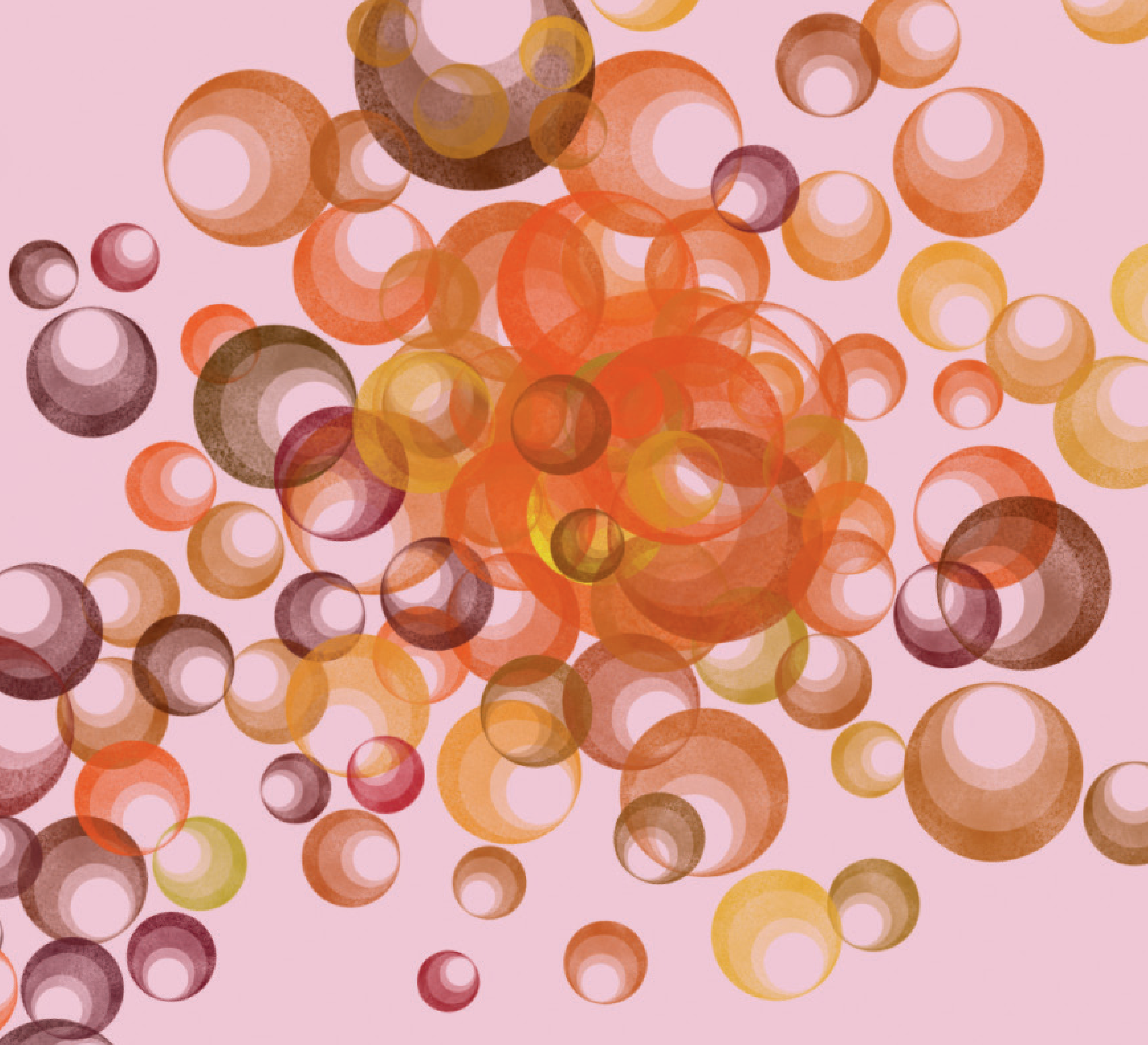


# Vowels in development

## Speech sound perception in Dutch infants with and without a family risk of dyslexia

**Maartje de Klerk**

# Vowels in development
## Speech sound perception in Dutch infants with and without a family risk of dyslexia

**Klinkers in ontwikkeling**
De perceptie van spraakklanken in Nederlandstalige baby's
met en zonder een familiair risico op dyslexie
(met een samenvatting in het Nederlands)

# Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op vrijdag 3 september 2021
des middags te 12.15 uur

door

**Maria Kitty Anna de Klerk**

geboren op 13 mei 1976 te De Bilt

**Promotoren**
Prof. dr. F. N. K. Wijnen
Prof. dr. E. H. de Bree

*The research reported in this dissertation was funded by The Netherlands Organization for Scientific Research, grant NWO-360-70-270.*

# Vowels in development

## Speech sound perception in Dutch infants with and without a family risk of dyslexia

**Maartje de Klerk**

**Leden van de beoordelings- en promotiecommissie**
Prof. dr. W.B.T. Blom
Prof. dr. E. Gerrits
Dr. C.M.M. Junge
Prof. dr. R.W.J. Kager
Prof. dr. C.C. Levelt
Prof. dr. J.E. Rispens



Cover illustration by Maartje de Klerk
Layout: Maartje de Klerk & Liesbeth van Woerkens

ISBN: 978-94-6103-089-4

Printed by: Ipskamp drukkers, Enschede

doe het toch maar
zeg dat maar tegen jezelf
op die momenten dat je niet meer weet
waar je het voor doet

doe het gewoon
want ergens weet je
dat dit het enige is waardoor je
in vrede met jezelf en de wereld kan leven

Uit *doe het toch maar* van Babs Gons

–

Voor Jonne, Louis en Benjamin

# Contents

Chapter 1

# General introduction

## 1.1 Speech sound categories in language acquisition

Speech sounds are the smallest elements of human language. The acquisition of speech sounds starts as soon as infants first see the light of life. In the case of vowels, acquisition possibly starts in the womb, when the fetus' hearing develops rapidly in the third trimester of pregnancy (Moon et al., 2013). Speech sounds are like snow crystals. Snow crystals all look the same until they are examined closely, it turns out they all are unique. The same is true for speech sounds: If you were to determine the acoustic properties of the 'e'-sound in the English word 'bed' as pronounced by 20 people, you would find them to be all different. Nonetheless, a native speaker of English will effortlessly identify each of these sounds as 'e'. The same would hold for the 'b' and 'd' sounds, and for any other speech sound. Many factors influence the acoustic properties of a speech sound. For example, the 'e' in 'bed' is acoustically different from the 'e' in 'led', as neighboring speech sounds have an influence on its articulation. Listeners treat these variants of 'e' as the same elements. Such classes of speech sounds that are perceived as equivalent are called *phonemes*. Hence, a phoneme is the mental (abstract) representation of a speech sound category.

The acquisition of speech sound categories is important for language development and use (e.g., Kuhl, 2004). An example that illustrates this is that speech sound categories are needed to distinguish between words, such as minimal word pairs. Minimal word pairs are words that differ in one speech sound category, or phoneme, such as *bed* and *bad*, or *pear* and *bear*. Phonemes, such as English /ɛ/ and /æ/, differentiate meaning between words /bɛd/ (*bed*) and /bæd/ (*bad*). Hence, infants need to be able to discriminate between speech sounds of their native language in order to learn to differentiate between spoken words and their meaning. This sounds very obvious and not problematic at all.

However, the challenge that infants face is that they have to respond to differences between *bed* and *bad*, but not to differences between different pronunciations of, for instance, bed. Factors influencing acoustic variations of speech sounds within the same speaker are, other than the neighboring speech sounds, the intensity with which is spoken, intonation, speaker affect (e.g., the difference between a happy or angry voice) or the physical condition, for example, whether the speaker smokes or not (e.g., Suwandi et al., 2020). Next to differences within the same speaker, variations between the speech sounds are also caused by different speakers. This inter-speaker variation stems, for example, from variations in speaking rate, a speaker's accent/dialect, age and vocal tract dimensions (e.g., Fisher & Linville., 1985).

All the ensuing acoustic variations across realizations of a specific speech sound category that do not alter word meaning are called allophonic differences. The listener needs to cluster these allophones to one single phonemic category, because the differences between these allophones are irrelevant. A phonemic category can therefore be defined as a group or cluster of phonetic realizations, allophones, of a particular language (Kuhl, 2004). Whereas allophonic differences should be ignored in word identification, other acoustically small differences that mark phonemic differences should not. In other words, a particular acoustic difference may reflect two allophonic tokens of /ɛ/, while another difference, possibly of a comparable acoustic magnitude, may mark the distinction between English /ɛ/ and /æ/. The challenge infants face is thus to cluster allophones into speech sound categories.

Acquisition of speech sounds entails both perception and production. In order to be able to utter specific speech sounds, a speaker must be able to perceive those speech sounds. A very well-known example is that Japanese adults have difficulty pronouncing the English /l/ and /r/ sounds as in *rice* and *lice*. These difficulties stem from the fact that Japanese does not have this /l/ and /r/ contrast, but only have one category, the Japanese /r/. As a consequence, Japanese native speakers cannot perceive the difference between these speech sounds very well (Miyawaki et al., 1975). Although perception of English /l/ and /r/ is poor in Japanese adults, experimental research has shown that perception can be improved by exposing the participant to speech sound tokens from a speech sound continuum (i.e., speech sound tokens ranging from /r/ to /l/ in acoustically equal steps) produced by different speakers (e.g., Lively et al., 1993). Moreover, a study by Bradlow and colleagues (1999) showed that perception of English /l/ and /r/improves production and vice versa. Hence, if listeners perceive the sounds better, they are also better able to produce them (and vice versa). Together, these examples illustrate the relation between perception and production and emphasizes the importance of well-developed speech sound categories.

The acquisition of speech sound categories is one of the major stepping stones of language acquisition (Kuhl, 2004). The ability to discriminate between speech sounds is needed to build speech sound categories. Studies have shown a significant correlation between early speech sound discrimination and later language skills: The stronger the discrimination performance of native, but not non-native speech, sound contrasts in 7.5-month-old infants, the higher the score on several language tasks (vocabulary size, sentence complexity and mean length of utterances) at 24-30-months of age (Kuhl et al., 2008). It could be inferred that the better native speech sounds are discriminated, the stronger or better the speech sound categories are developed. Hence, these results imply that there is a relation between (well-developed) speech sound categories and later language skills. Another example of the importance of well-developed speech sound categories for language-related development is their role in literacy: the phonemes need to be mapped onto the correct visual symbols, graphemes (Vellutino et al., 2004). In sum, the development of speech sounds is highly important and marks one of the first steps in native language development.

### 1.2 Speech sound categories and dyslexia

While children acquire spoken language spontaneously and effortlessly, they need to invest attention and effort to acquire written language. Instruction and practice are needed to become a fluent reader. For some children the process of learning to read and/or spell is a difficult road, full with struggles and frustration, that doesn't necessarily lead to fluent reading. Children who have persistent difficulties in learning to read and/or spell, despite proper literacy education, normal intelligence and no other neuro-developmental disorders can be classified as having developmental dyslexia (henceforth dyslexia, Lyon et al., 2003; Peterson & Pennington, 2015).

Dyslexia is considered to be a multifactorial disorder (Pennington, 2006; Peterson & Pennington, 2015), which implies that multiple risk- and protective factors are involved. The severe and persistent reading and/or spelling problems of people with dyslexia have been proposed to be caused by a phonological deficit (Ramus et al., 2003; Vellutino, et al., 2004). This generally accepted phonological deficit hypothesis states that individuals with dyslexia have less well-developed phoneme representations (e.g., Vellutino et al., 2004), which complicates the construction of the phoneme-grapheme associations required for reading. Evidence that supports the phonological deficit comes from studies that showed that individuals with dyslexia typically have poor phonological awareness skills, referring to the ability to manipulate speech sounds, for example to remove the /r/ from tree (Bus & IJzendoorn, 1999; van Viersen et al., 2018). They also perform more poorly than their peers on 1) verbal short-term memory, measured with nonword repetition (e.g., de Bree et al., 2010; Moll, et al., 2013), 2) letter-sound association (Blau et al., 2009; Mittag et al., 2013) and 3) rapid automatized naming, in which pictures, colors, letters or numbers have to be named as quickly as possible (Araújo et al., 2014; Donker et al., 2016).

One line of research has focused on the question of whether the phonological deficit stems from a speech perception deficit (e.g., Boets et al., 2007; Hakvoort et al., 2016; Richardson et al., 2003; Serniclaes et al., 2004; Werker & Tees, 1987). If speech sounds are not perceived adequately, this might impede the formation of phoneme representations, which, in turn, might have a negative effect on learning to associate phonemes to graphemes (e.g., Blomert, 2011, Fraga Gonzalez et al., 2015). Speech perception can be measured through speech sound categorization and discrimination, with categorization referring to the ability to cluster phonetically distinct speech sounds to the same category (Kuhl, 2004) and discrimination referring to the ability to differentiate between two speech sounds. A general finding is that adults and children with dyslexia perform more poorly on categorization and discrimination tasks (Hakvoort et al., 2016; Maassen et al., 2001; Mody et al., 1997; but see Messaoud-Galusi et al., 2011 for contrasting findings). On the other hand, there are also indications that individuals with dyslexia have a heightened sensitivity to allophonic differences. The theory of an allophonic mode of speech perception entails that individuals with dyslexia maintain sensitive to phonemic distinctions irrelevant of their native language (Noordenbos et al., 2012; Serniclaes et al., 2004). Whether or not individuals have poorer or heightened discrimination skills, a recent meta-analysis does show support for a categorical perception deficit in dyslexia (Noordenbos & Serniclaes, 2015).

As dyslexia runs in families (Snowling & Melby-Lervåg, 2016), speech perception in children of dyslexic parents (family risk; FR) can be evaluated. Studies that investigated whether categorization and discrimination difficulties also occur prior to the acquisition of literacy skills, found that FR children and infants also perform more poorly on speech sound discrimination and categorization tasks (e.g., Leppänen et al., 1999; Noordenbos & Serniclaes, 2015; Richardson et al., 2003; Thiede, et al., 2019; van Leeuwen et al., 2006). These results could imply that individuals with (an FR of) dyslexia have less well-developed speech sound categories, or phoneme representations. Yet, the developmental trajectory of the acquisition of speech sound categories in FR infants has never been investigated throughout the first year of life, the timespan in which it is assumed that infants tune into their native language.

### 1.3 Speech sound categories and perceptual attunement

During the first year of life infants develop from ‘universal listeners’ to ‘language-specific listeners’ (Best, 1994; Kuhl et al., 2008; Maye et al., 2008; Werker & Curtin, 2005). The assumption is that infants are born with the ability to distinguish all the possible speech sound contrasts of the world’s languages, including those that are not relevant in their native language phonology, and that this ability declines as they mature. Earlier in this introduction I gave the example of Japanese adults, who have difficulty distinguishing the English /l/ from the English /r/ (Miyawaki et al., 1975) because the /l/ - /r/ contrast is not part of the phonology of Japanese. However, Japanese infants can initially discriminate this contrast. Infants between 6-8-months of age discriminate /r/ from /l/. However, as they get older, they gradually lose sensitivity to this contrast: already at 10-

12- months of age they are less sensitive to the same contrast (Kuhl, et al., 2006; Tsushima et al., 1994). In contrast, their American peers show increased sensitivity to this contrast (Kuhl et al., 2006). The phonology of the native language determines whether sensitivity to speech sound contrast maintains or decreases.
Hence, the ability to discriminate speech sound contrasts that are irrelevant in the native language phonology declines during the first year of life, but perception of native contrasts improves. This developmental change is often referred to as *perceptual attunement* (Maurer & Werker, 2014). Werker & Tees (1984) were the first to show cross-linguistic evidence for this developmental shift, and many others followed (e.g., Best, 1994; Cheour et al., 1998; Kuhl et al., 1992; Polka & Werker, 1994). These studies showed that infants can discriminate non-native consonantal contrasts before the age of 8 months, but around 10-12 months the sensitivity to non-native contrasts declines, whereas native contrast are discriminated throughout development (e.g., Cheour et al., 1998; Werker & Tees, 1984) or are not discriminated from birth but discrimination improves (Liu & Kager, 2016; Narayan et al., 2010; Polka et al., 2001; Sato et al., 2010). For vowels this shift from universal to language specific has been reported to take place earlier, around the ages of 6-8 months (Mazuka et al., 2014; Polka & Bohn, 1996; Polka & Werker, 1994; see also, Tsuji & Cristia, 2014).
Most of what is known about the development of speech perception comes from research conducted with English learning infants and consonantal contrasts. From a cross-linguistic perspective it is important that many types of contrasts (consonants, vowels, tones) and language learners (e.g., Dutch, Chinese, Kenyan) are the topic of investigation, in order to be able to draw conclusions about (proposed) universal maturational patterns, and to establish if the developmental trajectory is affected by language specific factors. Several studies have found that perceptual attunement is less straightforward than it was originally claimed to be. For instance, not all speech sounds are discriminated from birth (e.g., Liu & Kager, 2016; Narayan et al., 2010). Also, for some non-native consonantal and vowel contrasts a decline in sensitivity is not found, counter to expectations (Best & Faber, 2000; Mazuka et al., 2014; Polka & Bohn, 1996; Tyler, Best, Goldstein & Antoniou, 2014). Assessing native and non-native vowel discrimination in Dutch infants would contribute to cross-linguistic insights of the development of speech sound perception.

## 1.4 Speech sound categories and (individual) discrimination assessment

Although it is important to study the developmental trajectory of speech sound discrimination, obtaining reliable data is a challenge. It is not possible to ask infants whether they perceive the difference between two speech sounds. Instead, studies assessing infants' knowledge of language use indirect measures of perceptual discrimination. A frequently used behavioral method is the habituation paradigm (Oakes, 2010), in which looking time to a visual stimulus that is paired to an auditory stimulus is the outcome measure. Thus, 'looking time' is taken to reflect listening time (Aslin, 2007) and this term (looking time) will be used throughout the dissertation. In experiments using such a design, the infant sits on the caregiver's lap looking at a screen on which visual stimuli are displayed. Auditory stimuli are played simultaneously. Generally, infants are habituated on a set of stimuli (e.g., the nonce word *sen*), followed by a test phase in which infants are tested on new set of stimuli (e.g., *saen*). If infants are sensitive to the difference between the habituation stimulus (e.g., *sen*), longer looking times are expected to the novel stimuli (e.g., *saen*, Sokolov, 1963).

Habituation studies often employ designs with only 2-4 test trials (see Colombo & Mitchell, 2009 for a review). However, looking time data are noisy: They are influenced by external (e.g., sounds outside the lab) and internal (e.g., bowel movements) distractions and cognitive abilities (e.g., attention span). Hence, in order to reliably assess discrimination performance, designs require data of many infants (e.g., Houston et al., 2007). This leads to logistical problems (parent and infant recruitment, duration of data collection). Furthermore, and more importantly, such a design only allows for finding general (group level) patterns of discrimination and not for interpretation at the individual level.

To overcome these methodological issues, Houston et al. (2007) designed the Hybrid Visual Fixation paradigm (HVF). This paradigm presents 14 test trials per infant. Adding more test trials to the test phase increases the likelihood of actually *learning* to discriminate between stimuli. However, in the HVF design old (familiarized) and new (unfamiliarized) stimuli alternate *within* a trial (alternating trials) and are far less frequent (4 trials) than the remaining 10 non-alternating trials in which only familiarized stimuli are presented. The combination of alternating/non-alternating trials and imbalance in frequency of occurrence prevents infants from learning the new stimuli during the test phase. Studies using this HVF have showed high test-retest results (Houston et al., 2007; Cristia et al., 2016). This is reassuring, as replicability of research findings is not self-evident in infant studies (Frank et al., 2017). Furthermore, on the basis of this design, Houston et al. (2007) showed that the design and subsequent statistical analyses allow individual speech sound discrimination performance.

