. The comparison of speech and non-speech processing in near-term foetuses (e.g. some time after the 28th week of gestation) would show whether, in the initial stage, humans process speech differently from comparable non-speech stimuli, that is, whether there is anything like universal perceptual categorization for speech (e.g. Kuhl et al., 1992, which we argued to be unlikely). Tracing the development of native against non-native speech perception up until birth could then reveal whether prenatal auditory exposure to the ambient language leads to language-specific perceptual warping of the auditory sensitivities.

The last point, and a most promising one, on the agenda for future research that we would like to bring up is the combination of mathematical and computational models with laboratory human studies. Especially because of the very sensitive population, it is vital that experimenters make specific and informed predictions about the foetal and neonate population. Those can be created by a priori building a multilayer model that, on the input level, simulates the physical properties of the auditory signal that reaches the foetal ear, and also, at the 'ear and brain' level approximates as closely as possible the foetal peripheral and central auditory system (as was done in Seebach et al., 1994). Feeding the model with stimuli that mimic the natural language environment of the developing individual can help create informed hypotheses on how the perceptual organization of speech sounds unfolds in its earliest stages.

7 Conclusion

The aim of this paper was to revisit the universal listener, a concept that is often referred to in the developmental literature but is much less often unambiguously defined. A review of the findings to date suggests that what characterizes the initial stages of an infant's language development is not universal *categorical* perception of speech sound contrasts attested in the languages of the world but instead a fine perceptual sensitivity to speech sounds' auditory properties. Arguably, the extent to which the young universal listener discriminates a particular phonetic difference might be entirely dependent on the psychophysical distance between the stimuli to be discriminated and on the perceptual saliency of the auditory dimension itself. Attempting to pinpoint the age at which an individual would exhibit such an accurate, auditory-based discrimination led to the revelation that even at birth, for some speech sounds at least, infants might already listen in a language-specific way. A discussion of the prenatal speech sound hearing and learning abilities reveals that it could well be in the prenatal period that a human starts learning about the segmental speech categories that make up her language.

The overall conclusion is that the stage at which humans display entirely language-general, i.e. universal, speech sound perception is most probably not at birth but earlier, at some point between the functional completeness of the temporal cortex in the 28th week of gestation and birth. It remains to be tested at which gestational age speech sound category formation begins, which speech sound contrasts it affects, and whether it entails sensitization, desensitization, or both. The field for speech perception experiments with newborns and foetuses is wide open and, thanks to the techniques now available, the studies to come are bound to bring us closer to understanding when and how a human starts learning from the surrounding speech signal and thereby sets out to be a native speaker of her language.

Study	Method	Native language	Age	Stimuli		Outcome
Aldridge et al. (2001, Exp. 3)	HAS	Mixed Spanish/English	0	[i]-[i/y]		
				[y]-[i/y]		
				[u]-[u/ɯ]		
				[ɯ]-[u/ɯ]		
Benavides-Varela et al. (2012)	fNIRS	Italian	0	[lala]-[lili]		
				[lala]-[tata]		
Bosch & Sebastián-Gallés (2003)	HPP (cntg.)	Spanish	4	[deði]-[dɛði]		
			8			
		Catalan	4	[deði]-[dɛði]		L1
			8			L1
		Bilingual Spanish & Catalan	4	[deði]-[dɛði]		L1
			8			L1
			12			L1
Cheour-Luhtanen et al. (1995)	EEG	Finnish	0	[i]-[y]		L1
				[i]-[ɨ]		
Dehaene-Lambertz (2000)	EEG	French	4	[ba]-[ga]		L1
Dehaene-Lambertz & Dehaene (1994)	EEG	Am. English	2–3	[ba]-[ga]		L1
Eimas et al. (1971)	HAS	*Am. English*	1–4	[ba]-[pʰa]	-20 vs. 0	
					+40 vs. +60	
					+20 vs. +40	L1
Eimas et al. (1974)	HAS	*Am. English*	2-3	[dæ]-[gæ]		L1
				[bæ]-[dæ]		L1
Eilers (1977)	HAS	Am. English	1–4.5	[sa]-[za]		L1
				[as]-[aːz]		L1
				[aːs]-[aːz]		
				[at]-[aːd]		L1
				[aːt]-[aːd]		
				[at]-[aːt]		
Lacerda (1992)	HAS	Swedish	0–6	[bɑ]-[ba]		L1
				[bɑ]-[bɔ]		L1
Lasky et al. (1975)	ECG	Spanish	4–6.5	[b̬a]-[pʰa]	-60 vs. -20	
					-20 vs. +20	(L1)
					+20 vs. +60	
Lecanuet et al. (1987)	CTG	*French*	35–38 wGA	[babi]-[biba]		L1
Levitt et al. (1988)	HAS	English	2	[fa]-[θa]		L1
				[va]-[ða]		L1
				[af]-[aθ]		L1
Kuhl & Miller (1982)	HAS	*Am. English*	1–4	[a]-[i]		L1
Kujala et al. (2004)	MEG	Finnish	0	[aː]-[iː]		L1
Mahmoudzadeh et al. (2013)	fNIRS	*French*	0 (pre-term)	[ba]-[ga]		L1
Moon et al. (2013)	HAS	Am. English & Swedish	0	changes in: [i], [y]	within native	L1
					within non-native	
Narayan et al. (2010)	CF (non-cntg.)	Filipino	6–8	[na]-[ŋa]		L1
			10–12			L1
		English	4–5	[ma]-[na]		L1
				[na]-[ŋa]		
			6–8	[ma]-[na]		L1
				[na]-[ŋa]		
			10–12	[ma]-[na]		L1

				[na]-[ŋa]		
Polka & Werker (1994, Exp. 1)	CF	English	4	[dit]-[dat]		L1
				[dyt]-[dut]		
				[dɤt]-[dʊt]		
			6	[dit]-[dat]		L1
				[dyt]-[dut]		
				[dɤt]-[dʊt]		
Sebastián-Gallés & Bosch (2009)	HPP (cntg.)	Monolingual Spanish or Catalan	4	[doði]-[duði]		L1
				[deði]-[duði]		L1
			8	[doði]-[duði]		L1
				[deði]-[duði]		L1
		Bilingual Spanish & Catalan	4	[doði]-[duði]		L1
				[deði]-[duði]		L1
			8	[doði]-[duði]		L1
				[deði]-[duði]		L1
			12	[doði]-[duði]		L1
Shahidullah & Hepper (1994)	USG	English	27 wGA	[ba]-[bi]		L1
			35 wGA			L1
Streeter (1976)	HAS	Kikuyu	2	[ba]-[pʰa]	-30 vs. 0	(L1)
					+10 vs.+40	
					+50 vs. +80	
Sundara et al. (2018)	CF (cntg.)	English	4	[na]-[ŋa]		
			6	[na]-[ŋa]		
				[n̪a]-[ɳa]		
				[la]-[ɭa]		
		French	6	[n̪a]-[ɳa]		
				[la]-[ɭa]		
Swoboda, Morse, & Leavitt (1976)		Am. English	2	[i]-[ɪ]		L1
				within [i], within [ɪ]		
Trehub (1973)	HAS	*Can. English*	1-4	[a]-[i]		L1
				[i]-[u]		L1
				[pa]-[pi]		L1
				[ta]-[ti]		L1
Trehub (1976)	HAS	Can. English	2–4	[pa]-[pã]		
		Can. Eng./French		[ʒa]-[r̝a]		
Weikum et al. (2012)	ECG	*Can. English*	36 wGA	[d̥a]-[tʰa]		L1
				[a]-[i]		L1

Table 1: A summary of studies that assessed (native or non-native) vowel or consonant discrimination in foetuses or infants younger than six months. The list of studies is meant as a representative, though not exhaustive, overview of speech sound discrimination in very young populations. Outcome column: shading denotes evidence for discrimination; "L1" marks a native (-like) contrast. Age in months unless stated as wGA (weeks of gestation age). Methods: CF = central fixation, cntg. = contingent on infant looking, CTG = cardiotocography, ECG = electrocardiography, EEG = electroencephalography, fNIRS = functional near-infrared spectroscopy, HAS = high-amplitude sucking, HPP = headturn preference procedure, MEG = magnetoencephalography, USG = ultrasonography. Language in italics if not specifically named in the cited study.