Assessing individual discrimination performance is of great value for studies into language development as well as studies into language(-related) problems and disorders. Currently studies (prospectively) predict other measures of language or literacy skills at *group level*, whereas in the case of dyslexia, for instance, it would be of interest to be able to use infants' individual discrimination outcomes to prospectively predict later language outcomes. Such an approach would strengthen and deepen our knowledge about the different risk factors that contribute to the development of the phonological difficulties, letter-sound associations and subsequent literacy difficulties.

### 1.5 Speech sound categories and distributional learning

One main question in speech perception is thus whether perceptual attunement is attested in vowels and in Dutch infants with and without an FR. Answering this question will provide insight into the construction of speech sound categories. A related question is what learning mechanism contributes to the shift from universal to language specific perception.

One mechanism that has been proposed to influence language learning is distributional learning. A large body of research has shown that adults, children and infants are sensitive to distributional (statistical) properties of their language: they are capable of tracking regularities in language input, such as the number of phoneme (co-)occurrences and sequences, to detect word-boundaries (e.g., Saffran et al., 1996), syntactic categories (e.g., Gerken et al., 2005), and speech sound categories (e.g., Capel et al, 2011; Liu & Kager, 2014, 2017; Maye et al., 2008; Wanrooij et al., 2014; Yoshida et al., 2010).

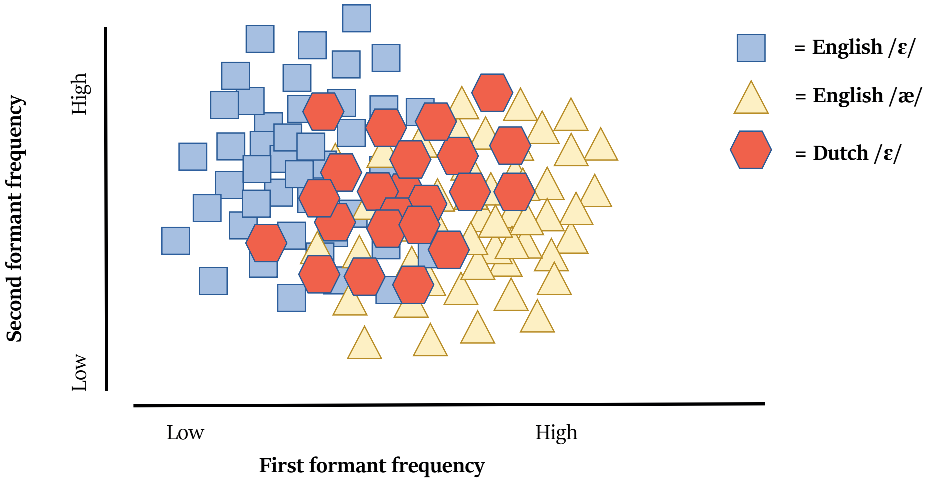
The distribution of speech sounds that listeners are exposed to reflect the phonemic categories used in the listeners' native language(s): Allophones cluster around acoustic values that are most frequently produced, and hence, together they reflect the phonemic categories used in a language (Kuhl, 2004). For example, English has two categories, the English /ɛ/ and /æ/, each with their own range of allophones and prototypical acoustical values. Dutch, on the other hand, has one category, the Dutch /ɛ/. Speech realizations of /æ/ are rather rare in Dutch compared to productions of /ɛ/ and, hence, are clustered around the frequently occurring, prototypical, /ɛ/ sound. If we visualize these productions of the English /æ/ and /ɛ/ and the Dutch /ɛ/ based on the first and second formant frequency values [1], we get a picture like Figure 1. It must be noted that Figure 1 is a simplification of speech in the real world: speech sounds are multi-dimensional, as they vary on multiple acoustic features due to inter- and intra-speaker variation. The assumption of distributional learning is that infants who acquire English as their native language are exposed to a bimodal distribution of the /ɛ/ - /æ/ speech

1 Formant frequencies are the peak resonance frequencies, caused by the vibration of the vocal folds and filtered by the vocal tract. From the formant frequencies that can be measured from a sound wave, the first two formant frequencies contribute most to identification of vowels.

sounds, whereas Dutch infants are exposed to a unimodal distribution. By tracking the frequencies of occurrences of the speech sounds, infants learn their native categories. A meta-analysis of Cristia (2018) showed that infants, in a lab setting, are able to learn new categories based on phonetic distributional information. Moreover, it has been proposed that the timepoint at which this mechanism has the strongest effect, is around the same timepoint that infants become language-specific listeners (Reh et al., 2021).

Figure 1

*A fictional illustration of speech sounds spoken by multiple speakers of English*




*Note.* F1 and F2 refer to first and second formant frequencies.

Although evidence has thus been reported for distributional learning of speech sounds (e.g., Cristia, 2018), most of this pertains to consonants, while attention to vowels has been rare. Furthermore, it has not been investigated yet whether this learning mechanism of distributional learning is accessible to FR infants. Yet, this is of interest as adults and children with (an FR of) dyslexia perform poorly on categorization tasks (Steffens et al., 1992; Werker & Tees, 1987). This might mean that development of speech sound categories is delayed or compromised in dyslexia, which might be due to different sensitivity to speech sounds and differences in distributional learning. We know of one study that looked into distributional learning in children with dyslexia. Vandermosten et al., (2019) found that that distributional learning was not as effective for grade 3 children with dyslexia as it was for their non-dyslexic peers. This is in line with studies that measure distributional learning in the context of phonotactic learning in 8-9-year-old children with dyslexia (Bonte et al., 2007) and artificial grammar learning in children and adults with dyslexia (e.g., Gabay et al., 2015; Pavlidou et al., 2010) and FR toddlers (Kerkhoff et al., 2013). Difficulties in distributional (statistical) learning might thus affect the language and literacy development of with people (an FR of) dyslexia.

**1.6 This dissertation**

In this dissertation, four (experimental) studies are devoted to gaining more knowledge about early speech sound acquisition. The research question of the first study (Chapter 2) is whether Dutch infants' discrimination performance follows the developmental trajectory as predicted by the theory of perceptual attunement. We tracked the development of two vowel contrasts in 6-8-10 month-old typically developing infants. We assessed perception of a salient, acoustically and articulatory highly distinctive, native contrast, Dutch /aː/ - /eː/, and a non-salient, non-native contrast, English /ɛ/ - /æ/. As the native contrast is salient, we expect that even very young children will be able to discriminate this contrast. Thus, the native contrast serves as a control condition. The English /ɛ/ - /æ/ contrast, on the other hand, is difficult to perceive for Dutch adults (Broersma & Cutler, 2011) as Dutch has one category, the Dutch /ɛ/. According to the theory of perceptual attunement, Dutch infants will initially show sensitivity to the English /ɛ/ and /æ/contrast, but this sensitivity will decline with age as Dutch infants will learn to perceive English /æ/ as an allophone of the Dutch /ɛ/. Previous studies have not always reported a decline in sensitivity for non-native consonantal contrasts (e.g., Tyler, Best, Goldstein & Antoniou, 2014) and vowel contrasts (e.g., Best & Faber, 2000) counter to predictions. Therefore, although a continued or increased sensitivity to the native vowel contrast combined with a decrease in sensitivity to the non-native contrast is predicted, it is an open question whether such effects will be found.

The second question is whether it can reliably be assessed which infants can be classified as discriminators, and which infants cannot. The data of the same infants as reported on in Chapter 2 were used to assess individual discrimination performance of the native Dutch /aː/ - /eː/ contrast. The statistical approach taken by Houston et al (2007) was replicated and extended with a more advanced approach, namely Bayesian hierarchical modeling (Chapter 3).

The third question (Chapter 4) is whether the discrimination performance of infants with a family risk of dyslexia follows predictions from the perceptual attunement theory. Here too, perception of the native /aː/ - /eː/ and English /ɛ/ - /æ/ was assessed in 6-8-10 month-old infants, specifically FR and no-FR infants.

We anticipate that the FR infants will discriminate the native contrast, as this is a salient contrast. However, on the basis of the phonological deficit hypothesis (Vellutino et al., 2004) and findings of speech perception difficulties in infants and children with (an FR of) dyslexia (e.g., Leppänen et al., 1999; Richardson et al., 2003; Thiede, et al., 2019; van Leeuwen et al., 2006), it is to be expected that FR infants will not show evidence of discriminating the non-native contrast. Another possibility is that infants will not lose the sensitivity to the irrelevant non-native contrast (Noordenbos, et al., 2012).

The final question is whether infants with and without an FR of dyslexia could potentially use phonetic distributional information in the input to acquire speech sound categories. On the basis of the meta-analysis by Cristia (2018) we expect that no-FR infants will discriminate the contrast in the bimodal, but not the unimodal condition. In contrast, on the basis of both the phonological deficit of dyslexia and findings of poorer distributional learning in dyslexia (Banai & Ahissar, 2018; Bonte et al., 2007; Kerkhoff et al., 2013; Vandermosten et al., 2019; van Witteloostuijn et al. 2017; Wijnen, 2013), we expect that distributional learning of speech sounds will be poorer in FR infants (Chapter 5). These four studies are followed by a general discussion, in which the findings of the four studies are interpreted, integrated and evaluated (Chapter 6).

# Chapter 2

# Lost and found: decline and reemergence of non-native vowel discrimination in the first year of life

**Abstract**
Our aim was to investigate perceptual attunement (PA) in vowel perception of Dutch-learning infants (6-8-10-month-olds) using the hybrid visual fixation paradigm (Houston et al., 2007). Infants were habituated to one phoneme and subsequently tested on items in which a token of the habituated phoneme alternated with either another token of the same phoneme, or a token from another phonemic category. Habituation involved tokens of multiple speakers. Infants were tested on a native (/aː/ - /eː/) and non-native (/ɛ/ - /æ/) contrast. The 6-month-olds (*n* = 38), 8-month-olds (*n* = 44) and 10-month-olds (*n* = 35) discriminated the native contrast. The non-native contrast was discriminated by the group of 6-month-olds (*n* = 42) but not the 8-month-olds (*n* = 47), in line with PA. However, the 10-month-olds (*n* = 39) also showed discrimination. We conclude that discrimination of phonetic categories can occur after perceptual attunement; discrimination performance is sensitive to tasks applied.

Published as de Klerk, M., de Bree, E., Kerkhoff, A., & Wijnen, F. (2019). Lost and found: decline and reemergence of non-native vowel discrimination in the first year of life. *Language Learning and Development*, 15(1), 14-31. htpps://doi.org/10.1080/15475441.2018.1497490

EdB, AK, MdK and FW designed the study. MdK tested the infants and conducted the analyses. MdK wrote and revised the paper, with input and feedback from EdB, AK and FW.

## 2.1 Introduction

In acquiring the sound system of their native language, infants learn which acoustic variations indicate phonemic contrasts and which are phonologically irrelevant, such as those resulting from inter- and intra-speaker variation. As a corollary of this learning process, infants' speech perception changes from language-general to language-specific in the first year of life: sensitivity to native speech sound contrasts increases whereas sensitivity to (most) non-native speech sounds decreases (e.g., Cheour et al., 1998; Werker & Tees, 1984). This process is often referred to as perceptual attunement (PA, see Maurer & Werker, 2014 for a recent review). A central prediction of PA is that sensitivity to non-native speech sound contrasts that are assimilated to one native category by adults declines in the first year of life. Although many studies report data that are in agreement with this prediction, not all do (e.g., Best & Faber, 2000; Mazuka et al., 2014; Polka & Bohn, 1996; Tyler, Best, Goldstein, & Antoniou, 2014). Given the lack of uniformity in the literature, further investigation of speech sound discrimination in infancy is warranted. Here, we assess the developmental trajectory of the discrimination of a salient native contrast (serving as a control experiment) and a non-salient non-native contrast in Dutch infants aged six, eight and ten months.

Werker and Tees (1984) were the first to report evidence for PA. They found that English infants discriminated Hindi dental-retroflex plosive (/t̪a/ - /ʈa/) and Nthlakampx velar-uvular ejective /k'i/ - /q'i/) contrasts at 6-8 months and 8-10 months, but were not able to do so at 10-12 months of age. In contrast, 11-12-month-old Hindi and Salish learning infants discriminated their native consonant contrasts. Subsequent studies supported PA in consonant perception (e.g., Best et al., 1995; Sundara et al., 2008; Werker & Lalonde, 1988). Although the number of studies on vowel discrimination is limited, findings also show PA, but at an earlier age than for consonants, i.e. around 6-8 months of age (Bosch & Sebastián-Gallés, 2003; Kuhl et al., 1992; Polka & Werker, 1994; Tsuji & Cristia, 2014).

PA predicts monotonic developmental trajectories. For native speech sounds, early discrimination of highly salient contrasts is anticipated. Early sensitivity is not always found for the less salient native contrasts (Liu & Kager, 2016; Narayan etal., 2010; Polka et al., 2001; Sato et al., 2010). When early sensitivity is not attested, gradual acquisition is expected as a result of continued exposure to the native language. This is referred to as enhancement of discrimination or facilitation (Kuhl et al., 2008, Narayan et al., 2010; Tyler, Best, Goldstein & Antoniou, 2014). For non-native contrasts, PA predicts a decline in discrimination if the speech sounds of the non-native contrast can be assimilated to one native speech sound. Several studies support this prediction (e.g., Best, 1994; Best &

McRoberts, 2003; Cheour et al., 1998; Polka & Werker, 1994; Werker & Lalonde, 1988; Werker & Tees, 1984).

However, a decline in discrimination of non-native speech sounds is not always found (e.g., Best & Faber, 2000; Polka & Bohn, 1996; Mazuka et al., 2014; Tyler, Best, Goldstein and Antoniou, 2014). Tyler, and colleagues (2014), for instance, found that English-learning infants at 6 and 11 months discriminated two non-native fricative velar-uvular /χ/ - /x/ and uvular-pharyngeal /χ/ - /ħ/) contrasts from Nuu-Chah-Nulth, a language spoken on Vancouver Island, Canada. PA predicts that the older group would not be able to do so. A similar pattern has been reported for vowel perception. Polka and Bohn (1996) assessed discrimination of a German /u/ - /y/ and an English /ɛ/ - /æ/ contrast by English-and German-learning infants. They found consistent discrimination by both the German and English-learning infants at 6-8 as well as 10-12 months. Hence, a decline is not always attested.

One factor that might influence discrimination performance is the phoneme used as the standard (habituation) stimulus. Polka and Bohn (1996) found that both English- and German-learning infants showed better discrimination when the habituation stimulus was /y/ than when it was /u/. Similarly, discrimination of the English /ɛ/ - /æ/ was attested when the habituation stimulus was /ɛ/, not when it was /æ/. Polka and Bohn explain these findings with their Natural Referent Vowel framework (NRV, Polka & Bohn, 2011). They propose that vowels in the most peripheral positions of the vowel space (based on their first two formant frequencies, i.e. /i/, /a/ and /u/), function as points of reference in the acquisition of the native vowel system. Due to their distinct acoustic and articulatory features, these vowels attract infants' attention (more than non-peripheral vowels do). Consequently, when infants have been habituated to a less peripheral vowel (e.g., German /y/), and are subsequently presented with a peripheral vowel (German /u/), they would show a stronger discrimination response than in the reverse situation. Moreover, the NRV framework proposes that, in general, discrimination is better if the order of presentation is from a less peripheral vowel to a more peripheral vowel, than the reverse. For example, English /æ/ is more peripheral than the English /ɛ/, and this would explain why discrimination is better when participants are habituated on English /ɛ/ (and subsequently hear English /æ/), than when the habituation stimulus is English /æ/. Although NRV seems to give a plausible explanation for the asymmetries found in speech perception, not all studies assessing discrimination of non-native vowel contrast find discrimination asymmetries that align with the NRV framework (e.g., Best & Faber, 2000; Mazuka et al., 2014; Tyler, Best, Faber & Levitt, 2014).

**Current Study**

The literature shows that a decline in non-native discrimination over age does not always occur. The present study aims to provide more cross-linguistic data on non-native speech perception. We tracked the development of two types of contrasts: a salient (acoustically and articulatory highly distinctive) native contrast, i.e. the Dutch /aː/ - /eː/ and a non-salient, non-native contrast, i.e. the English /ɛ/ - /æ/ in infants aged six, eight and ten months old. As the native contrast is salient, we expect that even very young children will be able to discriminate this contrast. Thus, the native contrast serves as a control condition, to assess whether the hybrid visual habituation paradigm (HVF, Houston et al., 2007) was suitable for assessing speech sound discrimination skills. Selection of a salient native contrast was preferred over a less salient native vowel contrast, such as Dutch /ɪ/ - /iː/: Younger infants might not show evidence of discrimination, as less salient contrasts take longer to acquire (Liu & Kager, 2016). Consequently, using a less salient contrast would not be appropriate for determining the sensitivity of the HVF procedure. To establish whether the sensitivity to a non-native and non-salient (acoustically and articulatory less distinctive) speech sound contrast declines, we chose the English /ɛ/ - /æ/ contrast. We expected that this would be a difficult contrast for the older infants, as (native) Dutch adult listeners assimilate both the English /ɛ/ and /æ/ to the Dutch /ɛ/ (Broersma and Cutler, 2011; Schouten, 1975).

We used the HVF procedure, which comprises more test trials (fourteen) than traditionally used in speech discrimination research and showed good test-retest reliability (Houston et al., 2007). It is a habituation-dishabituation procedure that combines elements of two other variants of visual fixation procedures. The first is the oddity variant, in which during test the old habituated stimulus is presented less frequently than the new stimulus. The second is the Stimulus Alternation Preference Procedure (SAPP, Best & Jones, 1998), which comprises non-alternating and alternating trials in the test phase. In our study, the procedure starts with habituation to one of the phonemes (e.g., /æ/ or /ɛ/), and this is followed by a test phase with eight non-alternating (e.g., /æ-æ/ or /ɛ-ɛ/) and four alternating pairs (e.g., /ɛ-æ/or /æ-ɛ/)[1].

We used tokens from four different female speakers during habituation. Speaker variability has been argued to enhance generalization of abstract features in the process of developing phonetic categories (Lively et al., 1993; Potter & Saffran, 2017; Rost & McMurray, 2009). In day to day speech perception infants need to extract acoustic information that is relevant to phonemic contrasts, while redundant information, not contributing to meaningful differences, needs to be ignored. Hence, the use of multiple speakers makes the task more comparable to the

demands of natural speech. Moreover, previous studies assessing discrimination of native and non-native vowel contrasts with multiple speakers have shown that infants are able to extract the relevant acoustic features to distinguish the contrast (Bosch & Sebastian-Gallés, 2003; Sebastian-Gallés & Bosch, 2009).
Two questions were addressed. The first was whether infants discriminate the native contrast at all ages, which we expected to be the case. The second was whether infants show a decline in discrimination performance of the non-native contrast. Here, expectations were less clear-cut. Based on the results of previous studies, both a decline in discrimination and its absence are conceivable (e.g., Polka & Werker, 1994; Polka & Bohn, 1996; Tyler, Best, Goldstein & Antoniou, 2014).

## 2.2 Method

*Participants*

Infants were recruited via the municipality of Utrecht (the Netherlands), and were divided into three age groups: 6- 8- and 10-month-olds. Caregivers were asked to fill out a questionnaire, which asked about birth weight, gestational age, health issues, and family background. Infants were included if: (a) they were raised only in Dutch; (b) their gestational age at birth was considered average, i.e., between 37 and 43 weeks; (c) their birth weight was considered average, i.e., between 2500-5000 grams; (d) there were no complications during the pregnancy or delivery; (e) did not have a history of known hearing loss or reduced vision and (f) they did not have reported neurological problems.
The aim was to include a minimum of 30 participants who finished both experiments in each age group, divided across habituation stimulus and contrast order. Given the number of anticipated drop-outs this number differs slightly for each age group and contrast (see Table 1, column 'Data included'). In total, 354 infants participated; see Table 1 for an overview of the age ranges and drop-out rates per contrast. One hundred and nine infants (31%) were tested but their data was not included in the data analysis. There were different reasons for this: behavior (crying, extreme restlessness, $n = 58$); failure to meet the habituation criterion ($n = 23$; see Procedure); technical errors ($n = 22$); having an ear infection at the time of testing ($n = 3$); parental interference ($n = 2$), or failure to meet the pre- and posttest attention criterion ($n = 1$; see *Procedure*). Although this drop-out rate is substantial, it can be considered normal for habituation studies (e.g., Narayan et al., 2010; Tyler, Best, Goldstein & Antoniou, 2014), especially in a design in which two contrasts were presented subsequently.

Table 1

*Numbers of participants, mean ages and age ranges, and drop-out rates per age group, for each of the two contrasts (native – non-native)*

| Contrast | Age group | Age range | Age (days) | Data tested | Drop out 1 | Drop out 2 | Data included | Habituation stimulus 1 |
|---|---|---|---|---|---|---|---|---|
| | | *month.days* | *M (SD)* | *N* | *n* | *n* | *n (female)* | *n* |
| Native | 6 | 6.1 - 6.30 | 203 (8.4) | 59 | 9 | 12 | 38 (18) | 17 |
| (/a:/ - /e:/) | 8 | 8.0 - 8.30 | 259 (6.5) | 66 | 8 | 14 | 44 (29) | 25 |
| | 10 | 10.3-10.30 | 320 (12.9) | 45 | 3 | 7 | 35 (18) | 28 |
| | Subtotal | | | 170 | 20 | 33 | 117 (66) | 70 |
| Non-native | 6 | 6.1 - 6.29 | 202 (8.2) | 65 | 9 | 14 | 42 (12) | 25 |
| (/æ/ - /ɛ/) | 8 | 8.3 - 8.29 | 261 (8.3) | 70 | 9 | 14 | 47 (26) | 27 |
| | 10 | 10.3 - 10.30 | 325 (6.8) | 49 | 3 | 7 | 39 (18) | 18 |
| | Subtotal | | | 184 | 21 | 35 | 128 (56) | 70 |
| | Total | | | 354 | 41 | 68 | 245 (122) | 140 |

*Note*. *Drop out 1* refers to infants who did not finish the first experiment or were excluded afterwards. *Drop out 2* refers to infants who did not start or did not finish the second experiment of the session. The column ‘Habituation Stimulus 1’ shows the numbers of participants who were habituated to Stimulus 1, i.e., *faap* in the native condition, and *sæn* in the non-native condition.

## Procedure and stimuli

### General procedure

The participant was seated on the caregiver's lap, in a three-walled canvas test booth with a canvas ceiling placed in a sound-attenuated room. The distance between the computer monitor (Philips LCD 150P4) on which the visual stimuli were displayed and the child's head was approximately 1.35 meters. The loudspeaker (Tannoy i8) through which the auditory stimuli were played was hidden behind the canvas of the booth and placed underneath the TV screen that showed the visual stimuli. Caregivers wore headphones (Telex, Echelon 20, over-ear headphones with claimed passive noise attenuation of 20dB), through which music was played in order to prevent them from hearing the stimuli and (potentially) influencing their child's behavior. The experiment was monitored and recorded through a video camera that was placed underneath the TV screen.

Caregivers consented to participate during their visit to the lab. In the lab, prior to testing, we explained to the caregiver that two short experiments would be conducted and that the child would hear native and non-native speech sounds, but not in which order this would take place. It was stressed that if a caregiver felt that his/her child was no longer comfortable, they could ask the experimenter to discontinue the experiment at any time. It was also explained that the experimenter could stop the experiment for that same reason. The caregiver was explicitly instructed to 1) not interfere with the experiment, e.g., by pointing to the computer screen, 2) not move their infant during the experimental trials, 3) soothe his/her child nonverbally when necessary. The aim was to test the infants on both contrasts (native and non-native) within one session. Children with odd numbers were assigned to the native contrast first and the non-native contrast second; children with even numbers were presented with the non-native contrast first and the native contrast second

Similar to Houston et al.'s (2007) study, the experiment (both native and non-native conditions) consisted of a habituation phase, in which the infant was habituated on one of the vowels of the pair, a test phase, in which looking times to sequences of trained vowels were compared to those of trained and contrasting vowels, and a pre- and posttest (to measure participants' attentiveness), in which general looking times were measured. Each of these phases included both auditory as well as visual stimuli.

### Stimuli

***Auditory and visual stimuli pre-and posttest.*** During the pre-and posttest infants were presented with both auditory (beep sounds, 330 Hz, duration 250ms, ISI 1000ms) and visual stimuli. Auditory stimuli were played at ~65 dB(A). The visual stimuli were series of three cartoon pictures (e.g., train, car, book) displayed for two seconds on a light blue screen. These pictures series were drawn randomly from a bank of 25 pictures. They could appear in nine different spots, one per row, within an invisible 3 x 3 grid, see Figure 1. After two seconds new pictures appeared at different locations.

***Visual stimuli habituation and test.*** Visual stimuli were eight still pictures of smiling female faces. Half of these pictures were used during habituation and the other half during test. Pictures were presented in randomized order per block of four trials. Between habituation trials a visual attention getter was displayed: a movie of a cute laughing baby. In between test trials a movie of a toddler going down a slide was used as an attention getter, see also Figure 1.

Figure 1

*Visual stimuli presented during the pre- and posttest, habituation and test phase*


*Note:* Picture 1 is an example of the visual stimuli during pre-and posttest; 2 is an example of a female face used during habitation and test trials; 3 is a still of the attention getter between habituation trials and 4 is a still of the attention getter between test trials.

***Auditory stimuli habituation and test.*** Vowel stimuli were presented in CVC syllables (/faːp/ - /feːp/, /sæn/ - /sɛn/). These targets were pseudowords. Tokens of four different female native speakers were obtained. Auditory stimuli were recorded in a sound-attenuated booth of the phonetics lab of Utrecht University,

using a Sennheiser microphone (ME-64) and a digital audio tape recorder (Tascam DA-40). In transferring the recordings onto a computer they were downsampled from 48 kHz to 22.05 kHz. The vowels /a/ and /e/ were presented in /fVp/-syllables, pseudowords *faap* (/faːp/) and *feep* (/feːp/). The four female Dutch speakers were aged between 25 and 35 years of age. They all spoke Standard Dutch and came from the *Randstad area*, a mostly urban area in the central-western Netherlands. Speakers were asked to read out loud a list of 52 words, containing the target pseudowords, as well as other monosyllabic pseudowords and monosyllabic Dutch words with the same vowels (e.g., *gaap* – yawn, *feest* - party). The English [æ] and [ɛ] were presented in /sVn/-syllables, pseudowords /sæn/ and /sɛn/. Tokens were recorded by four female native English speakers, aged between 25 and 35 years. They came from different regions: South-East London, Belfast, Preston (Lancashire) and Manchester. The pseudowords /sæn/ and /sɛn/ were read out loud from a list of 52 words containing the target words and real words (e.g., *have* and *pet*) as was done for the native /aː/ - /eː/ contrast.
Each speaker produced four tokens of each target pseudoword (e.g., /faːp/ and /feːp/). From all four speakers, one token of each target pseudoword per contrast was selected, except from one speaker from whom 2 tokens per target word were selected. This resulted in five tokens of four different speakers for both contrasts. Four tokens were used during habituation and the fifth token (token 2 from speaker 1, see also Figure 2) was used in the test phase (see Procedure), hence the fifth token presented during test was from a familiar speaker because participants heard a different token from that speaker during habituation. All auditory stimuli were played at ~65 dB(A). Tokens selected were the most child-friendly in prosody and speech affect.
The first and second formant frequencies (Hz) were measured with the software program PRAAT (Boersma & Weenink, version 5.4.06) and can be found in Table 2. They were measured at the midpoint of the vowel, where the acoustics are minimally influenced by the surrounding consonants. The Dutch tokens are representative for typical /aː / and /eː/ vowels spoken by a female as was demonstrated by a study of Adank, van Hout & Smits (2004). The English recordings of the four speakers had been created to assess categorical perception of these vowels in children and adults (see Heeren, 2006 for a similar approach on /ɑ/-/aː/). The stimuli we used were the end points of these continua. The English tokens were also judged by two native English listeners (from the London area) and rated as good exemplars of the /ɛ/ and /æ/. These tokens are representative for female British-English /ɛ/ and /æ/ (Deterding, 1997).

Table 2
*Acoustic characteristics of the stimuli*

| Stimulus | Total duration | Vowel duration | F1 | F2 | F3 | F0 | F0 range | Intensity |
|---|---|---|---|---|---|---|---|---|
| **fa:p1** | **536** | **210** | **923** | **1637** | **2871** | **266** | **201-323** | **72** |
| fa:p2 | 587 | 238 | 1071 | 1628 | 2813 | 235 | 185-309 | 72 |
| fa:p3 | 559 | 204 | 969 | 1807 | 2685 | 291 | 197-325 | 72 |
| fa:p4 | 572 | 201 | 941 | 1615 | 3033 | 219 | 142-331 | 72 |
| **fa:p5** | **547** | **205** | **993** | **1646** | **2765** | **265** | **198-346** | **72** |
| **fe:p1** | **570** | **186** | **540** | **2349** | **2894** | **267** | **204-334** | **72** |
| fe:p2 | 621 | 205 | 618 | 2168 | 2851 | 267 | 208-308 | 72 |
| fe:p3 | 631 | 191 | 619 | 2286 | 3061 | 271 | 196-314 | 72 |
| fe:p4 | 622 | 190 | 492 | 2358 | 3046 | 210 | 146-325 | 75 |
| **fe:p5** | **593** | **201** | **513** | **2383** | **2873** | **257** | **192-334** | **72** |
| **sæn1** | **556** | **172** | **962** | **1658** | **3032** | **221** | **200-261** | **74** |
| sæn2 | 605 | 228 | 1036 | 1863 | 3337 | 250 | 217-309 | 70 |
| sæn3 | 484 | 160 | 1000 | 1642 | 2873 | 200 | 186-243 | 70 |
| sæn4 | 594 | 168 | 1015 | 1676 | 2529 | 212 | 180-250 | 71 |
| **sæn5** | **507** | **181** | **976** | **1543** | **2961** | **220** | **200-268** | **76** |
| **sɛn1** | **459** | **148** | **892** | **2016** | **3371** | **245** | **215-297** | **75** |
| sɛn2 | 552 | 167 | 760 | 2061 | 3340 | 267 | 261-293 | 72 |
| sɛn3 | 545 | 163 | 741 | 1782 | 3033 | 234 | 225-266 | 74 |
| sɛn4 | 665 | 167 | 786 | 2252 | 3062 | 222 | 202-259 | 72 |
| **sɛn5** | **499** | **145** | **889** | **2005** | **3315** | **241** | **207-285** | **76** |

*Note.* Vowel duration is given in milliseconds. F1-3 refers to the first three formant frequencies, measured at the midpoint of the vowel. F0, the fundamental frequency (pitch), F0 range (minimum – maximum F0) and Intensity are measured over the total duration of the vowel. Stimuli in **bold** are used during test.

**Experimental procedure**

**Pre- and posttest.** Pre- and posttest were used to gauge participants' general attentiveness. The pretest started immediately when the participant began to look at the screen and had a fixed duration of approximately 24 seconds. The posttest immediately followed the test phase. Looking times to the screen were measured and were taken to refer to listening times (Aslin, 2007). If total looking time to the posttest stimulus was less than 50% of the total looking time to the pretest stimulus, the participant was considered to be showing a general loss of attention. Data of this participant were excluded from further analyses ($n$ = 1, see Participants).

**Habituation and test.** The habituation phase consisted of a maximum of 12 trials, with a maximum of 30 repetitions of a token per trial (ISI of 1 second) resulting in a total duration of approximately 48 seconds. A moving window was used to determine whether the participant had habituated: the mean of trials 1-3 was compared to the mean of trials 4-6. If the mean looking time had decreased with 35%, this was taken as indication that the child had habituated. If looking time had not decreased with 35%, then the mean of the first three trials was compared to the mean looking time of trials 5-7, then 6-8 up to 10-12, as 12 was the maximum number of habituation trials.

The habituation phase started with the attention getter (movie of a cute laughing baby). As soon as the participant looked towards the screen, the experimenter started the first trial. At trial initiation, the visual stimulus changed to one of the smiling female faces, auditory stimuli were played and looking time, was measured. As soon as the participant looked away, the experimenter stopped this measurement and restarted when the infant oriented again to the screen. When the infant looked away for more than two seconds, the trial was terminated and either the next trial started or, if the habituation criterion was reached, the test phase commenced. In the test phase, trials were started and stopped following the same procedure as in the habituation phase. Participants were habituated on either a repetition of /faːp/ or /feːp/ tokens. Within one trial, one token of one speaker was used. Participants were presented with all four voices, in randomized order, i.e. in each block of four trials the participant heard all four voices but in randomized order within the blocks. The order of habituation stimuli (*faap* (/faːp/) or *feep* (/feːp/) was counterbalanced between infants.

The test phase had a fixed number of 12 trials, with a maximum number of 30 tokens per trial, resulting in a total duration of approximately 48 seconds per trial. Test trials consisted either of alternating pseudoword pairs (i.e. /faːp/-/feːp/) or non-alternating pairs (i.e. /faːp/$_i$-/faːp/$_j$; see *Stimuli*. The alternating and non-alternating trials were presented in a semi-fixed order: the first trial could be either alternating or non-alternating, which was counterbalanced. The second trial was non-alternating if trial 1 was alternating and alternating if trial 1 was non-alternating. Three subsequent alternating trials occurred at positions: 5, 8 and 12. The other trials were non-alternating. During the test phase a new token of a familiar speaker was introduced. This was done to ensure that the non-alternating trials (e.g., /fa:p/-/fa:p/) had both a new token (/fa:p/ token-2 from speaker-1) and a familiar token (/fa:p/ token-1 from speaker-1), just like the alternating trials had a new token (/fe:p/ token-1 from speaker-1) and a familiar token (/fa:p/ token-1 from speaker-1) was used. See Figure 2 for a schematic of the procedure.

Figure 2
*Schematic of the testing procedure*

| Pretest | Habituation Phase | Test Phase | Posttest |
| --- | --- | --- | --- |
| 1 | 1 | 1 | 1 |
| Beep sounds | Trial 1 /fa:p/ (T1.S1) | Trial 1 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | Beep sounds |
| 330 Hz | Trial 2 /fa:p/ (T1.S3) | **Trial 2 /fe:p/-/fa:p/ (T1.S1 – T1.S1)** | 330 Hz |
| 250 ms | Trial 3 /fa:p/ (T1.S2) | Trial 3 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | 250 ms |
| ISI 1000 ms | Trial 4 /fa:p/ (T1.S4) | Trial 4 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | ISI 1000 ms |
| | Trial 5 /fa:p/ (T1.S3) | **Trial 5 /fe:p/-/fa:p/ (T1.S1 – T1.S1)** | |
| | Trial 6 /fa:p/ (T1.S2) | Trial 6 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 7 /fa:p/ (T1.S4) | Trial 7 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 8 /fa:p/ (T1.S1) | **Trial 8 /fe:p/-/fa:p/ (T1.S1 – T1.S1)** | |
| | Trial 9 /fa:p/ (T1.S1) | Trial 9 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 10 /fa:p/ (T1.S2) | Trial 10 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 11 /fa:p/ (T1.S4) | Trial 11 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 12 /fa:p/ (T1.S3) | **Trial 12 /fe:p/-/fa:p/ (T1.S1 – T1.S1)** | |

*Note.* In this schematic, the first test trial is non-alternating and the second alternating. The alternative version contains a reversal of these first two trials. In all cases, the remaining three alternating trials have a fixed trial number, namely the 5th, 8th and 12th trial. Alternating trials are printed in bold. In the habituation phase, speakers are presented in randomized orders per block of 4 trials. Token is abbreviated and 'T' and speakers as 'S'.

The test phase started with the attention getter (movie of the toddler on a slide). As soon as the participant looked towards the screen, the experimenter initiated the first trial by pressing a button, which started the trial and looking time measurement. Looking time measurement was the same as during habituation. The changes we made in the design compared to Houston et al.'s study (2007), are summarized in the endnote[1].

**Data coding: online and offline**

**Online coding.** The experimenter sat in a room adjacent to the test room and watched the caregiver and infant through a closed-circuit TV. Looking times to trials were captured online by pressing buttons on a button-box connected to a computer (Asus P4PE). An experiment control application (Zep; Veenker, 2008) was used for presentations of the auditory and visual stimuli and for the data registration.

**Offline coding.** A random subset (approximately 42% of the entire set) of the video recordings was recoded frame-by-frame (one frame had a duration of 30 ms) using Psycode software (developed by Gervain & Filippin), by 2 trained coders who were naive regarding to the design and the purpose of the experiment. The results of the raw and recoded data correlated strongly, $r(100) = .99$, $p < .001$.

**Data analysis and screening**

**Test phase.** To answer the questions whether 1) there was an effect of trial type (alternating versus non-alternating, 2) there were differences between the age groups, and 3) the contrasts (native or non-native), the looking times to alternating and non-alternating trials were analyzed using random effect modeling (SPSS, version 23). The raw looking times to alternating and non-alternating trials were not normally distributed; for this reason, a log transformation (Log10) was performed. After this transformation the skewness (.05) and kurtosis (-.37) values were acceptable. Looking times are reported in Table 3 and Figure 3.

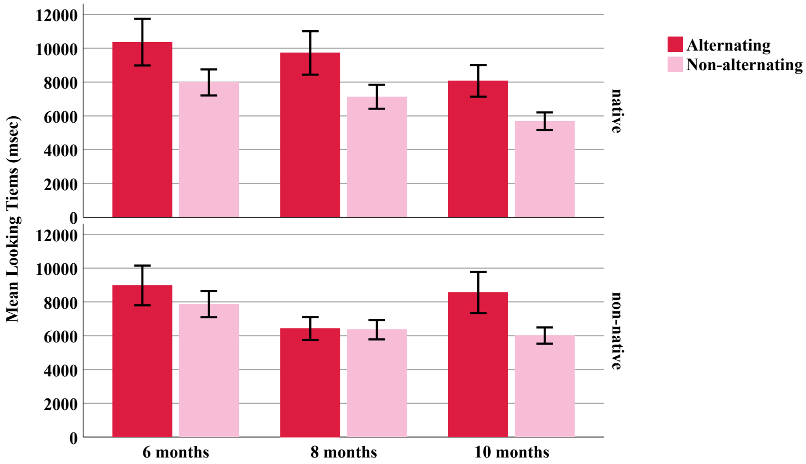
## 2.3 Results

**The role of contrast**

The aim of the study was to investigate the developmental patterns of vowel perception in the first year of life. Our main interest was 1) whether the HVF paradigm could be used to assess the discrimination of speech sound contrasts (rather than word contrasts, as in Houston et al. 2007), and 2) whether non-native discrimination results yielded by the HVF paradigm would agree with PA. Thus, a positive answer to question (1) is a precondition for answering question (2). As infants were tested on both contrasts, we treat contrast (native, non-native) as a within-subject factor. Interactions of contrast with other factors would lead us to analyze the results per contrast separately.