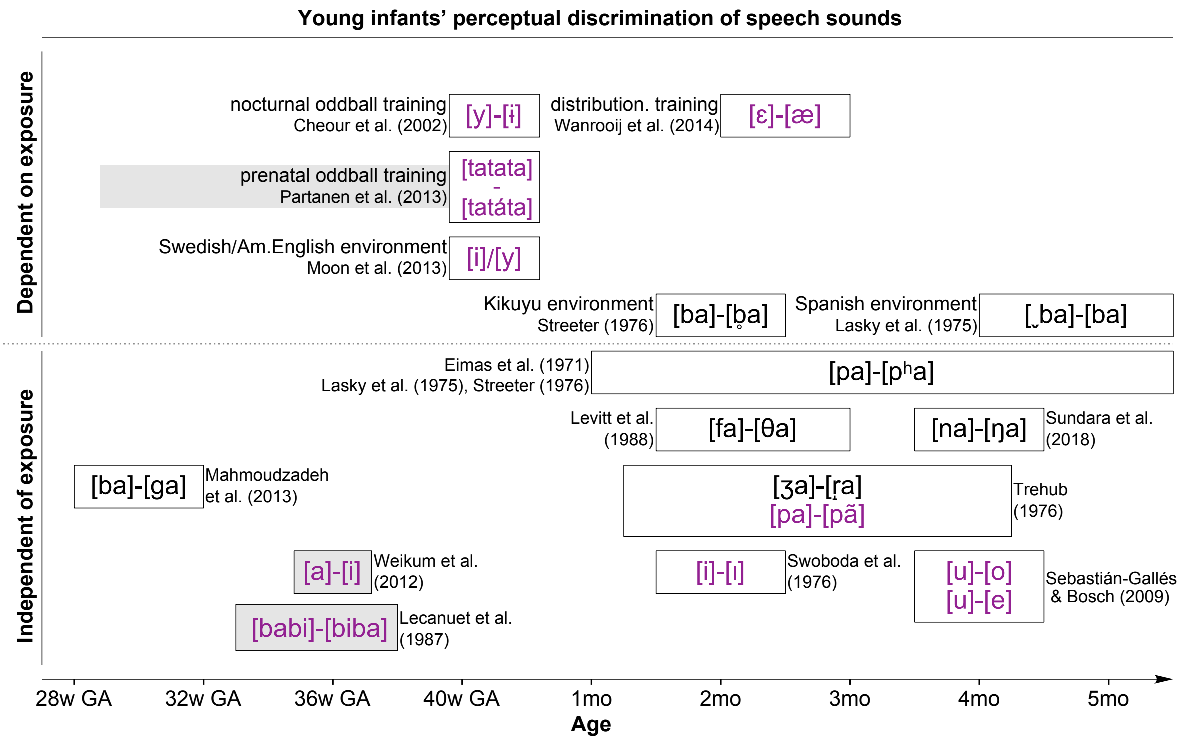

Figure 1: A timeline sketching selected speech sound discrimination abilities before six months of age. The lower part shows discrimination independent of native-language or training exposure, the upper part shows discrimination dependent on prior exposure. Vowel distinctions in purple, consonantal ones in black. Grey shading shows intrauterine perception or training. The width of the rectangles spans the infant ages at which discrimination of the depicted contrasts has been attested by the respective studies. The figure shows that from very early on, humans are able to distinguish not only large but also relatively small differences in vowels and in consonants. Note that the timeline does not contain all discriminable contrasts and does not show failures to discriminate.