Figure 3

*Mean looking times to alternating and non-alternating trials of both contrasts per age group*




*Note.* Error bars represent the confidence interval (95%).

Table 3

*Looking times to alternating and non-alternating trials of both contrasts*

| Contrast | Age | Alternating trials | | Non-alternating trials | | Statistics | | | Participants | Preference for alternating trials* | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *M* | *(SD)* | *M* | *(SD)* | *F* | *p* | *Cohen's d* | *N* | *n* | % |
| Native | 6 | 10.4 | (8.6) | 7.9 | (6.8) | 13.55 | < .001 | .31 | 38 | 26 | 68 |
| | 8 | 9.7 | (8.6) | 7.1 | (6.7) | 21.74 | < .001 | .32 | 44 | 30 | 68 |
| | 10 | 8.1 | (5.6) | 5.7 | (4.5) | 29.24 | < .001 | .45 | 35 | 27 | 77 |
| | **All** | **9.4** | **(7.9)** | **7.0** | **(6.3)** | **62.70** | **< .001** | **.32** | **117** | **83** | **71** |
| Non-native | 6 | 9.0 | (7.7) | 7.9 | (7.2) | 4.59 | .032 | .15 | 42 | 29 | 69 |
| | 8 | 6.4 | (4.7) | 6.3 | (5.7) | .66 | .416 | | 47 | 25 | 53 |
| | 10 | 8.6 | (8.0) | 6.0 | (4.3) | 21.56 | < .001 | .37 | 39 | 28 | 72 |
| | **All** | **7.9** | **(6.8)** | **6.7** | **(5.9)** | **18.16** | **< .001** | **.18** | **128** | **82** | **64** |

*Note.* Looking times are given in seconds. **Preference for alternating trials* refers to the number of infants who had on average longer looking times to alternating trials then non-alternating trials

**Results of the effect of contrast**

Looking times per trial type (alternating vs. non-alternating) are presented in Table 3. A random effect modeling analysis included participant as random factor and trial number as a repeated effect (covariance structure AR1). The fixed factors were trial type (alternating and non-alternating trials), age (six, eight and ten months) and contrast (native first vs non-native). The model that best fitted the data included the fixed factors trial type, age, contrast and Trial Type*Contrast*Age $F(8, 650) = 3.05$, $p = .002$. The 3-way interaction shows that the effect of trial type on looking time differs across contrasts and ages. We will present separate analyses per contrast in the next sections. The main effect of trial type, $F(1, 1931) = 78.50$, $p < .001$, indicates that infants looked longer to alternating trials than to non-alternating trials, and the main effect of age, $F(2, 242) = 4.50$, $p = .012$, means that overall looking time decreased as age increased. No main effect of contrast was found, $F(2, 233) = 1.78$, $p = .184$, indicating the overall looking times was not significantly different for both contrasts.

**The native contrast**

**Data analysis and screening**

**Habituation phase.** In order to assess whether total looking time and number of trials needed to habituate change as a function of age, univariate ANOVAs and non-parametric tests were conducted. The mean of the total looking times to habituation trials as well as the number of trials required for habituation were assessed across age. The looking times were not normally distributed. Log transformation (Log10) resulted in a distribution that does not differ significantly from a normal distribution (skewness = .09, kurtosis = -.71). The mean number of trials needed to habituate was not normally distributed after log transformation. Therefore, non-parametric testing was conducted on this measure.

**Test phase.** To answer the questions whether 1) infants are able to discriminate the Dutch /a/ - /e/ contrast, 2) there are differences between the age groups, and 3) the habituation stimulus influences discrimination, the looking times to alternating and non-alternating trials were analyzed using random effect modeling (SPSS, version 23). The raw looking times to alternating and non-alternating trials were not normally distributed; for this reason, a log transformation (Log10) was performed and after this transformation the skewness (.12) and kurtosis (.15) values were acceptable.

**Results native contrast**

**Habituation phase**

Although mean looking times showed a tendency to decrease as a function of increasing age (see Table 4), there was no significant main effect of age in a univariate ANOVA with log-transformed mean looking times to habituation trials as dependent variable, $F(2, 109) = 2.49$, $p = .087$. The number of trials needed to habituate did not differ across age groups (Kruskal-Wallis test, $H(2) = 1.84$, $p = .912$).

Table 4
*Looking times to habituation stimuli and numbers of habituation trials in the native and non-native contrasts*

| Age group | Contrast | Total looking time | | Nr. habituation trials | |
|---|---|---|---|---|---|
| | *Native* | *M* | *SD* | *M* | *SD* |
| 6 | | 108 | 56 | 6.6 | 1.3 |
| 8 | | 102 | 58 | 6.9 | 1.6 |
| 10 | | 82 | 42 | 6.7 | 1.3 |
| Total | | 98 | 54 | 6.8 | 1.4 |
| | *Non-Native* | | | | |
| 6 | | 117 | 66 | 6.7 | 1.6 |
| 8 | | 94 | 47 | 7.4 | 1.8 |
| 10 | | 83 | 33 | 6.8 | 1.5 |
| Total | | 101 | 53 | 7.0 | 1.6 |

*Note.* Looking times are given in seconds.

**Test phase**

Looking times are reported in Table 3 and Figure 3. A random effect modeling analysis included Participant as random factor and trial number as a repeated effect (covariance structure AR1). The fixed factors were trial type (alternating and non-alternating trials), age (six, eight and ten months) and habituation stimulus (/faːp/ or /feːp/). Evidence for continuous discrimination would be visible as a main effect of trial type and the absence of a significant Trial Type*Age interaction. Evidence for directional asymmetry would surface as an interaction between habituation stimulus and trial type, or in a three-way interaction of habituation stimulus, trial type and age.

The model that best fitted the data included the fixed factors trial type (alternating and non-alternating trials) and age (six, eight and ten months), which comprises a significant effect of 1) trial type on looking time, $F(1, 916) = 62.59$, $p < .001$, indicating that infants looked longer to alternating trials than to non-alternating trials, and 2) age, $F(2, 111) = 6.04$, $p = .003$, meaning that overall looking time decreased as age increased. The Trial Type*Age interaction was not significant, $F(2, 916) = .90$, $p = .406$. Nonetheless, planned post hoc comparisons (Bonferroni-adjusted) were conducted to assess whether each age group discriminated the contrast. As can be seen in Table 3 and Figure 3, all age groups discriminated the contrast. Moreover, the effect size of trial type increases with age, which implies that discrimination becomes more robust as age increases. The models that included the fixed factor Habituation Stimulus yielded no effect of habituation stimulus, $F(1, 111) = .46$, $p = .500$, no interaction was found between trial type and habituation stimulus, $F(1, 926) = .08$, $p = .783$, and no interaction between trial type, habituation stimulus and age, $F(7, 431) = .90$, $p = .501$.

**Summary native contrast**

Our goal was to determine whether the HVF paradigm (Houston et al., 2007) could be used to tap discrimination of speech sounds. For this reason, we chose an acoustically salient (i.e. an acoustically and articulatorily highly distinctive) vowel contrast. The expectation was that infants across the entire age range (six, eight and ten months) would be able to discriminate the native /a:/-/e:/ contrast. This expectation is confirmed by our results. Importantly, this result entails that infants were able to make generalizations over speakers and attend to those acoustic features that differentiate between /a:/ and /e:/, regardless of the habituation stimulus. Discrimination performance becomes more robust as age increases, as is indicated by an increasing effect size. These results show that speech sound discrimination in infants can be measured successfully through HVF.

**The non-native contrast**

**Data analysis and screening**

**Habituation phase.** Data analysis was the same as for the native contrast. Looking times to habituation trials were not normally distributed. Log transformation (Log10) rendered a distribution which does not differ significantly from a normal distribution (skewness = .32, kurtosis = -.07).

**Test phase.** The raw looking times to alternating and non-alternating trials were not normally distributed, but were not significantly different from a normal distribution after log transformation (skewness = .20, kurtosis = -.04).

**Results non-native contrast**

**Habituation phase**

Habituation times are reported in Table 4. The numerical decrease of habituation time is not supported by a significant effect of age on mean (Log10) looking times to habituation trials (Univariate ANOVA), $F(2, 120) = 1.99$, $p = .141$. The number of trials to habituate did also not differ across age groups, $H(2) = 4.39$, $p = .112$.

**Test phase**

A change over age in discrimination performance is attested when the interaction Trial Type*Age is significant. Whether discrimination is better when trained on one stimulus type would surface as an interaction between habituation stimulus and trial type, or a three-way interaction between habituation stimulus, trial type and age. Table 3 and Figure 3 display the results of the test phase.
The model that best fitted the data included the fixed factors trial type (alternating and non-alternating trials) and age (six, eight and ten months). The significant Trial type*Age interaction $F(2, 1021) = 4.21$, $p = .015$, was explored by Bonferroni-adjusted pairwise comparisons. Infants aged 8 months did not show a significant difference between alternating and non-alternating trials, whereas the other two age groups did, see Table 3 and Figure 3. As can be seen in Table 3, the effect size of trial type is larger for the 10-month-olds than for the 6- and 8-month-olds. The effect of trial type on looking time, $F(1, 1021) = 20.08$, $p < .001$, indicates that infants looked longer to alternating trials than to non-alternating trials. The effect of age, $F(2, 126) = 2.69$, $p = .072$, was marginally significant. The pattern points in the direction of a decrease in looking time as age increased, as was found for the native contrast. This finding aligns with results of the total looking time to habituation trials. The models that included the fixed factor habituation stimulus yielded an effect of habituation stimulus, $F(1, 121) = 18.15$, $p < .001$, indicating that infants had overall longer looking times when trained on /sæn/. The interaction between Trial Type*Habituation Stimulus, $F(1, 1022) = .75$, $p = .388$, and between Trial type*Habituation Stimulus*Age, $F(5, 287) = .56$, $p = .727$, were not significant; whether infants were trained on either /sæn/ or /sɛn/ had no influence on discrimination.

**The effect of the order in both contrasts**

To evaluate whether assessing both native and non-native discrimination within one session impacted infants' performance, we conducted additional analyses in which interactions with contrast order (first or second) were included. Again, random effect modeling was used to analyze the data. Fixed factors were trial type (alternating or non-alternating), age (6, 8 or 10 months) contrast (native or non-native) and contrast order (first or second). The model that best fitted the data included the interaction Trial Type*Contrast*Age*Contrast Order, $F(19, 566) = 2.26$, $p = .002$. The four-way interaction suggests that the effect of trial type on looking times is not the same for both contrasts at all ages. The interaction between Trial Type*Contrast Order*Contrast was marginally significant, $F(5, 727) = 2.09$, $p = .064$, and suggests that the effect of trial type on looking times is not the same for both contrast; see Figure 4. It must be noticed, however, that the interactions Trial Type*Contrast Order and Trial Type*Contrast Order*Age were not significant (all $p < .2$), which means that this four-way interaction should be interpreted with caution.

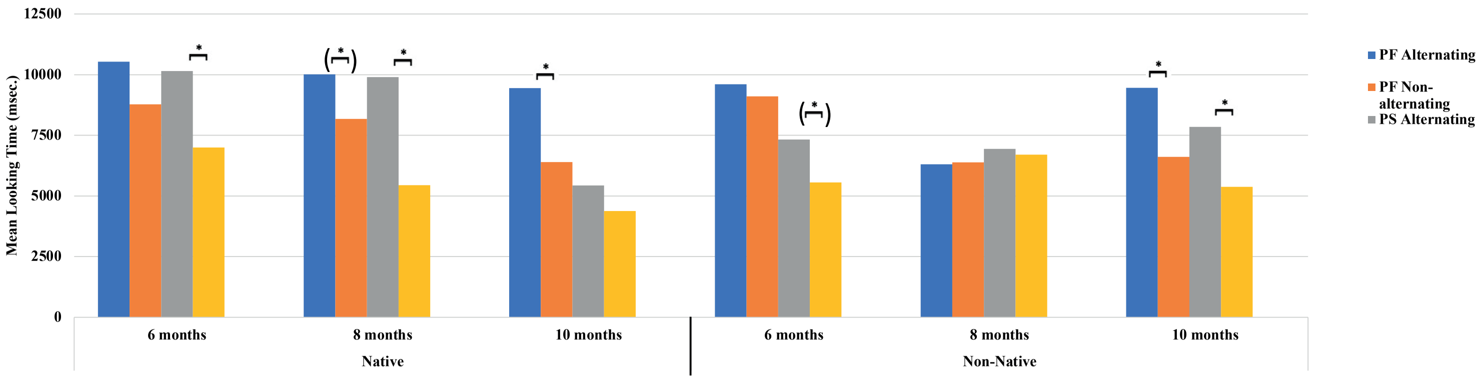
Table 5
*Numbers of participants, and numbers of participants who received a contrast first and second, per age group and contrast.*

| Contrast | Age group | Both contrasts | Contrast presented 1st | Contrast presented 2nd |
|---|---|---|---|---|
| | | *n* | *n* | *n* |
| Native | 6 | 31 (0) | 15 | 16 |
| (/a:/ - /e:/) | 8 | 37 (5) | 22 | 15 |
| | 10 | 26 (0) | 15 | 11 |
| | Subtotal | 94 (5) | 52 | 42 |
| Non-native | 6 | 31 (1) | 16 | 15 |
| (/æ/ - /ɛ/) | 8 | 37 (3) | 15 | 22 |
| | 10 | 26 (1) | 11 | 15 |
| | Subtotal | 94 (5) | 42 | 52 |
| Both contrasts | Total | 188 (10) | 94 | 94 |

*Note. Both contrasts* refers to the number of infants who finished the native and the non-native contrast. The number in parentheses refers to those infants who already heard the contrast at a younger age (at least 2 months earlier) but did not finish the experiment at that earlier session, these infants were not included in the analysis. *Contrast presented first* refers to the number of participants who received the contrast first during the session.

Figure 4

*Mean looking times to alternating and non-alternating trials for both contrasts and both orders of presentation*




*Note.* The abbreviations in the legend, can be read as follows: PF stands for *presented first*, which means that the contrast is presented first during the session. PS stands for: *presented second*, which means that the contrast is presented second during the session. Error bars represent standard errors.

In order to interpret this four-way interaction we used paired sample t-tests to analyze the data per contrast, contrast order and age group. For the 8-month-olds in the non-native contrast condition we can conclude that no matter whether the contrast was presented first, $t(23) = .18$, $p = .862$, or second, $t(22) = .25$, $p = .803$, the contrast was not discriminated. For the 10-month-olds, results are also robust: they discriminated the contrast when the contrast is presented first, $t(23) = 2.77$, $p = .011$, and when it is presented second, $t(14) = 2.89$, $p = .012$. The 6-month-olds, however, seem to discriminate the non-native contrast only when it was presented second, although the effect is only marginally significant, $t(13) = 2.02$, $p = .064$, and not when it was presented first, $t(27) = 1.12$, $p = .273$

Order of contrast presentation also affected 6-month olds' performance on the native contrast. They only discriminated the native contrast when it was presented second, $t(16) = 5.04$, $p < .001$, not first, $t(20) = 1.41$, $p = .174$. The 8-month-olds discriminate the contrast whether it is presented first (albeit marginally so), $t(28) = 1.99$, $p = .055$, or second $t(14) = 3.464$ $p = .004$. For the 10-month-olds the mean difference between alternating and non-alternating trials is only significant if the contrast is presented first, $t(21) = 4.03$, $p = .001$, not second, $t(11) = 1.76$, $p = .107$. From these additional analyses, we conclude that the main findings still hold: 8-month-olds do not discriminate the non-native contrast, whereas the 10-month-olds do.

**Summary non-native contrast**

The data are suggestive of a decline in non-native /ɛ/ - /ae/ discrimination, as predicted by the perceptual attunement hypothesis (e.g., Werker & Tees, 1984; Kuhl et al., 1992; Polka & Werker, 1994): the 6-month-olds discriminated the contrast whereas the 8-month-olds did not. However, the picture is more complex. First, while a significant difference between alternating and non-alternating trials was found for the 6-month-olds, the effect size was small. The claim that 6-month-olds can discriminate the non-native vowel contrast should therefore be made with caution. This is also supported by the additional analyses, which showed that the 6-month-olds did not discriminate the contrast when it was presented first and only marginally so when presented second. Still, this result need not be interpreted as contradictory to PA. Polka & Werker's study (1994) showed that younger infants (4-month-olds) successfully discriminated a non-native vowel contrast, whereas the performance of the 6-month-olds was poorer than predicted. Hence, it is possible that perceptual attunement for vowels starts before or around the age of 6 months.

Secondly, the decline in non-native vowel discrimination was not stable: the 10-month-olds, in contrast to the 8-month-olds, clearly discriminated English

/æ/ and /ɛ/. This aligns with other studies which failed to show a decline in discrimination of non-native speech sounds (e.g., Best & Faber, 2000; Polka & Bohn, 1996; Mazuka et al., 2014; Tyler, Best, Goldstein & Antoniou, 2014). In combination with the results of the 6-month-olds, these findings are suggestive of a U-shaped developmental trajectory. A similar pattern is also reported by Best and Faber (2000). They assessed discrimination abilities of English learning infants, aged 3-5, 6-8 and 10-12 months, using a non-native Norwegian (/i/ - /y/) contrast with which adult listeners had shown difficulty in an earlier study. The 3-5 and 10-12-month-olds did show evidence of discrimination, but the group of 6-8-month-olds did not. The developmental pattern found in the study of Best and Faber thus also shows a 'dip' in performance.

**2.4 Discussion**

We aimed to assess whether perceptual attunement occurs in Dutch-learning infants' vowel perception. Six to ten-month-old infants were tested on a salient native /aː/ - /eː/ contrast and a non-salient, non-native /ɛ/ - /æ/ contrast. We predicted that the native contrast would be discriminated at all ages, since the contrast we used was a salient (acoustically and articulatorily highly distinctive) contrast. Predictions for the non-native contrast were less straightforward. Based on PA, a decrease in discrimination was to be expected. However, some studies have not found a decline in non-native contrasts (e.g., Polka & Bohn, 1996; Best & Faber, 2000).

The outcome of the first study shows that the HVF paradigm designed by Houston et al. (2007) can be used to assess speech sound discrimination abilities. At all three ages (six, eight and ten months) infants clearly discriminated the native /aː/ - /eː/ contrast. These results align with earlier findings that salient native contrasts are discriminated by young infants and that this sensitivity is maintained throughout development (e.g., Best et al., 1995; Werker & Tees, 1984).

The findings of the non-native contrast condition are suggestive of a decline in sensitivity between 6 and 8 months of age. This pattern of discrimination performance matches that of PA (e.g., Polka & Werker, 1994). However, in contrast to the PA prediction, our 10-month-old participants showed sensitivity to the non-native vowel contrast. The 10-month-olds discriminated the non-native contrast regardless of whether the contrast was presented first or second. For the 6-month-olds, however, this was not the case. They only discriminated the non-native contrast when it was presented second. The same was found for the 6-month-olds in the native contrast; here too they performed better when it was presented second. These outcomes suggest that the younger infants need some training with the paradigm.

Furthermore, we did not find evidence for discrimination asymmetry (Polka & Bohn, 1996, 2011). Discrimination was not better when children were habituated on /ɛ/, a less-peripheral vowel, than when they were habituated with the more peripheral /æ/. However, it should be noted that vowel asymmetries are claimed to surface when stimulus presentation changes from the less peripheral vowel to the more peripheral vowel. The HVF procedure might not be suitable to test this, as one vowel type (less or more peripheral) is followed by the other within the same trial. An effect of vowel asymmetry might therefore only be seen in the first non-alternating trial.