References

Abel, S. M. (1972). Duration Discrimination of Noise and Tone Bursts. *The Journal of the Acoustical Society of America, 51*(4B), 1219–1223. doi:10.1121/1.1912963

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

Adank, P., van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical Society of America, 116*(3), 1729–1738. doi:10.1121/1.1779271

Aldridge, M. A., Stillman, R. D., & Bower, T. G. R. (2001). Newborn categorization of vowel-like sounds. *Developmental Science, 4*(2), 220–232. doi:10.1111/1467-7687.00167

Aslin, R. N., & Pisoni, D. B. (1980). Some Developmental Processes in Speech Perception. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology: Perception and Production* (pp. 67-96). New York: Academic Press.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of Voice Onset Time by Human Infants: New Findings and Implications for the Effects of Early Experience. *Child Development*, *52*(4), 1135-1145. doi:10.2307/1129499

Aslin, R. N., Werker, J. F., & Morgan, J. L. (2002). Innate phonetic boundaries revisited (L). *The Journal of the Acoustical Society of America, 112*(4), 1257–1260. doi:10.1121/1.1501904

Baru, A. V. (1975). Discrimination of Synthesized Vowels [a] and [i] with Varying Parameters (Fundamental Frequency, Intensity, Duration and Number of Formants) in Dog. In G. Fant & M. A. A. Tatham (Eds.), *Auditory Analysis and Perception of Speech* (pp. 91–101). New York: Academic Press. doi:10.1016/b978-0-12-248550-3.50010-6

Benavides-Varela, S., Hochmann, J.-R., Macagno, F., Nespor, M., & Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences*, *109*(44), 17908–17913. doi:10.1073/pnas.1205413109

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA, US: The MIT Press.

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

Bion, R. A. H., Miyazawa, K., Kikuchi, H., & Mazuka, R. (2013). Learning Phonemic Vowel Length from Naturalistic Recordings of Japanese Infant-Directed Speech. *PLoS One, 8*(2), e51594. doi:10.1371/journal.pone.0051594

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life. *Language and Speech, 46*(2-3), 217–243. doi:10.1177/00238309030460020801

Burdick, C. K., & Miller, J. D. (1975). Speech perception by the chinchilla: discrimination of sustained [a] and [i]. *The Journal of the Acoustical Society of America, 58*(2), 415–427. doi:10.1121/1.380686

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics, 7*(3), 207–239. doi:10.1017/s0142716400007542

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics, 28*(3), 455–474. doi:10.1017/s0142716407070257

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The Roots of Bilingualism in Newborns. *Psychological Science, 21*(3), 343–348. doi:10.1177/0956797609360758

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*(5), 351–353. doi:10.1038/1561

Cheour, M., Martynova, O., Näätänen, R., Erkkola, R., Sillanpää, M., Kero, P., … Hämäläinen, H. (2002). Speech sounds learned by sleeping newborns. *Nature, 415*(6872), 599–600. doi:10.1038/415599b

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., … Näätänen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research, 82*(1), 53–58. doi:10.1016/0378-5955(94)00164-l

Chiosáin, M. N., & Padgett, J. (2012). An acoustic and perceptual study of Connemara Irish palatalization. *Journal of the International Phonetic Association, 42*(2), 171–191. doi:10.1017/s0025100312000059

Chládková, K., Escudero, P., & Boersma, P. (2011). Context-specific acoustic differences between Peruvian and Iberian Spanish vowels. *The Journal of the Acoustical Society of America, 130*(1), 416–428. doi:10.1121/1.3592242

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America, 118*(3), 1661–1676. doi:10.1121/1.2000774

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176. doi: 10.1126/science.7375928

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*(2), 133–150. doi: 10.1016/0163-6383(86)90025-1

Dehaene-Lambertz, G. (2000). Cerebral Specialization for Speech and Non-Speech Stimuli in Infants. *Journal of Cognitive Neuroscience, 12*(3), 449–460. doi:10.1162/089892900562264

Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature, 370*(6487), 292–295. doi:10.1038/370292a0

Dewson, J. H. (1964). Speech Sound Discrimination by Cats. *Science, 144*(3618), 555–556. doi:10.1126/science.144.3618.555

Eilers, R. E. (1977). Context-sensitive perception of naturally produced stop and fricative consonants by infants. *The Journal of the Acoustical Society of America, 61*(5), 1321–1336. doi:10.1121/1.381435

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental Changes in Speech Discrimination in Infants. *Journal of Speech and Hearing Research, 20*(4), 766–780. doi:10.1044/jshr.2004.766

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics, 16*(3), 513–521. doi:10.3758/bf03198580

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4*(1), 99–109. doi:10.1016/0010-0285(73)90006-6

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science, 171*(3968), 303–306. doi:10.1126/science.171.3968.303

Emberson, L. L., Liu, R., & Zevin, J. D. (2013). Is statistical learning constrained by lower level perceptual organization? *Cognition, 128*(1), 82–102. doi:10.1016/j.cognition.2012.12.006

Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. A. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America, 126*(3), 1379–1393. doi:10.1121/1.3180321

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., … Yurovsky, D. (2017). A Collaborative Approach to Infant Research: Promoting Reproducibility, Best Practices, and Theory-Building. *Infancy, 22*(4), 421–435. doi:10.1111/infa.12182

Galle, M. E., & McMurray, B. (2014). The development of voicing categories: A quantitative review of over 40 years of infant speech perception research. *Psychonomic Bulletin & Review, 21*(4), 884–906. doi:10.3758/s13423-013-0569-y

Gervain, J., & Mehler, J. (2010). Speech Perception and Language Acquisition in the First Year of Life. *Annual Review of Psychology, 61*(1), 191–218. doi:10.1146/annurev.psych.093008.100408

Gervain, J., & Werker, J. F. (2008). How Infant Speech Perception Contributes to Language Acquisition. *Language and Linguistics Compass, 2*(6), 1149–1170. doi:10.1111/j.1749-818x.2008.00089.x

Granier-Deferre, C., Ribeiro, A., Jacquet, A.-Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science, 14*(2), 336–352. doi:10.1111/j.1467-7687.2010.00978.x

Graven, S. N., & Browne, J. V. (2008). Auditory Development in the Fetus and Infant. *Newborn and Infant Nursing Reviews, 8*(4), 187–193. doi:10.1053/j.nainr.2008.10.010

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America, 100*(2), 1111–1121. doi:10.1121/1.416296

Holmberg, T. L., Morgan, K. A., & Kuhl, P. K. (1977). Speech perception in early infancy: discrimination of fricative consonants. *The Journal of the Acoustical Society of America, 62*(S1), S99–S99. doi:10.1121/1.2016488

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception & Psychophysics, 72*(5), 1218–1227. doi:10.3758/app.72.5.1218

Junge, C., Boll-Avetisyan, N., & Benders, T. (2019). Speech perception and discrimination: from sounds to words. In J. S. Horst & J. von Koss Torkildsen (Eds.), *International handbook of language acquisition* (pp. 153–172). (Routledge International Handbooks). New York, NY: Routledge, Taylor and Francis Group. doi:10.4324/9781315110622-9

Jusczyk, P. W. (1995). Language Acquisition: Speech Sounds and the Beginning of Phonology. In J. L. Miller & P. D. Eimas (Eds.), *Speech, Language, and Communication* (pp. 263–301). San Diego, CA, US: Academic Press. doi:10.1016/B978-012497770-9.50010-8

Kisilevsky, B. S., Hains, S. M. J., Brown, C. A., Lee, C. T., Cowperthwaite, B., Stutzman, S. S., … Wang, Z. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behavior and Development, 32*(1), 59–71. doi:10.1016/j.infbeh.2008.10.002

Kronrod, Y., Coppess, E., & Feldman, N. H. (2016). A unified account of categorical effects in phonetic perception. *Psychonomic Bulletin & Review, 23*(6), 1681–1712. doi:10.3758/s13423-016-1049-y

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America, 66*(6), 1668–1679. doi:10.1121/1.383639

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831–843. doi:10.1038/nrn1533

Kuhl, P., Conboy, B., Padden, D., Nelson, T., & Pruitt, J. (2005). Early Speech Perception and Later Language Development: Implications for the "Critical Period." *Language Learning and Development, 1*(3), 237–264. doi:10.1207/s15473341lld0103&4_2

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America, 63*(3), 905–917. doi:10.1121/1.381770

Kuhl, P. K., & Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception & Psychophysics, 31*(3), 279–292. doi:10.3758/bf03202536