It is conceivable that the developmental fluctuations in discrimination attested in this study result from an interaction between the developmental differences between the age groups and the speaker variation used during training. We used multiple exemplars during habituation to facilitate phonetic learning. Variation stimulates phonetic learning as it demands abstraction of invariant features (e.g., Lively et al., 1993). The acoustic variation resulting from speaker variability might have influenced discrimination performance, but in different degrees in each age group. Indeed, there is evidence that the amount of variation needed in order to be helpful during a task differs between age groups (Estes & Lew-Williams, 2015; Singh et al., 2004; Singh, 2008; Vukatana et al., 2015). For instance, Singh et al. (2004) showed that 10.5-month-old-infants can recognize a word in a happy affect after having been trained on that same word in a different speech affect (neutral), whereas 7.5-month-olds could not. A follow-up study (Singh, 2008) showed that this latter group did succeed when more variation in speaker affect was offered during training. The amount of variation needed to yield successful (categorical) discrimination seems to vary along age groups. This might explain the U-shaped pattern suggested by our data; the variation may have been enough for the 10-month-olds to support learning, but not for the 8-month-olds.

We argue that the 6-month-olds discriminate the non-native contrast on the basis of their early perceptual abilities, rather than phonetic perception. In this view, the 6-month-olds in our study have not been able to use the speaker variation to discriminate the non-native contrast. The 10-month-olds, who have acquired native phonetic categories, are able to use their native /ɛ/ category during the experiment: the limited variation offered during training was sufficient for them to make a good estimate of the vowel that was presented during training and maintain a stable representation during test. The 8-month-olds, on the other hand, cannot rely on their early perceptual abilities, nor on their phonetic categories, possibly because they are in the very early stages of PA, i.e. they are in between perceptual strategies. Pursuing this line of reasoning, the amount of variation offered during habituation might not have been sufficient for the 8-month-olds.

This leads to the prediction that the 8-month-olds will be able to discriminate the contrast when (much) more speaker variation is introduced. Another prediction is that 10-month-olds will not perform well when there is less variation during training, i.e. when a single speaker is used. We also predict that the amount of variation will not influence the discrimination performance of 6-month-olds. These predictions remain to be tested.

The findings of our study are suggestive of a U-shaped developmental trajectory. Such a pattern has been observed in earlier work investigating the development of native vowel perception of bilingually raised (henceforth bilingual) infants (Bosch & Sebastián-Gallés, 2003; 2009, but see: Burns et al., 2007; Sebastián-Gallés & Bosch, 2009; Sundara, Polka & Molnar, 2008). Bosch and Sebastian-Gallés (2003) tested 4- and 8-month-old Catalan and Spanish monolinguals and Catalan-Spanish bilinguals on a Catalan, but not Spanish, non-salient (acoustically close), /e/ - /ɛ/ contrast (presented in /deði/ - /dɛði/ pseudowords). They found that Catalan-Spanish bilingual 8-month-olds could not discriminate the contrast, whereas the 4-month-olds could. Their monolingual peers, however, showed the pattern predicted by PA: Monolingual Spanish infants showed a decline in discrimination (as did the bilingual infants), whereas monolingual Catalan infants did not. Subsequently, Bosch and Sebastian-Gallés (2003) tested 12-month-old bilingual Catalan-Spanish infants. This group of infants was able to discriminate the non-salient native /e/ - /ɛ/ contrast. Taken together, the discrimination pattern of the bilinguals over time was U-shaped. In a follow up study with bilingual Catalan-Spanish 6-12-month-old infants (Sebastián-Gallés & Bosch, 2009), however, the U-shaped pattern was only found with another acoustically close (non-salient) contrast /o/ - /u/ and not with the salient /e/ - /u/. The U-shaped patterns in the studies of Bosch and Sebastián-Gallés (2003; 2009) and our study might both be explained by an interaction between developmental processes such as PA, the salience of the contrast and the experimental design employed.

Indeed, there are indications that the failure of the bilingual Catalan-Spanish 8-month-olds to discriminate the native contrasts is related to the experimental paradigm employed in relation with non-salient stimuli, such as the Catalan /e/ - /ɛ/ contrast. Albareda-Castellot, Pons and Sebastián-Gallés (2011) tested 8-month-old monolingual Catalan and Spanish and bilingual Catalan-Spanish infants on the same vowel contrast /e/ - /ɛ/ as was used in Bosch and Sebastián-Gallés (2003). Instead of a familiarization preference procedure, they used anticipatory eye movement to measure discrimination performance. In their experiment, the performance of the bilingual infants was similar to that of their monolingual peers; both the Catalan monolinguals and the Catalan-Spanish bilinguals discriminated the contrast, while in the study of Bosch and Sebastián-

Gallés (2003) the bilinguals failed. The difference between the findings of these two studies might stem from the fact that the familiarization preference paradigm, used in Bosch and Sebastián-Gallés (2003), relies on recovery of attention (increase in looking time) elicited by a vowel change, e.g., a change from /deði/ to /dɛði. However, Albareda-Castellot et al. (2011) indicate that an estimated 66% of all Catalan words have Spanish cognates. Cognates are similar sounding words which often include a vowel difference, e.g., /ʃukulatə/ - /tʃokolate/ (*chocolate*). Hence, vowel change does not alter word meaning in many cases and for Catalan-Spanish learning infants, these vowel changes are very common. A paradigm based on the surprise effect of a vowel change might thus not have captured the bilingual 8-month-olds' true sensitivity to this non-salient contrast. So, we argue that the lack of discrimination of the 8-month-old Catalan-Spanish bilinguals in the study of Bosch and Sebastián-Gallés (2003, 2009) is due to the interaction between 1) the contrast being acoustically and articulatory highly similar, 2) PA and 3) insufficient sensitivity to the paradigm. The lack of discrimination shown in our 8-month-old monolingual Dutch infants is argued to also be due to type of contrast used (non-salient and non-native), PA and task elements, i.e., insufficient speaker variation during the habituation phase.

In our study, task effects might also explain the large variations in looking times (resulting in small effect sizes) for the 6- and 8-month-olds in native conditions, and the 6-month-olds in the non-native condition. One feature of our procedure that might explain this, is the relatively long ISI (1000 ms). The long ISI might have interfered with the younger groups' discrimination performance, due to their limited short-term memory. Some evidence for this interpretation comes from other studies on vowel perception using a habituation paradigm. Studies that did not find discrimination by very young infants of a non-salient native vowel contrast (Liu & Kager, 2016) or non-native vowel contrasts (Mazuka et al., 2014) had long ISIs (1500 ms). In contrast, a study that did find discrimination by very young infants of a non-salient non-native vowel contrast had a shorter ISI (750 ms, Best & Faber, 2000). Given that working memory capacity increases with age, the effect of shorter ISI duration might be most pronounced at 6 months (Pelphrey et al., 2004). Predictions following from this are that the 6-month-olds will show better performance as a group when ISI is reduced in both native and non-native contrasts.

The results of this study have shown that infants in the process of PA are still able to discriminate a non-native contrast. As Werker (1994, p.106) states, 'developmental changes do not result in a permanent loss' of discrimination abilities. However, during and after the process of PA, discrimination performance might depend to a greater extent on the experimental design.

**Ethics statement**

Informed consent was obtained from the parents before testing; consent and participation could be retracted at any time. The authors declare that the research was conducted in accordance with APA ethical standards as well as *The Netherlands Code of Conduct for Scientific Practice* issued in 2004 (revised in 2018) by the Association of Universities in the Netherlands (VSNU).

**Acknowledgements**

We thank the caregivers of our participants for their contributions to our study. We also thank Sule Kurtçebe, Tinka Versteegh, Lorijn Zaadnoordijk en Joleen Zuidema, for assistance in data collection, and Mattis van den Bergh and Kirsten Schutter for their statistical advice. This study was funded by The Netherlands Organization for Scientific Research (NWO, grant nr. 360-70-270).

1 We made six changes to the HVF paradigm as originally described by Houston et al. (2007). First, the test phase was reduced in length, as we know from experience that Dutch children are not always able to sit through experiments that have the same duration as those conducted with children from the US. So, instead of 14 test trials, we have 12. The number of alternating trials has remained the same, however.
Secondly, the target pseudowords were not presented with synchronized audiovisual presentation, as our lab equipment did not allow us to do so. Instead, we used still pictures of smiling female faces. Even if the smiles of the smiling female faces interfered with the perception of our CVC pseudowords, they would have affected all vowels equally, since none of the vowels we used are associated with a closed spread position of the mouth; see also Figure 1.
Thirdly, we used multiple speakers in the habituation phase rather than a single speaker. We used multiple speakers to make the task comparable to the demands of natural speech.
Fourth, the habituation criterion was set at 65% instead of 50%. Dijkstra & Fikkert (2010) who used HVF to assess consonant perception, also used the 65% criterion. Other studies assessing speech sound discrimination abilities have also relied on the 65% criterion (e.g., Liu & Kager, 2014, 2015, 2016; Mazuka et al., 2014; Pater, Stager & Werker, 2004). In our opinion, this criterion allows for tracing a decrease in attention without introducing a risk that infants tune out entirely (which would lead to unwanted data reduction).
Fifth, the pre-test and post-test had a fixed duration. Infants can have very short looking times in the initial phase of an experiment (Colombo & Mitchell, 2009). A fixed duration solves this problem and makes a pre- and posttest with fixed duration a good measure of arousal.
Sixth, we changed the look-away time criterion to 2 seconds instead of 1 second, in light of participants' ages and the stimuli we used. A one second criterion might be too short for the youngest infants to recover from their look away. Many studies assessing speech sound discrimination use the 2 second criterion (e.g., Best & Faber, 2000; Bosch & Sebastián-Gallés, 2003; Tyler, Best, Goldstein & Antoniou, 2014).

# Chapter 3

# A step forward: Bayesian hierarchical modelling as a tool in assessment of individual discrimination performance

**Abstract**

Individual assessment of infants' speech discrimination is of great value for studies of language development that seek to relate early and later skills, as well as for clinical work. The present study explored the applicability of the hybrid visual fixation paradigm (Houston et al., 2007) and the associated statistical analysis approach to assess individual discrimination of a native vowel contrast, /aː/ - /eː/, in Dutch 6 to 10-month-old infants. Houston et al. found that 80% (8/10) of the 9-month-old infants successfully discriminated the contrast between pseudowords *boodup* - *seepug*. Using the same approach, we found that 12% (14/117) of the infants in our sample discriminated the highly salient /aː/ - /eː/ contrast. This percentage was reduced to 3% (3/117) when we corrected for multiple testing. Bayesian hierarchical modeling indicated that 50% of the infants showed evidence of discrimination. Advantages of Bayesian hierarchical modeling are that 1) there is no need for a correction for multiple testing and 2) better estimates at the individual level are obtained. Thus, individual speech discrimination can be more accurately assessed using state of the art statistical approaches.

Published as de Klerk, M., Veen, D., Wijnen, F., & de Bree, E. (2019). A step forward: Bayesian hierarchical modelling as a tool in assessment of individual discrimination performance. *Infant Behavior and Development*, 57, 101345. https://doi.org/1016/j.infbeh.2019.101345

EdB, MdK and FW designed the study. MdK tested the infants and conducted the group level analyses. DV conducted the individual level analyses. MdK wrote and revised the paper, with input and feedback from EdB, DV and FW. DV wrote and revised the text concerning the individual analysis with input and feedback from MdK, EdB and FW.

## 3.1 Introduction

Early speech discrimination is assumed to be vital for children's language acquisition, as it is a first step into the formation of speech sound categories. These, in turn, are necessary for word learning (e.g., Tsao et al., 2004). These past decades have seen a significant increase in our understanding of the development of speech perception in infants (see for recent reviews Maurer & Werker, 2014 and Tsuji & Cristia, 2014 ). However, the majority of studies have based their conclusions on group data. It has thus far turned out difficult to make claims about individual performance and development, even though this type of information is critical for understanding individual developmental trajectories as well as clinical questions. It seems that only one study has addressed this matter so far (Houston et al., 2007). In the present study, we use a variant of Houston et al.'s hybrid visual fixation paradigm (HVF), and we describe and evaluate a new approach for assessing individual infants' phoneme discrimination.

Infant speech discrimination can only be measured indirectly. A frequently used behavioral method is a habituation paradigm. In such paradigms, looking time is the preferred dependent variable. Generally, in habituation paradigms infants are habituated on a set of stimuli (A), followed by a test phase in which infants are tested on new set of stimuli (B), i.e., the 'dishabituation' or 'change' trials. If infants are sensitive to the difference between A and B, longer looking times are expected to the novel stimuli (B) (Sokolov, 1963). Studies often employ designs with only 2-4 test trials (see Colombo & Mitchell, 2009 for a review). This can lead to interpretation difficulties, because infant data is, without exception, noisy. Group results often show large individual variation in looking times. This reflects substantial inter-individual variation, comprising overall long or short lookers. It also reflects intra-individual variation. This variation may result from a variety of factors, both infant-internal, such as gas in the digestive system, tiredness, developmental level, memory capacity, attentiveness, motivation, and external factors, such as sounds other than the stimuli, stimulus complexity, and task demands. Hence, the length of a look does not merely reflect the mental processing of the stimulus, and thus does not unequivocally mirror habituation or dishabituation (Oakes, 2010). In order to deal with the noise, researchers typically collapse data over individuals. However, the HVF paradigm (Houston et al., 2007) uses 14 test trials instead of 2-4 test trials, which in principle allows for individual assessment, as the higher number of test trials will boost the signal-to-noise ratio.

Recently, there has been a growing interest in explaining individual differences in infants' early speech perception, viz., word segmentation and speech sound discrimination skills (see Cristia et al., 2014 for a review). A frequently used approach to explain individual differences is to use follow-up data, such as later vocabulary size, reading scores or other skills to (retro- or prospectively) predict infants' looking times (e.g., Altvater-Mackensen et al., 2015; Cristia, 2011; Junge & Cutler, 2014; Melvin et al., 2017; Molfese 2000; Newman et al., 2006). For instance, Newman and colleagues (2006) found that 24-month-old toddlers with larger vocabulary sizes were better at speech perception tasks in infancy than their peers with smaller vocabularies. Although the reported correlations between looking time data and later language, cognitive or social measures, e.g., vocabulary size, social interaction, social economic status (e.g., Altvater-Mackensen & Grossmann, 2015; Melvin et al. 2017) are sometimes low to moderate, the meta-analysis of Cristia et al. (2014) shows that early speech perception skills have a predictive value of later language skills.

Even though there is a (weak) positive relation between early looking time data and later language, cognitive or social measures, this type of analysis does not provide information about an individual child's ability to discriminate speech sounds or segment words. There are three reasons why individual data collected with the traditional discrimination paradigms cannot provide this information. First, individual data is likely to show that some infants have, on average, longer looking times to the familiarized, than to the new stimuli (Houston-Price & Nakai, 2004). This could be due to some infants having reached the habituation criterion without having fully encoded the stimulus (Aslin & Fiser, 2005); as a consequence they do not look longer to the new stimulus. However, such a looking pattern does not imply that they cannot discriminate A from B (e.g., Aslin & Fiser, 2005; Houston-Price & Nakai, 2004). This implies that the *direction* of the difference in raw looking times cannot be used to infer discrimination. Second, it is not a priori clear that a larger looking time difference between stimuli A and B is evidence for better discrimination performance, and a smaller difference reflects poorer discrimination (Aslin & Fiser, 2005), because there is no clear conceptualization of looking time duration and discrimination. Third, although Houston and colleagues found high test-retest reliability (Houston et al., 2007), this test-retest reliability was found to be extremely variable across different experiments in a multi-center study by Cristia, Seidl, Singh and Houston (2016). Across the three participating labs 12 speech perception experiments were conducted, which included testing and retesting of 5-12-month-old infants within 18 days. Some of the labs found significant correlations between performance of the infants tested on two separate days, whereas others did not. One of the labs used the HVF paradigm to assess speech sound discrimination skills of a vowel contrast (/i - u/), a consonant contrast (/sa - ʃa/) and a word contrast (*boodup-seepug*). Here too, test-retest reliability was extremely variable across experiments; there were high test-retest correlations for vowel and consonant contrasts, but not for the word contrast. In conclusion, it appears highly challenging, if not impossible, to infer discrimination at the individual level, based on raw looking time data.

Evidence for discrimination at the individual level might be found if infant data could be modeled taking into account the individual variances as well as the autoregressive effect, i.e. the correlations in noise between trials. Houston and colleagues attempted to tackle these issues by using the HVF paradigm and applying statistical analyses on the individual data and test trials. However, the statistical approach by Houston and colleagues (2007), testing each infant individually using a classical frequentist approach, ignores chance findings based on multiple testing, and misses the opportunity to gain strength in analyses by taking the hierarchical structure of the data into account. Bayesian hierarchical modelling could be a solution to overcome the multiple testing impracticality (Gelman et al., 2012). Additionally, adding (hierarchical) information to the individual estimates reduces noise, and also reduces the number of cases for which estimated effects are found in the wrong direction, type-S (sign) errors, and inflated estimated effects, type-M (magnitude) errors (Gelman & Tuerlinckx, 2000).

Houston and colleagues (2007) developed the HVF paradigm to assess discrimination skills at the individual level. HVF is a habituation paradigm that includes more test trials (14 trials) than typically used in habituation studies, facilitating individual analysis. In their study, Houston et al. tested ten 9-month-olds on the pseudowords *boodup* and *seepug*. These stimuli could a priori be regarded as highly discriminable for infants this age. Infants were habituated on one of the words (e.g., *boodup*) and then tested on alternating (*boodup-seepug*) and non-alternating (*boodup-boodup*) trials. Data was analyzed using a linear regression model with autoregressive (AR1) error structure. Eight out of the ten infants were able to discriminate the contrast, as indicated by a significant difference in looking time between alternating (*boodup-seepug*) and non-alternating test trials (*boodup-boodup*, *seepug-seepug*). The paradigm has successfully been used by other researchers assessing speech (sound) discrimination skills of infants *at group level* (Chapter 2; Cristia, et al., 2016; Dijkstra & Fikkert, 2011; Horn et al., 2007; Liu & Kager, 2015). The design and analysis applied by Houston et al (2007) might be suitable for assessing individual performance in speech sound discrimination as well.

**Current study**

In the current study, we applied an adapted variant of Houston et al.'s procedure to infants' speech sound discrimination: we used a Dutch vowel contrast (/aː/-/eː/).

Smits, Warner, McQueen and Cutler (2003) found that when native adults speakers of Dutch were presented with /a:/ and /e:/ in syllable medial position, vowel /e:/ was classified only once as /a:/ out of 1548 instances and the opposite error never occurred. This indicates that the contrast is easy to discriminate by adults. Results of Chapter 2 showed that groups of Dutch learning 6, 8, and 10-month-old infants can indeed discriminate this contrast; moreover, performance increased with age (see *Results*, 3.1). These findings are in line with theories of speech perception which predict good or age-related enhancement of discrimination of highly distinctive native speech sounds contrasts (Tsuji & Cristia, 2014; Maurer & Werker, 2014). The current study investigates outcomes at the *individual* level rather than the group level, using the data from our previously published paper (Chapter 2). The primary research question is whether we can obtain similar results at the individual level as Houston, et al. (2007). We expect that a large percentage of individual infants will show evidence of discrimination, mirroring the findings reported by Houston et al. (2007).

In addition, we explore the application of Bayesian hierarchical modeling to our discrimination data, and compare it to Houston et al.'s statistical approach. Bayesian hierarchical modeling might provide better estimates of individual infants' discrimination performance than classical regression modeling: Using a Bayesian hierarchical analysis allows us to obtain estimates for each of the individual and group parameters in one model without the need to correct for multiple testing (Gelman et al., 2012). If it can be assumed that infants within the same age group belong to the same population -i.e. infants are exchangeable *within* age groups but not *between* age groups- a hierarchical (multilevel) structure is thus a more powerful approach.

## 3.2 Method

### Participants

A total of 117 typically developing, monolingual Dutch 6-10-month-old infants participated. In addition, 53 infants (31% of total recruited) were tested, but their data was not included for analysis because of behavior during test (crying, extreme restlessness, $n$ = 31), technical errors ($n$ = 12), failure to meet the habituation criterion ($n$ = 5; see Procedure), parental interference ($n$ = 3), or ear infection at time of testing ($n$ = 3). An overview of the ages and drop-out rates is provided in Table 1. Note that none of the infants were excluded for failing to meet the pre- and posttest criterion (see *Procedure*). Parents provided active informed consent before participation.

Table 1
*Numbers of participants, mean ages and age ranges, and drop-out rate per age group*

| Age group | Age range | Age (days) | Infants tested | Infants included | Drop-out rate |
|---|---|---|---|---|---|
| | *month.days* | *M (SD)* | *N =* | *n =* | *n = (%)* |
| 6 | 6.1 - 6.30 | 203 (8.4) | 59 | 38 | 21 (35) |
| 8 | 8.0 - 8.30 | 259 (6.5) | 66 | 44 | 22 (33) |
| 10 | 10.3-10.30 | 320 (12.9) | 45 | 35 | 10 (22) |
| Total | | | 170 | 117 | 53 (31) |

**Stimuli**
Both auditory as well as visual stimuli were presented in each phase of the procedure. Similar to Houston et al.'s (2007) study, the experiment consisted of a habituation phase, a test phase, and a pre- and posttest to measure participants' general attentiveness. For more detailed information about the stimuli we refer to Chapter 2.

During the *pre-and posttest* infants were presented with both auditory (beep sounds, 330 Hz, played at 65 dB(A), duration 250ms, ISI 1000ms, total duration of ~24 seconds) and visual stimuli. The visual stimuli were three cartoon pictures pseudo-randomly selected from a set of 25 (e.g., train, car, book), displayed for two seconds on a light blue background. These pictures appeared in three different, randomly selected positions within an invisible 3 x 3 grid, see Figure 1. Every two seconds new pictures appeared at different locations.

In both the *habituation* and *test phase* participants heard a speech token repeatedly (with a maximum of 30 repetitions) while being shown one of six still pictures of smiling female faces. The faces were displayed in a random order, one face per trial. Houston and colleagues used movies of females producing the words: we could not do the same because of technical limitations. Between habituation trials a visual attention getter was displayed: a video of a cute laughing baby. The attention getter shown between test trials was a video clip of a toddler going down a slide (see Figure 1 for the visual stimuli). Auditory stimuli were native vowels /a:/ and /e:/, embedded in pseudowords *faap* (/fa:p/) and *feep* (/fe:p/). Five tokens of four female Dutch native speakers (aged between 25 and 35 years of age) were obtained. From three speakers one token was selected. From the fourth speaker two tokens were selected, one of which was used during the habituation and test phase and the other only during test phase (see Figure 3 for an overview).

The four different speakers that were used during the habituation phase were presented per block of 4 trials, in randomized order. All auditory stimuli were played at ~65 dB(A). Tokens were spoken in a child-friendly manner.

Figure 1
*Visual stimuli presented during the pre- and posttest, habituation and test phase*


*Note:* Picture 1 is an example of the visual stimuli during pre-and posttest; 2 is an example of a female face used during habitation and test trials; 3 is a still of the attention getter between habituation trials and 4 is a still of the attention getter between test trials.

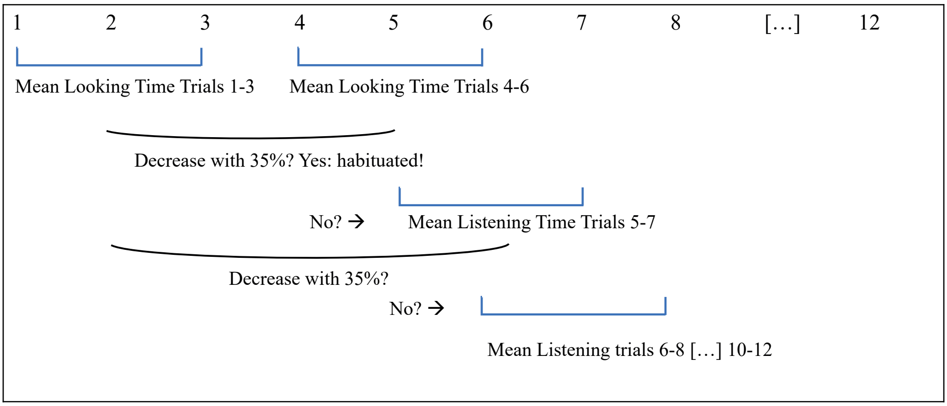
**Procedure**
Infants were seated on their caretaker's lap in a sound-attenuated booth. As soon as infants looked towards the computer screen in front of them, the experimenter started the first trial. In each trial, the time the participant was looking at the screen was measured. Whenever the participant looked away for 2 consecutive seconds, the trial was ended; a new one started when the infant oriented to the screen again. There was no minimum looking time to the screen. Looking times were coded online using a button box connected to the computer controlling the experiment and acquiring data.
Pre- and posttest were used to gauge participants' general attentiveness. If total looking time to the posttest stimulus was less than 50% of the total looking time to the pretest stimulus, the participant was considered to be showing a general

loss of attention and was discarded for analysis. This was never the case in our sample (see *Participants*).

The habituation phase consisted of a maximum of 12 trials, with a maximum of 30 repetitions of a token per trial (ISI of 1 second) resulting in a total duration of approximately 48 seconds. A 65% habituation criterion was used to determine whether the participant had habituated. To determine whether the habituation criterion was met, a moving window was used (Figure 2). The mean looking times of the first three trials (1-3) was compared to the subsequent three trials (4-6): if looking time had decreased by (minimally) 35%, the criterion was met. If not, the mean looking time of trial 1-3 was compared to 5-7, 6-8, etc., and the same criterion applied, up until the final subset 10-12. Infants who did not meet the habituation criterion were not included in data analysis (n = 5, see *Participants*). The selection of habituation stimuli (*faap* (/fa:p/) or *feep* (/fe:p/)) was counterbalanced between infants. Infants were presented with all four voices, in randomized order: in each block of four trials the infant heard all four voices but in randomized order within the blocks (see Figure 3).

Figure 2

*Visual depiction of the assessment of the (65%) habituation criterion*




The test phase included a fixed number of 12 trials, with a maximum number of 30 tokens per trial, resulting in a duration of approximately 48 seconds per trial. Houston et al. (2007) used 14 test trials (10 non-alternating and 4 alternating). We reduced the number of test phase trials and thus duration, because we know from experience that Dutch infants are not always able to sit through experiments that have the same duration as those conducted with infants in the US. Of these 12 test trials, four were alternating (e.g., /fe:p/-/fa:p/), and 8 non-alternating (e.g.,

/faːp/-/faːp/). The alternating and non-alternating trials were presented in a semi-fixed order: the first trial could be either alternating or non-alternating, which was counterbalanced. Three subsequent alternating trials occurred at positions: 5, 8 and 12. During the test phase a new token of one familiar speaker was introduced, either non-alternating or alternating (see Figure 3).

Figure 3

*Schematic overview of the experimental procedure with reference to the auditory stimuli only*

| Pretest | Habituation Phase | Test Phase | Posttest |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| Beep sounds | Trial 1 /fa:p/ (T1.S1) | Trial 1 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | Beep sounds |
| 330 Hz | Trial 2 /fa:p/ (T1.S3) | **Trial 2 /fe:p/-/fa:p/ (T1.S1 – T1.S1)** | 330 Hz |
| 250 ms | Trial 3 /fa:p/ (T1.S2) | Trial 3 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | 250 ms |
| ISI 1000 ms | Trial 4 /fa:p/ (T1.S4) | Trial 4 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | ISI 1000 ms |
| | Trial 5 /fa:p/ (T1.S3) | **Trial 5 /fe:p/-/fa:p/ (T1.S1 – T1.S1)** | |
| | Trial 6 /fa:p/ (T1.S2) | Trial 6 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 7 /fa:p/ (T1.S4) | Trial 7 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 8 /fa:p/ (T1.S1) | **Trial 8 /fe:p/-/fa:p/ (T1.S1 – T1.S1)** | |
| | Trial 9 /fa:p/ (T1.S1) | Trial 9 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 10 /fa:p/ (T1.S2) | Trial 10 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 11 /fa:p/ (T1.S4) | Trial 11 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 12 /fa:p/ (T1.S3) | **Trial 12 /fe:p/-/fa:p/ (T1.S1 – T1.S1)** | |

*Note.* In this example, the first test trial is non-alternating and consequently the second is alternating. The remaining three alternating trials have a fixed number, viz. the 5th, the 8th and 12th trial. Alternating trials are printed in bold. Token is abbreviated as 'T' and Speakers as 'S'.

### 3.3 Results

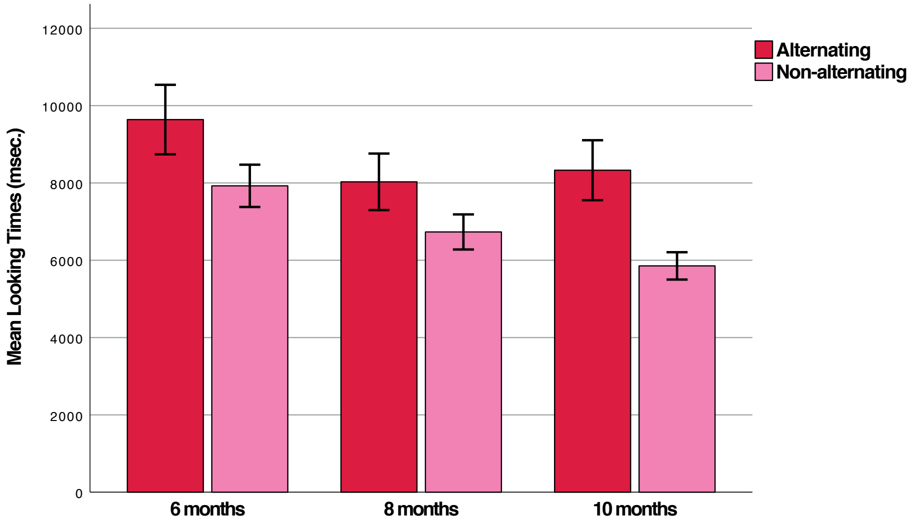
#### Summary of the group data (Chapter 2)

The group-based data is presented in Figure 4 and Table 2. Mixed Modeling using SPSS (version 23) with participants as random factor, trial number as a repeated effect (covariance structure AR1), and trial type (alternating vs. non-alternating) and age as the fixed factors showed that at group level, infants between 6-10 months of age discriminated /faːp/ from /feːp/, at group level. In the current study we focus on the individual data.

Table 2
*Looking times (seconds) to alternating and non-alternating trials*

| Infants | Age group | Alternating trials | | Non-alternating | | Non-alternating | | |
|---|---|---|---|---|---|---|---|---|
| *N* | | *M* | *(SD)* | *M* | *(SD)* | *F* | *P* | *Cohen's d* |
| 38 | 6 | 10.4 | (8.6) | 7.9 | (6.8) | 13.55 | < .001 | .31 |
| 44 | 8 | 9.7 | (8.6) | 7.1 | (6.7) | 21.74 | < .001 | .32 |
| 35 | 10 | 8.1 | (5.6) | 5.7 | (4.5) | 29.24 | < .001 | .45 |
| **117** | **All** | **9.4** | **(7.9)** | **7.0** | **(6.3)** | **62.70** | **< .001** | **.32** |

Figure 4
*Raw mean looking times (milliseconds) to alternating and non-alternating trials per age group*




*Note.* Error bars represent the confidence intervals (95%).

**Data screening**

The raw looking times to alternating and non-alternating trials were not normally distributed; for this reason, a log transformation (Log10) was performed. After this transformation the skewness (.123, *SE* = .065) and kurtosis (.150, *SE* = .131) values were acceptable. We refer to the supplementary files for histograms of the raw and log transformed data (https://osf.io/ebrxy/).

**Analysis 1: Linear regression model with autoregressive (AR1) error structure**

To assess individual performance, we used the same regression model with autoregressive effect as Houston et al. (2007, Figure 5). Looking times and statistical outcomes per infant are reported in Appendix A. Individual analyses show that condition effects were significant for 14 participants, implying that only 12% of the infants were able to discriminate between alternating and non-alternating trials. When we correct for multiple testing using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995), this number decreases to 3 infants (3/117), a mere 3%.

Figure 5

*The model used for the linear regression with autoregressive errors*

$$y_t = b_0 + b_1 C_t + a_t$$

$$a_t = \begin{cases} \phi_1 a_{t-1} + e_t, \text{ if } t \geq 1 \\ 0, \text{ otherwise} \end{cases}$$

*Note.* Subscript $t$ denotes the trial number $t=1,\ldots,T$; $y$ denotes the looking time of the trial; $C$ denotes the condition (alternating or non-alternating) of the trial; $e$ denotes the error term; $\phi_1$ denotes the autoregressive factor. In this model $b_1Ct$ accounts for the influence of the condition and $b_1$ is interpreted as the difference in looking time for the two conditions. The dependence on the looking time of the previous trial is found in the specification of the error structure, $\phi_1 a_{t-1}$. The error in the current time, $a_t$, point is dependent on the error of the previous time point, $a_{t-1}$, except for $a_1$, because $a_1$ is the first trial. There is no carry-over effect from the previous trial and no autoregressive effect.

Our results do not align with the results of the study of Houston et al. (2007), in which 80% (8/10) of the 9-month-old infants successfully discriminated the contrast. Applying the Benjamini-Hochberg correction for multiple testing to Houston et al.'s (2007) data did not make a difference in their outcomes, because of the few participants tested and the large effect of condition on looking times. Nevertheless, an analysis without having to correct for multiple testing is desirable and Bayesian modeling could be a solution.

**Analysis 2: Hierarchical bayesian modeling**

The analyses used in the paper by Houston et al. (2007) rely on separate regression analyses for each individual child. However, if we assume that infants are exchangeable within the same age group, that is, that they come from the same population, an alternative and more powerful approach is to model their looking times in a hierarchical (multilevel) structure. By modeling both the individual and group effects in one analysis instead of doing so for 117 separate analyses, one for each individual, part of the observed variance could be explained at the group level instead of trying to explain all variance at the individual level. As a result, we will have reduced uncertainty in our estimates for the individual parameters (Gelman, 2006). Moreover, by using a Bayesian hierarchical analysis, we are able to obtain estimates for each of the individual and group parameters in one model without the need to correct for multiple testing (Gelman et al., 2012).

In our Bayesian hierarchical regression, we modelled the individual infant data in three groups based on their age (6, 8 and 10 months). We used the same model as before, namely a regression model with an AR1 error structure, with Log10 transformed looking times as outcomes and condition (alternating or non-alternating trial) as predictor. For all groups we obtained both group and individual estimates for the intercept (looking time alternating trials), the condition (difference in looking time between alternating and non-alternating trials) and the AR1 effect. Details on the priors, estimation, model fit and sensitivity analyses are given in the supplementary files on the Open Science Framework webpage for this study at (https://osf.io/ebrxy/). In short, we achieve a good model fit.

The parameter of interest was the condition parameter. This parameter allowed us to establish whether the looking times differed between the alternating and non-alternating condition for the individual infants. To keep the decision criterion as similar as possible to the previously described analyses, we checked how many of the infants included the value 0 in their 95% credibility interval (CI) for the condition parameter. For the 95% CI (the 0.025 and 0.975 quantiles of the posterior sample) we regard this interval as having a 95% probability of containing the unknown parameter value. In contrast, the 95% confidence interval in frequentist statistics relates to (potential) replications of the experiment and expresses the expectation that the interval contains the true parameter estimate in 95% of the experiments. In our study, the percentages of infants whose 95% CI did not include 0 are displayed per age group in Table 3. For the 10-months-olds we found that 77% discriminated between the alternating and non-alternating condition, and 53% of the 6-month-olds did, whilst for the 8-month-old infants this was only 27%.

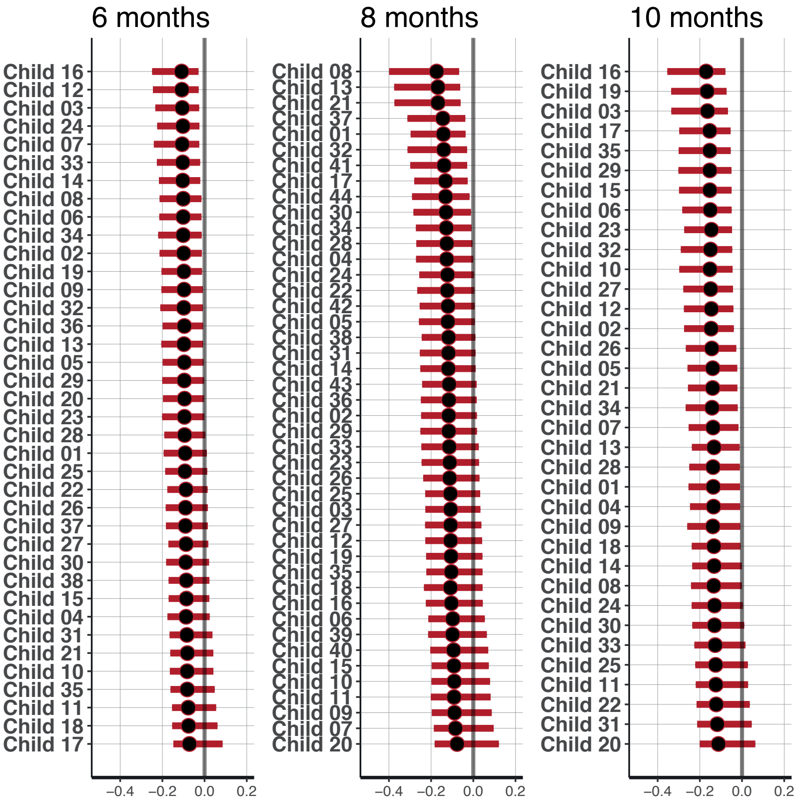
Table 3
*Number and percentage of infants that discriminate the contrast significantly per age group and of infants that did not include the value 0 in their 95% credibility interval (CI)*

| | | Frequentist (non-hierarchical) modeling | | Bayesian hierarchical modeling |
|---|---|---|---|---|
| Age group | Participants | Uncorrected successful discrimination (%) | Corrected successful discrimination (%) | Infants without 0 in their 95% CI (%) |
| 6 | 38 | 2 (5) | 0 (0) | 20 (53) |
| 8 | 44 | 4 (9) | 2 (5) | 12 (27) |
| 10 | 35 | 8 (23) | 1 (3) | 27 (77) |
| Total | 117 | 14 (12) | 3 (3) | 59 (50) |

Figure 6 shows the results of the hierarchical model for each individual per age group. Credibility intervals for the 8-month-old infants show larger uncertainty for the estimates than for the other two age groups, especially the 6-month-olds. The group-estimated effect of condition, depicted in the left panel of Figure 7, increases with age. The estimated random effect for condition is largest in the 8-month-old group, which can be seen from the variance estimates in the right panel of Figure 7. Because the infants of the 8-month-old group differ more from one another than the infants in the other age groups, less shrinkage of estimates occurs and we remain more uncertain about their estimated condition effects. This outcome is visible in the larger credibility intervals for the infants in age group 8 compared to the other two age groups.