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics, 32*(6), 542–550. doi:10.3758/bf03204208

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. doi: 10.1111/j.1467-7687.2006.00468.x

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*(5044), 606–608. doi:10.1126/science.1736364

Kujala, A., Huotilainen, M., Hotakainen, M., Lennes, M., Parkkonen, L., Fellman, V., & Näätänen, R. (2004). Speech-sound discrimination in neonates as measured with MEG. *NeuroReport, 15*(13), 2089–2092. doi:10.1097/00001756-200409150-00018

Lacerda, F. (1992). Young infants prefer high/low vowel contrasts. *PERILUS*, *15*, 85–90.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT Discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, *20*(2), 215–225. doi: 10.1016/0022-0965(75)90099-5

Lecanuet, J.-P., Granier-Deferre, C., Decasper, A., Maugeais, R., Andrieu, J. A., & Busnel, C. (1987). Fetal perception and discrimination of speech stimuli; demonstration by cardiac reactivity; preliminary results. *Comptes Rendus De L'Academie Des Sciences. Serie III, Sciences de la Vie, 305*(5), 161–164.

Lecanuet, J.-P., & Schaal, B. (1996). Fetal sensory competencies. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, *68*, 1–23. doi: 10.1016/0301-2115(96)02509-2

Levitt, A., Jusczyk, P. W., Murray, J., & Carden, G. (1988). Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 361-368. doi: 10.1037/0096-1523.14.3.361

Lim, S.-J., Fiez, J. A., & Holt, L. L. (2019). Role of the striatum in incidental learning of sound categories. *Proceedings of the National Academy of Sciences, 116*(10), 4671–4680. doi:10.1073/pnas.1811992116

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD, 20*(3), 384–422. doi:10.1080/00437956.1964.11659830

Liu, L., & Kager, R. (2016). Perception of a native vowel contrast by Dutch monolingual and bilingual infants: A bilingual perceptual lead. *International Journal of Bilingualism, 20*(3), 335–345. doi:10.1177/1367006914566082

Maddieson, I. (1986). The size and structure of phonological inventories: Analysis of UPSID. In J. J. Ohala (Ed.), *Experimental Phonology* (pp. 105–123). New York: Academic Press.

Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S., Dubois, J., … Wallois, F. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proceedings of the National Academy of Sciences, 110*(12), 4846–4851. doi:10.1073/pnas.1212220110

Marean, G. C., Werner, L. A., & Kuhl, P. K. (1992). Vowel categorization by very young infants. *Developmental Psychology, 28*(3), 396–405. doi:10.1037/0012-1649.28.3.396

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106*(3), 1367–1381. doi:10.1016/j.cognition.2007.07.002

Maurer, D., & Werker, J. F. (2013). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178. doi: 10.1002/dev.21177

May, L., Gervain, J., Carreiras, M., & Werker, J. F. (2017). The specificity of the neural response to speech at birth. *Developmental Science, 21*(3), e12564. doi:10.1111/desc.12564

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101–B111. doi:10.1016/s0010-0277(01)00157-3

Mazuka, R., Hasegawa, M., & Tsuji, S. (2013). Development of non-native vowel discrimination: Improvement without exposure. *Developmental Psychobiology, 56*(2), 192–209. doi:10.1002/dev.21193

McCarthy, K. M., Skoruppa, K., & Iverson, P. (2019). Development of neural perceptual vowel spaces during the first year of life. *Scientific Reports, 9*, 19592. doi:10.1038/s41598-019-55085-y

McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology*, *27*(4), 1546–1571. doi: 10.1044/2018_ajslp-17-0100

McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition, 95*(2), B15–B26. doi:10.1016/j.cognition.2004.07.005

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33–B42. doi:10.1016/s0010-0277(02)00157-9

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

Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007). Neural Attunement Processes in Infants during the Acquisition of a Language-Specific Phonemic Contrast. *Journal of Neuroscience, 27*(2), 315–321. doi:10.1523/jneurosci.1984-06.2007

Molemans, I., van den Berg, R., Van Severen, L., & Gillis, S. (2012). How to measure the onset of babbling reliably? *Journal of Child Language*, *39*(3), 523–552. doi: 10.1017/s0305000911000171

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*(4), 495–500. doi:10.1016/0163-6383(93)80007-u

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experiencedin uteroaffects vowel perception after birth: a two-country study. *Acta Paediatrica, 102*(2), 156–160. doi:10.1111/apa.12098

Mugitani, R., Pons, F., Fais, L., Dietrich, C., Werker, J. F., & Amano, S. (2009). Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology, 45*(1), 236–247. doi:10.1037/a0014043

Mulak, K. E., Bonn, C. D., Chládková, K., Aslin, R. N., & Escudero, P. (2017). Indexical and linguistic processing by 12-month-olds: Discrimination of speaker, accent and vowel differences. *PLoS One, 12*(5), e0176762. doi:10.1371/journal.pone.0176762

Nakisa, R. C., & Plunkett, K. (1998). Evolution of a Rapidly Learned Representation for Speech. *Language and Cognitive Processes, 13*(2-3), 105–127. doi:10.1080/016909698386492

Narayan, C. R. (2019). An acoustic perspective on 45 years of infant speech perception, Part 1: Consonants. *Language and Linguistics Compass, 13*(10), e12352. doi:10.1111/lnc3.12352

Narayan, C. R. (2020). An acoustic perspective on 45 years of infant speech perception. II. Vowels and suprasegmentals. *Language and Linguistics Compass*, e12369. doi:10.1111/lnc3.12369

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*(3), 407–420. doi:10.1111/j.1467-7687.2009.00898.x

Palmer, S. B., Fais, L., Golinkoff, R. M., & Werker, J. F. (2012). Perceptual Narrowing of Linguistic Sign Occurs in the 1st Year of Life. *Child Development, 83*(2), 543–553. doi:10.1111/j.1467-8624.2011.01715.x

Partanen, E., Kujala, T., Naatanen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences, 110*(37), 15145–15150. doi:10.1073/pnas.1302159110

Peña, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier Speech Exposure Does Not Accelerate Speech Acquisition. *Journal of Neuroscience, 32*(33), 11159–11163. doi:10.1523/jneurosci.6516-11.2012

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13*(2), 253–260. doi:10.3758/bf03214136

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America, 89*(6), 2961–2977. doi:10.1121/1.400734

Polka, L., & Bohn, O.-S.. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America, 100*(1), 577–592. doi:10.1121/1.415884

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics, 39*(4), 467–478. doi:10.1016/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*(2), 421–435. doi:10.1037/0096-1523.20.2.421

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crèpin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology, 28*(3), 191–212. doi:10.1016/0028-2243(88)90030-5

Recasens, D. (2012). A cross-language acoustic study of initial and final allophones of /l/. *Speech Communication, 54*(3), 368–383. doi:10.1016/j.specom.2011.10.001

Repp, B. H. (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). Orland, FL: Academic Press. doi:10.1016/b978-0-12-608610-2.50012-1

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science, 8*(2), 162–172. doi:10.1111/j.1467-7687.2005.00403.x

Sato, H., Hirabayashi, Y., Tsubokura, H., Kanai, M., Ashida, T., Konishi, I., … Maki, A. (2012). Cerebral hemodynamics in newborn infants exposed to speech sounds: A whole-head optical topography study. *Human Brain Mapping, 33*(9), 2092–2103. doi:10.1002/hbm.21350

Sato, Y., Sogabe, Y., & Mazuka, R. (2010a). Development of Hemispheric Specialization for Lexical Pitch–Accent in Japanese Infants. *Journal of Cognitive Neuroscience, 22*(11), 2503–2513. doi:10.1162/jocn.2009.21377

Sato, Y., Sogabe, Y., & Mazuka, R. (2010b). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology, 46*(1), 106–119. doi:10.1037/a0016718

Schwartz, J.-L., & Escudier, P. (1989). A strong evidence for the existence of a large-scale integrated spectral representation in vowel perception. *Speech Communication, 8*(3), 235–259. doi:10.1016/0167-6393(89)90004-6