Figure 6

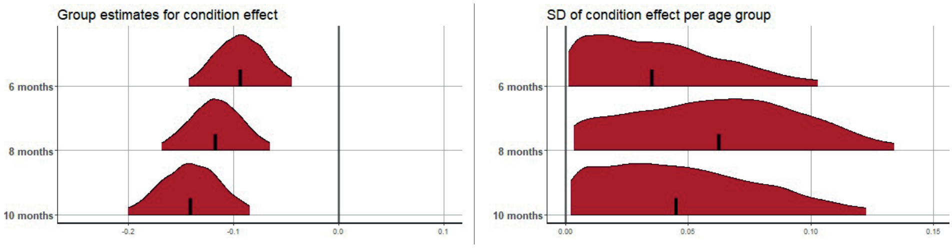
*Results of the hierarchical model for each individual per age group*




*Note.* The black dots represent the median; the red bars represent the 95% credibility intervals.

Figure 7

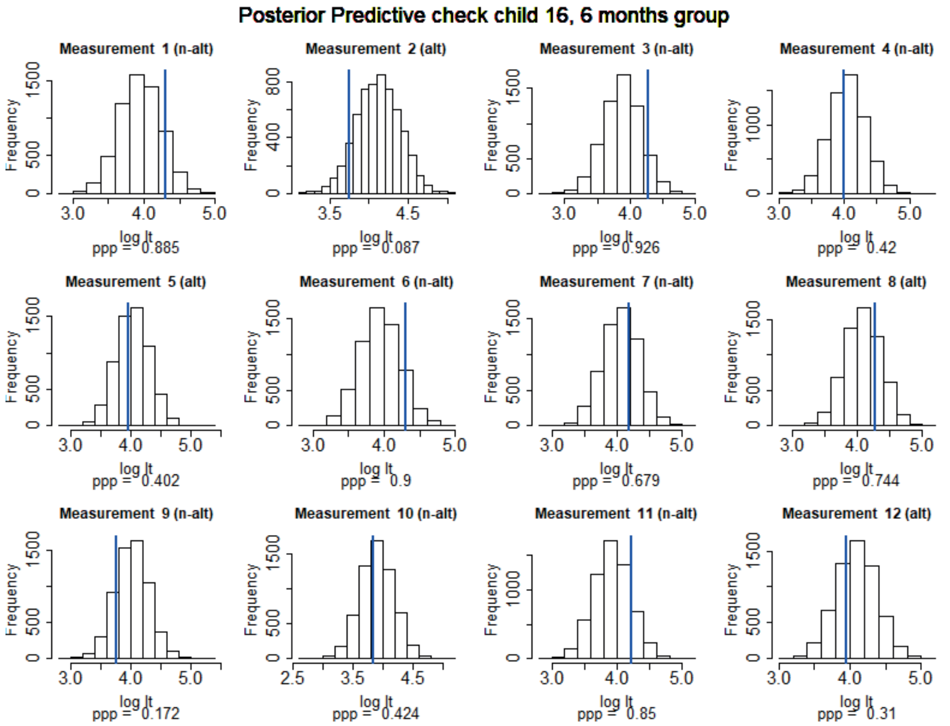
*Group estimates for condition effects and variation per age group*




*Note.* The left panel shows the group estimates for condition effects. The right panel shows the standard deviation of the condition effect per age group. The densities, presented in red, represent the 95% credibility interval.

Figure 8

*Posterior predictive simulations for child 16 in age group 6 for all 12 observed trials*



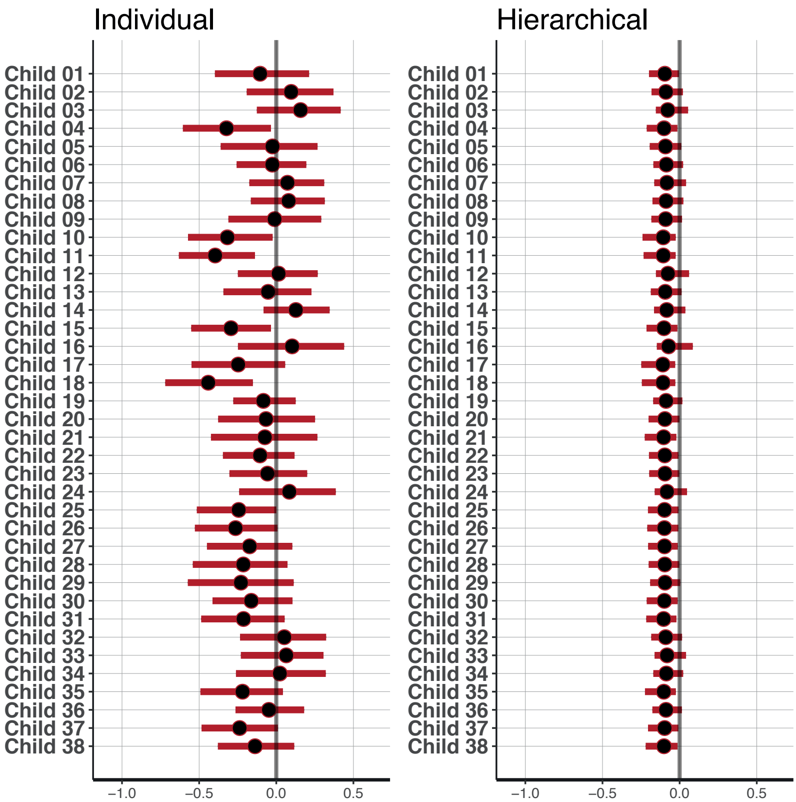

*Note.* Each histogram contains 6000 simulated values for that particular observation of that specific child based on the posterior parameter estimates. The blue vertical line denotes the actually observed value for the specific measurement.

As part of the model assessment we conducted posterior predictive checks. These checks provide insight into the plausibility to the hypothesized and estimated model by drawing simulations from the posterior model. Figure 8 shows how well the model fits the data of a particular child, in this case child 16 in age group 6. Simulations are based on the posterior parameter estimates for this specific child at each specific measurement, taking into account the child-specific estimated looking times for (non-)alternating trials, the child-specific condition effect and the child-specific autoregressive effect. The posterior predictive p-value (ppp) indicates the proportion of simulated values for this measurement that are smaller than the observed value. If ‘ppp’ falls between 0.025 and 0.975 we conclude that

our model provides an accurate prediction for this specific observation. Note that this specific child 16 is classified as non-discriminator and that all measurements are accurately captured by the model as shown by the blue bars in each histogram (Figure 8). For an example of a child classified as non-discriminator with less accurate model descriptions for the observed measurement see for instance child 17 from age group 10, measurements (trials) 5 and 7 (see https://osf.io/ebrxy/).
To evaluate the effects of the hierarchical regression compared to modelling the individual regressions, we also ran Bayesian regression analyses with AR1 error structure without the multilevel structure. Figure 9 shows the estimates with their uncertainty for the condition parameter for all infants in age group 6 (only); the other groups show similar patterns. The figure shows that including the hierarchical structure reduced the uncertainty of the estimates markedly.

Figure 9

*Comparison of results of individual and hierarchical analyses for condition parameter of each infant in the 6-month-olds group*



*Note.* The Hierarchical model reduces the uncertainty (95% CI represented by red bar, median represented by the black dot) for the parameter estimates.

Table 4 displays the mean log-transformed looking time differences between the alternating and non-alternating trials for all individuals that did not include the value 0 in their 95% CI for the condition effect in the hierarchical regression. These raw data show the direction of the average difference in looking time between alternating and non-alternating trials, as well as the magnitude of the average difference between trial types. As can be seen, both looking time difference directions are present, meaning that the data set includes infants with on average longer looks to alternating trials as well as infants with on average longer looks to non-alternating trials. In addition, Table 4 shows that the magnitude of looking time differences between alternating and non-alternating trials shows considerable variation.

Table 4

*The mean looking time difference between alternating and non-alternating trials for the infants whose confidence interval (95%) did not cross the value 0*

| Subject | Group | Difference | Subject | Group | Difference |
|---|---|---|---|---|---|
| | Age | alternating - nonalternating | | Age | alternating - nonalternating |
| child 02 | 6 | -.10 | child 41 | 8 | -.27 |
| child 03 | 6 | -.14 | child 44 | 8 | -.07 |
| child 05 | 6 | -.03 | child 01 | 10 | .12 |
| child 06 | 6 | .03 | child 02 | 10 | .13 |
| child 07 | 6 | -.10 | child 03 | 10 | -.09 |
| child 08 | 6 | -.01 | child 04 | 10 | .25 |
| child 09 | 6 | -.02 | child 05 | 10 | .11 |
| child 12 | 6 | -.05 | child 06 | 10 | .09 |
| child 13 | 6 | .02 | child 07 | 10 | .15 |
| child 14 | 6 | -.16 | child 08 | 10 | .26 |
| child 16 | 6 | -.13 | child 09 | 10 | .16 |
| child 19 | 6 | -.03 | child 10 | 10 | .11 |
| child 20 | 6 | .19 | child 12 | 10 | .20 |

| | | | | | |
|---|---|---|---|---|---|
| child 23 | 6 | .03 | child 13 | 10 | -.02 |
| child 24 | 6 | .09 | child 14 | 10 | .15 |
| child 29 | 6 | .11 | child 15 | 10 | -.12 |
| child 32 | 6 | .06 | child 16 | 10 | -.15 |
| child 33 | 6 | -.07 | child 17 | 10 | -.04 |
| child 34 | 6 | -.04 | child 18 | 10 | .14 |
| child 36 | 6 | .08 | child 19 | 10 | -.23 |
| child 01 | 8 | -.24 | child 21 | 10 | .03 |
| child 08 | 8 | -.36 | child 23 | 10 | .14 |
| child 13 | 8 | -.15 | child 26 | 10 | -.02 |
| child 17 | 8 | .19 | child 27 | 10 | .12 |
| child 20 | 8 | .48 | child 28 | 10 | .22 |
| child 21 | 8 | .04 | child 29 | 10 | -.10 |
| child 30 | 8 | -.10 | child 32 | 10 | .16 |
| child 32 | 8 | -.12 | child 34 | 10 | .36 |
| child 34 | 8 | -.11 | child 35 | 10 | .02 |
| child 37 | 8 | -.08 | | | |

*Note.* The mean log-transformed looking time differences are presented

**3.4 Discussion**

The primary aim of this study was to determine if speech discrimination performance can be reliably assessed for individual infants in a habituation design. This is crucial for understanding individual developmental trajectories and in addressing potential clinical questions. In order to do so we used the experimental design -- hybrid visual fixation (HVF) -- and statistical approach -- linear regression modeling with autoregressive error structure -- reported in Houston et al. (2007). Houston et al. found that 80% (8/10) of their 9-month-old participants discriminated the *boodup* - *seepug* contrast. Our study assessed individual native phoneme (/faːp - /feːp/) discrimination in Dutch infants aged 6, 8 and 10 months, using a slightly altered version of the HVF paradigm. When conducting the regression analysis that Houston et al. (2007) applied, we found that only 12% (14/117) of the infants discriminated the contrast. We were thus not able to replicate Houston et al.'s (2007) findings, using the same model as they did.

Houston et al. did not correct for multiple testing, but when such a correction is applied (as we did), it did not make a difference for the findings of the Houston et al. sample. For our study, however, the correction led to a reduction of the percentage of infants in whom discrimination could be attested to 3% (from 12%). Bayesian hierarchical modeling provides both group and individual estimates using the same model and therefore has the advantage that it does not require correction for multiple testing. Using a hierarchical model with both the autoregressive effect (looking time decreases during test) and the inclusion of group information led to reduced uncertainty of the estimates of the condition effects (alternating versus non-alternating) at both the group and the individual level. The analysis returned a higher percentage (50%) of infants that showed evidence of discrimination. Evidence of discrimination is defined as the 95% credibility interval that does not include value 0 for the condition effect. For the 10-months-olds we found that 77% discriminated between *faap* and *feep*, while 53% of the 6-month-olds and only 27% of the 8-month-olds did. These individual discrimination outcomes are still lower than expected. We expected that most infants would show evidence of discrimination, regardless of age and we predicted discrimination percentages comparable to those obtained by Houston et al. (2007). Seventy-seven percent of the 10-months-old infants discriminated the contrast. This is comparable to findings of 9-month-olds in the study of Houston et al. (2007). It is conceivable that the design (8 alternating and 4 non-alternating test trials) is more suitable for the older than for the younger infants.

Two design differences between the study by Houston et al. (2007) and ours could also account for the diverging results. First, Houston et al. used a word contrast, *boodup - seepug*, which differs markedly from the phonemic contrast /faːp - feːp/ we used. The more conspicuous word contrast may have elicited a larger difference between alternating and non-alternating trials. Second, Houston et al. used 14 test trials, two more non-alternating trials than we did. This might have caused a lower mean looking time to non-alternating trials, as infants' internal representation of the old (non-alternating) stimulus might become stronger during test, which is expected to result in a larger increase in looking time to new stimuli (Sokolov, 1963). Still, infants of all age groups showed evidence of discrimination (Chapter 2) and Figure 7 of this paper) and this does not seem to align with the lower percentage of infants significantly discriminating the contrast we observed in the current study. However, age-related enhancement of discrimination is shown by an increasing percentage of infants discriminating the contrast, which fits the theory of perceptual attunement (Tsuji & Cristia, 2014; Maurer & Werker, 2014).

Our individual analyses are an exploratory extension of the individual analyses done by Houston et al. (2007); we used Bayesian hierarchical modelling to assess if an infant can discriminate the two stimuli. The theoretical advantages of our approach have been discussed throughout the paper. The approach by Houston et al. (2007) and our approach lead to different conclusions for many infants in our study. Strictly speaking, our decision rule, i.e., discrimination is attested if the 95% CI does not include 0, is not an entirely proper method for hypothesis testing. Some shortcomings of forcing decision rules on parameter estimates are discussed in Lee (2018), where Bayes Factors are advocated. However, the application of Bayes Factors in the current setting would present serious challenges and there are arguments against them in general (Gelman et al., 2013). On the other hand, our approach is not unprecedented; Kruschke (2013), for example, used a similar approach as an alternative to t tests, and Gelman and Tuerlinckx (2000) show that this approach reduces the chance of Type S (sign) errors in comparison to the classical framework. The decision rule we used could be used to infer discrimination.

The Bayesian hierarchical model presents a more reliable statistical approach: If measurements contain (substantial) noise, this negatively affects the reliability of a measurement. That is, if we measure the same construct multiple times we obtain different results. If we are able to reduce the noise, our measurement becomes less variable and will measure the same construct in a more stable manner over multiple times. By including hierarchical structures in our model we can capture part of the noise in our estimated looking times (see Figure 9). The reduction of the noise leads to less variable representations of the measurements which can be seen as an improvement of the reliability of the measurements (Gelman et al., 2012).

The current study aimed at assessing individual outcomes because looking time data is noisy and often challenging to interpret (Aslin & Fiser, 2005; Oakes, 2010). Nevertheless, studies do attempt to interpret these individual variations by, for instance, examining follow-up data and in retrospect analyze the infant looking time data (e.g., Newman et al., 2006), which at group level give some insight in the relations between early perception skills and later language development (Cristia et al., 2014). However, raw looking time data cannot be used to infer success or failure. In order to classify individuals as discriminators, data should be modelled and advanced statistical methods need to be applied. The method presented in this study allows us to classify individual infants as discriminators or non-discriminators. Moreover, the procedure allows us to investigate how well our model performs for each trial for each individual child using posterior predictive checks, an example can be seen in Figure 8. However, more research needs to be done to investigate replicability of the current study. Factors that will influence outcomes are, for example, sample size, as estimates will be more accurate with increased sample size, and the total number of data points per subject. Future research should also focus on the question whether classification as presented in this study is indeed of clinical value: do infants classified as discriminators have better language performance measured at a later age?

Taken together, assessing individual discrimination performance with an autoregressive model per individual without correcting for multiple testing is not an approach to be favored. On the other hand, if multiple testing is corrected for, significant results rely on sample size, because with each infant that is added another test should be run. Sample size influences the corrected alpha-level, which is arbitrary. A model in which all these issues can be tackled is the Bayesian Hierarchical model: we can account for a decrease in looking time (autoregressive effect); it includes group information in the hierarchical model; it does not require correction for multiple testing, and it provides more confidence in classifying infants as being able to discriminate a stimulus contrast or not. Our findings thus provide a step forward in assessing infants' speech discrimination.

**Ethics statement**

Informed consent was obtained from the caregiver before testing and the caregiver was allowed to retract this consent and participation any time during testing. The authors declare that the research was conducted in accordance with APA ethical standards as well as *The Netherlands Code of Conduct for Scientific Practice* issued in 2004 (revised in 2018) by the Association of Universities in the Netherlands (VSNU).

**Acknowledgements**

We are grateful to the infants and their caregivers for participating. We would like to thank the student assistants Sule Kurtçebe, Tinka Versteegh, Lorijn Zaadnoordijk and Joleen Zuidema, who helped collecting data. We would like to thank Derek Houston for sharing some of his raw data with us (see Appendix A). This research was funded by The Netherlands Organization for Scientific Research (NWO). Grants nr. 360-70-270, awarded to. F.N.K. Wijnen and nr. VIDI-452-14-006, awarded to R. van de Schoot).

Chapter 4

# Speech discrimination in infants at family risk of dyslexia: Group and individual-based analyses

**Abstract**
Deficiencies in discriminating and identifying speech sounds have been widely attested in individuals with dyslexia as well as in young children at family-risk (FR) of dyslexia. A speech perception deficit has been hypothesized to be causally related to reading and spelling difficulties. So far, however, early speech perception of FR infants has not been assessed at different ages within a single experimental design. Furthermore, a combination of group- and individual-based analyses has not been made. In this cross-sectional study, vowel discrimination of 6-8-10-month old Dutch FR infants and their non-risk peers (no-FR) was assessed. Infants ($N$ = 196) were tested on a native /aː/-/eː/ and a non-native English /ɛ/-/æ/ contrast, using a hybrid visual habituation paradigm. Frequentist analyses were used to interpret group differences. Bayesian hierarchical modeling was used to classify individuals as speech sound discriminators. FR and no-FR infants discriminated the *native* contrast at all ages. However, individual classification of the no-FR infants suggests improved discrimination with age, but not for the FR infants. No-FR infants discriminated the *non-native* contrast at 6 and 10, but not at 8 months. The FR infants did not show evidence of discriminating the contrast at any of the ages: 0% were classified as discriminators. The group and individual-based data are complementary and together point towards speech perception differences between the groups. The findings also indicate that conducting individual analyses on hybrid visual habituation outcomes is possible. These outcomes form a fruitful avenue for gaining more understanding of development, group differences, as well as prospective relationships.

Published as de Klerk, M., de Bree, E., Veen, D., & Wijnen, F. (2021). Speech discrimination in infants at family risk of dyslexia: Group and individual-based analyses. *Journal of Experimental Child Psychology*. https://doi.org/10.1016/j.jecp.2020.105066

EdB, MdK and FW designed the study. MdK conducted the group level analyses. DV conducted the individual level analyses. MdK wrote and revised the paper, with input and feedback from EdB, DV and FW.

## 4.1 Introduction

Developmental dyslexia is a language-based learning disability characterized by severe word reading and/or spelling problems (Lyon et al., 2003; Peterson & Pennington 2015). These literacy difficulties can have a profound impact on educational/academic achievement, self-esteem, and social development (Livingstone et al., 2018). Therefore, it is of great interest to understand the precursors or risk factors that lead to the subsequent deficit. Dyslexia is considered to be a multifactorial disorder, which implies that multiple risk- and protective factors are involved. The disorder is considered to be highly heritable: Children with a dyslexic parent have a 29- 66% risk of developing dyslexia (Snowling & Melby-Lervåg, 2016). Assessing the abilities of children with a family risk (FR) of dyslexia, therefore, is a valuable approach for finding early markers of dyslexia (e.g., Caglar-Ryeng et al., 2019; Snowling & Melby-Lervåg, 2016; van Viersen et al., 2018).