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: is the distributional account all there is to it? *Developmental Science, 12*(6), 874–887. doi:10.1111/j.1467-7687.2009.00829.x

Seebach, B. S., Intrator, N., Lieberman, P., & Cooper, L. N. (1994). A model of prenatal acquisition of speech parameters. *Proceedings of the National Academy of Sciences, 91*(16), 7473–7476. doi:10.1073/pnas.91.16.7473

Shahidullah, S., & Hepper, P. G. (1994). Frequency discrimination by the fetus. *Early Human Development, 36*(1), 13–26. doi:10.1016/0378-3782(94)90029-9

Smits, R., Sereno, J., & Jongman, A. (2006). Categorization of sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 733-754. doi: 10.1037/0096-1523.32.3.733

Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature, 259*(5538), 39–41. doi:10.1038/259039a0

Sundara, M., Ngon, C., Skoruppa, K., Feldman, N. H., Onario, G. M., Morgan, J. L., & Peperkamp, S. (2018). Young infants' discrimination of subtle phonetic contrasts. *Cognition, 178*, 57–66. doi:10.1016/j.cognition.2018.05.009

Swoboda, P. J., Morse, P. A., & Leavitt, L. A. (1976). Continuous Vowel Discrimination in Normal and At Risk Infants. *Child Development, 47*(2), 459. doi:10.2307/1128802

Thiede, A., Virtala, P., Ala-Kurikka, I., Partanen, E., Huotilainen, M., Mikkola, K., … Kujala, T. (2019). An extensive pattern of atypical neural speech-sound discrimination in newborns at risk of dyslexia. *Clinical Neurophysiology, 130*(5), 634–646. doi:10.1016/j.clinph.2019.01.019

Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology*, *9*(1), 91–96. doi: 10.1037/h0034999

Trehub, S. E. (1976). The Discrimination of Foreign Speech Contrasts by Infants and Adults. *Child Development*, *47*(2), 466–472. doi: 10.2307/1128803

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, *56*(2), 179–191. doi: 10.1002/dev.21179

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., … Best, C. T. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: Language-specific developmental changes in speech perception abilities. *The 3rd International Conference on Spoken Language Processing, ICSLP 1994, Yokohama, Japan, September 18-22, 1994*. Retrieved from http://www.isca-speech.org/archive/icslp_1994/i94_1695.html

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences, 104*(33), 13273–13278. doi:10.1073/pnas.0705369104

Wanrooij, K., Boersma, P., & Van Zuijen, T. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. *Frontiers in Psychology*, *5*, 77. doi:10.3389/fpsyg.2014.00077

Weikum, W. M., Oberlander, T. F., Hensch, T. K., & Werker, J. F. (2012). Prenatal exposure to antidepressants and depressed maternal mood alter trajectory of infant speech perception. *Proceedings of the National Academy of Sciences, 109*(Supplement 2), 17221–17227. doi:10.1073/pnas.1121263109

Werker, J. F. (1995). Exploring developmental changes in cross-language speech perception. In L. R. Gleitman & M. Liberman (Eds.), *An invitation to cognitive science. Language: An invitation to cognitive science* (pp. 87–106). Cambridge, MA, US: The MIT Press.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development, 1*(2), 197–234. doi:10.1080/15475441.2005.9684216

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental Aspects of Cross-Language Speech Perception. *Child Development, 52*(1), 349-355. doi:10.2307/1129249

Werker, J. F., & Lalonde, C. (1988). Cross-Language Speech Perception: Initial Capabilities and Developmental Change. *Developmental Psychology*, *24*(5), 672–683. doi: 10.1037/0012-1649.24.5.672

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*(1), 49–63. doi: 10.1016/S0163-6383(84)80022-3

Werker, J. F., & Tees, R. C. (1992). The Organization and Reorganization of Human Speech Perception. *Annual Review of Neuroscience*, *15*(1), 377–402. doi: 10.1146/annurev.ne.15.030192.002113

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect 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

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-Month-olds use distinct objects as cues to categorize speech information. *Cognition, 113*(2), 234–243. doi:10.1016/j.cognition.2009.08.010