A phonological deficit has been proposed to be one of the main contributing risk factors in developing dyslexia (Ramus, et al., 2003; Vellutino, Fletcher, Snowling & Scanlon, 2004). It is hypothesized that people with dyslexia have poorly specified phonological representations, which, in turn, have a disruptive effect on the construction of phoneme-grapheme connections (e.g., Blomert, 2011; Mittag et al., 2013). Although the phonological deficit cannot account for all literacy problems in people with dyslexia (Pennington et al., 2012), there is extensive evidence for phonological problems in children and adults diagnosed with dyslexia (Ramus et al., 2003; van Bergen et al., 2012). It is also seen in FR children prior to the acquisition of literacy skills (see Snowling & Melby-Lervåg, 2016 for a meta-analysis and review), which is suggestive of a causal relation.

One potential cause of the phonological deficit is poor speech perception: if speech sounds cannot be perceived and categorized adequately, this will hamper the formation of phonological representations as well as grapheme-phoneme associations (e.g., Goswami, 2000). A large number of studies found that adults and children with dyslexia perform more poorly on tasks measuring speech perception skills than their peers (e.g., Schulte-Körne et al., 2001; Werker & Tees, 1987, but see Nittrouer et al., 2011; Ramus et al., 2003; Rosen & Manganari, 2001). This is also found for children and infants with an FR (e.g., Boets et al., 2007; Guttorm, et al., 2005; Richardson et al., 2003; van Alphen, et al., 2004). Moreover, some studies have found that children with lower (pre)reading skills showed poorer speech perception performance as infants (e.g., Guttorm et al., 2010; Molfese, 2000; van Zuijen et al., 2013). Hence, speech perception skills seem to be related to learning to read and spell effectively and efficiently.

Adults and children with dyslexia have been found to perform more poorly than their peers on speech sound categorization tasks (Hakvoort et al., 2016; Maassen et al., 2001; Mody et al., 1997), which has led to the proposal of a categorical speech perception deficit (e.g., Serniclaes et al., 2004). One explanation for the categorization deficit could be that ‘children with dyslexia maintain the sensitivity to phonemic distinctions which all newborns have irrelevant of their native language’ (Noordenbos et al., 2012, p. 1470). Although the reported results are not fully consistent (Blomert & Mitterer, 2004; Brandt & Rosen, 1980; Hazan et al., 2009; Messaoud-Galusi et al., 2011) a recent meta-analysis does show support for a categorical perception deficit in dyslexia (Noordenbos & Serniclaes, 2015). Poor categorical perception has also been found in kindergartners with an FR of dyslexia (Boets et al., 2007; Gerrits & de Bree, 2009; Noordenbos, et al., 2012).

Phonological categorization builds on a robust speech sound discrimination ability. The available evidence suggests that speech sound discrimination in FR infants is weaker in comparison to low risk (no-FR) peers (van Leeuwen, et al., 2006; Richardson et al., 2003, and Volkmer & Schulte-Körne, 2018 for a recent review on EEG studies). Van Leeuwen and colleagues (2006) conducted an EEG study with 2-month-old Dutch infants using an oddball paradigm in which /b/ and /d/ were presented in Dutch /bak/ (*box*) - /dak/ (*roof*) words. The tokens used were taken from a /b/-/d/ continuum. The FR infants showed a significantly less pronounced mismatch negativity response to the deviant stimulus, indicative of a delay in categorization. Such poorer phoneme discrimination in FR infants has been found for consonants as well as vowels (e.g., Guttorm et al., 2001; van Leeuwen, et al., 2006; Leppänen et al., 1999; Molfese, 2000; Pihko, et al., 1999; Thiede, et al., 2019). In sum, the literature shows that FR infants have more difficulty with discrimination between phonemes. This finding can be related to subsequent poor categorization and aligns with the notion of a speech perception deficit in dyslexia.

The studies that report discrimination difficulties in FR infants have thus far been limited in the sense that the speech sound contrasts under investigation were all native contrasts and were mostly assessed at one age. Because it is well established that speech perception changes in the first year of life due to language exposure, it is warranted to investigate how (native) speech perception develops in FR infants. In typically developing infants, speech perception changes from universal to language-specific (e.g., Werker & Tees, 1984). This means that the ability to discriminate native speech sound categories remains good or improves (for sounds that are initially difficult to discriminate), whereas the ability to detect speech sound distinctions that are not phonemic in the native language decreases (e.g., Tsuji & Cristia, 2014; Werker & Tees, 1984). This developmental transition is generally referred to as *perceptual attunement* (Maurer & Werker, 2014) and emerges around the age of 10-12 months for consonantal contrasts and at the age of 6-8 months for vowel contrasts (Kuhl et al., 1992; Polka & Werker, 1994). Perceptual attunement is the first step into the formation of (native) phoneme categories.

Although investigation of the developmental trajectory of native and non-native speech perception of FR infants is important for evaluating the process of perceptual attunement in FR, we know of no studies that have looked into this. There are, in contrast, some studies with no-FR infants. In one such study (Chapter 2), it was found that no-FR infants were able to discriminate between the salient native vowel contrast /a:/ and /e:/ at 6, 8 and 10 months of age and that this discrimination improved with age. In contrast, only the 6 and 10-month-olds were able to discriminate between non-native English /ɛ/ and /æ/; the 8-month-olds were not. These findings are indicative of perceptual attunement between the ages of 6 and 8 months. The finding that the 10-month-olds could discriminate the contrast was explained by an interaction between task demands and maturation.

**Current study**

The current study compares speech sound discrimination of 6-8-and-10-month-old FR infants to that of their no-FR peers. We used the hybrid visual habituation paradigm (Chapter 2; Houston et al, 2007), comprising test trials with similar phonemes (non-alternating, e.g., /a:/-/a:/) and different phonemes (alternating, /a:/-/e:/). We addressed two questions. The first is whether perceptual attunement occurs in FR infants. In other words, is there a change from a universal listener to a language-specific listener, also in FR infants? If perceptual attunement takes place, native contrasts are expected to be discriminated at all ages, whereas there is a decrease in the ability to discriminate non-native contrasts as infants mature (e.g., Kuhl, et al., 2008; Tsuji & Cristia, 2014). Hence, FR infants should be able to discriminate a salient (acoustically and articulatory highly distinctive) native contrast, such as Dutch /a:/ - /e:/. However, studies on speech sound discrimination skills of FR infants have often used non-salient native contrasts and found perception difficulties on these subtle contrasts (e.g., van Leeuwen et al., 2006). It is therefore possible that initial discrimination is weak(er) and the gradual improvement of discrimination proceeds more slowly than in no-FR infants. In other words, there could be a delay in perceptual attunement for FR infants, hence, investigation of a native salient contrast is warranted.

With respect to the discrimination of non-native contrasts, the question is whether FR infants' data will provide evidence of discrimination at 6 months of age, and a decrease in sensitivity at later ages to the English /ɛ/ - /æ/ contrast that we use in this study. This contrast is difficult to distinguish for Dutch adults (Broersma & Cutler, 2011); both vowels are perceived as Dutch vowel /ɛ/. On the basis of previous findings of poor speech sound discrimination of FR infants (e.g van Leeuwen et al, 2006; Leppännen et al., 2002; Richardson, et al., 2003), it is conceivable that the subtle non-native contrast will not be discriminated throughout development (e.g., van Leeuwen et al., 2006; Richardson, et al., 2003). Another possibility is that infants will not lose the sensitivity to the irrelevant non-native contrast (Noordenbos, et al., 2012).

Our second main question is whether it is possible to identify individual infants as being able to discriminate the speech sound contrast or not. Put differently, can infants be classified as 'discriminators' at the individual level on the basis of outcomes on behavioral speech perception tasks? This is an important question in the field of speech perception (Cristia et al., 2014; Houston, et al., 2007): Group-based findings are valuable for understanding a general pattern of discrimination, but identification of individual difficulties and future outcomes requires reliable analyses on individual-based data. Furthermore, studies have investigated the relation between early speech perception and later reading skills retrospectively (Guttorm, et al., 2010; van Zuijen, et al., 2013). If discriminators at infancy can be identified successfully, this would facilitate prospective studies into early speech perception and later language and reading skills.

To address this question, we take our previous study (Chapter 3) as starting point. We evaluated different methods of individual analyses and found that Bayesian hierarchical modeling was the most successful. This approach takes into account the hierarchical nature of the data: infants within the same age group are assumed to belong to the same population, meaning that infants are exchangeable *within* age groups but not *between* age groups. The advantage of Bayesian hierarchical modeling in comparison to frequentist approaches is that it yields estimates for all the individual and group parameters in one model without having to correct for multiple testing (Gelman et al., 2012). Furthermore, the consequence of hierarchically modeling the individual and group effects in one analysis is that part of the observed variance can be explained at the group level instead of trying to explain all the variance at the individual level (Gelman, 2006). In our previous study (Chapter 3) we showed that by adding the hierarchical structure we reduced the noise, which led to less variable representations of the measurements. This can be seen as an improvement of the reliability of the measurements (Gelman et al., 2012). Individual outcomes can provide more insight in developmental trajectories and are thus of great value for studies that relate early abilities to later language skills.

## 4.2 Method

### Participants

Participants were recruited via a letter sent to all the parents of newborns of (reference to city). Addresses were supplied by the municipality of Utrecht City. Before coming to the lab, parents were asked to fill out a questionnaire, consisting of questions about birth weight, gestational age, health issues and (medical) family background. Infants were included if: (a) they were raised only in Dutch; (b) their gestational age at birth was between 37 and 43 weeks; (c) their birth weight was between 2500-5000 grams; (d) there were no complications during the pregnancy or delivery; (e) they did not have a history of known hearing loss or reduced vision and (f) they did not have reported neurological problems.

In order to ascertain whether the FR infants could truly be categorized as such, three tests were administered to the parent who had indicated a history of reading problems. The first was a timed word reading test, the 'Een-minuut-test' (EMT; Brus & Voeten, 1972). In this test parents had to read out loud a list of known words as quickly and accurately as possible within one minute. The second test was a timed pseudoword reading test. Parents were asked to read out loud a list of pseudowords within two minutes ('de Klepel'; van den Bos et al., 1994). The third test, a verbal competence test (Analogies), was a subtest of the Dutch version of the Wechsler Adult Intelligence Scale (WAIS; Uterwijk, 2000). Infants were included in the family-risk group if parents had met one of the following criteria: 1) the percentile scores on one reading tests was ≤ 10, or 2) the percentile score was ≤ 20 on both of the reading tests, or 3) when the discrepancy between the one of the reading tests and the verbal competence test was 60 percentile points or more (Kuijpers et al., 2003). If the criteria were not met, the infant was not included in the study ($n$ = 7).

A flowchart of the data inclusion process can be found in Figure 1. In total, 117 FR infants were tested, potentially on both the native and non-native vowel discrimination experiment, rendering a potential of 234 datasets or records (2*117 records). However, 71 (30%) records were not included for the following reasons: 1) behavior invalidating the measurements (crying, extreme restlessness, $n$ = 29); 2) the second discrimination experiment was never started (the decision to proceed to the next experiment depended on behavior and well-being of the infant after the first experiment, $n$ = 23); 3) the parent was not classified as dyslexic ($n$ = 14 records, see above); 4) failure to meet the habituation criterion ($n$ = 11; see *Procedure*); or 5) a technical error ($n$ = 1). Figure 1 contains a capture of the inclusion criteria in a flow chart. In total 163 records were included. These 163 records came from 98 infants. Sixty-five of these infants finished both the native and non-native condition, hence, these 65 infants yielded 130 records (native $n$

= 65 and non-native *n* = 65). Some infants (*n* = 33) finished only one contrast (native, *n* = 14 or non-native, *n* = 19), see Table 1 (FR) and Appendix B and C.

The no-FR infants were selected from the data set presented in Chapter 2. The no-FR infants (*n* = 98) were matched to the FR infants on the following characteristics: 1) age; 2) the number of experiments that they had finished during the session (1 or 2); 3) the stimulus they were habituated on and 4) which contrast was presented first (native or non-native). In the no-FR selection too, *n* = 65 infants completed both experiments and *n* = 33 infants finished the task in one condition; see Table 2 and the Appendices B (native contrast) and C (non-native contrast) for more information regarding the number of infants per age groups that finished both contrasts.

All parents were native monolingual speakers of Dutch and lived in Utrecht City. Data on parental level of education and the family situation, i.e., the number of siblings and birth order (first born, second born, etc.) are summarized in Table 3. The educational level was coded ranging from 1 (primary school) to 6 (university level). The average educational level of FR infants' fathers was significantly lower than that of no-FR fathers, but for both groups, the educational level was high, see Table 3. The majority of FR (88%) and no-FR (95%) fathers had completed a university degree (bachelor or master level).

Figure 1

*Flow chart for data inclusion*

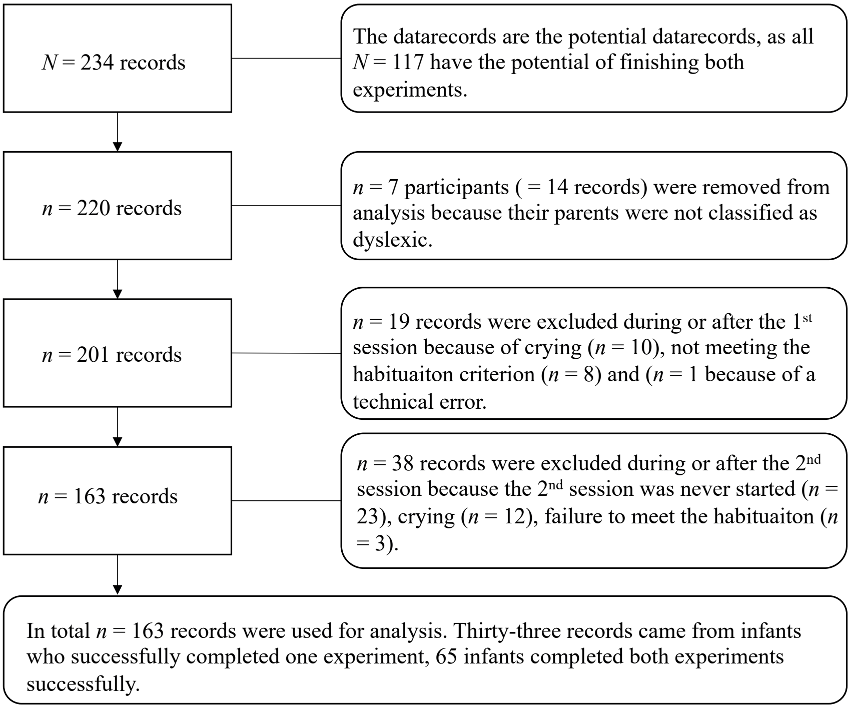


Table 1

*Numbers of FR participants and mean ages per age group, for the native and non-native contrast*

| Age group | Age | FR-infants tested | FR-total infants included | Of which both contrasts | Only native contrast | Only non-native contrast | Total included native contrast | Total included non-native contrast |
|---|---|---|---|---|---|---|---|---|
| | *FR infants* | *N* | *n (female)* | *n (female)* | *n (female)* | *n (female)* | *n (female)* | *n (female)* |
| 6 | (in days) | 38 | 33 (14) | 26 (11) | 2 (1) | 5 (2) | 28 (12) | 31 (13) |
| 8 | 194 (8.7) | 38 | 32 (15) | 19 (6) | 5 (3) | 8 (6) | 24 (9) | 27 (12) |
| 10 | 258 (7.8) | 41 | 33 (19) | 20 (10) | 7 (7) | 6 (2) | 27 (17) | 26 (12) |
| Total | 316 (8.2) | 117 | 98 (48) | 65 (27) | 14 (11) | 19 (10) | 79 (38) | 84 (37) |

Table 2

*Numbers of no-FR participants and mean ages per age group, for the native and non-native contrast*

| Age group | Age | Tested on native contrast only | Tested on non-native contrast only | Tested on both contrasts | Total included native contrast | Total included non-native contrast |
|---|---|---|---|---|---|---|
| | (in days) | *n (female)* | *n (female)* | *n (female)* | *n (female)* | *n (female)* |
| 6 | 203 (8.4) | 2 (0) | 5 (2) | 26 (10) | 28 (10) | 31 (12) |
| 8 | 259 (6.5) | 5 (3) | 8 (6) | 19 (11) | 24 (14) | 27 (17) |
| 10 | 320 (12.9) | 7 (4) | 6 (2) | 20 (10) | 27 (14) | 26 (12) |
| Total | | 14 (7) | 19 (10) | 65 (31) | 79 (38) | 84 (41) |

*Note.* These no-FR infants are a subsample of the sample presented in Chapter 2.

Table 3
*Background information of participants*

| Measure | no-FR | FR | |
|---|---|---|---|
| | *M (SD)* | *M (SD)* | *Mann-Whitney test (two-sided)* |
| Educational level father | 5.58 (.5) | 5.13 (.9) | $U$ = 3,040.00, $z$ = -3.44, $p$ = .001 |
| Educational level mother | 5.66 (.5) | 5.48 (.7) | $U$ = 3,647.00, $z$ = -1.77, $p$ = .077 |
| Nr. of siblings | .33 (.5) | .28 (.5) | $U$ = 4,000.50, $z$ = -.66, $p$ = .509 |
| Birth Rank | 1.31 (.5) | 1.27 (.5) | $U$ = 4,028.00, $z$ = -.57, $p$ = .570 |

*Note.* Educational level was measured on a scale from 1 (primary school) to 6 (university/PhD).

## Procedure and stimuli

### General procedure

Participants were tested in a three-walled canvas test booth placed in a sound-attenuated room. The infant was seated on the parent's lap, approximately 1.35 meter from the 17-inch computer screen (Philips LCD 150P4). The loudspeaker (Tannoy i8) through which the auditory stimuli were played was hidden behind the canvas of the booth and placed underneath the TV-screen that showed the visual stimuli. Parents wore headphones (Echelon Telex), through which music was played in order to prevent them from hearing the stimuli and (potentially) influencing their child's behavior. The experiment was monitored and recorded through a video camera that was placed underneath the TV screen. Looking time was tracked by pressing a button box for looking and looking away. Looking time was taken to reflect listening time (Aslin, 2007). The button-box was connected to a (Asus P4PE) computer. An experiment control application (*Zep*; Veenker, 2008) was used for presentations of the auditory and visual stimuli and for the data registration. Trials were initiated with a button press and were ended when either the infant looked away for two seconds or when maximum trial length was reached.

Prior to testing, parents provided written consent for participation and the experimental procedure was explained to the caregiver without telling them which of the conditions (native or non-native) was presented first, see (Chapter 2) for further instructions to the caregiver. The aim was to test the infants on both contrasts (native and non-native) within one session and the order was counterbalanced between infants.

Similar to the study of Houston and colleagues (2007), the experimental set up consisted of a habituation phase, in which infants were habituated to one of the vowels of the pair (e.g., /aː/ in / faːp/), a test phase, in which looking times to non-alternating (habituation) vowel pairs (e.g., /faːp/ - /faːp/) were compared to alternating vowel pairs, i.e. a pair consisting of a trained vowel and a contrasting, untrained, vowel (e.g., /faːp/-/feːp/). The experiment begins and ends with a pre- and posttest to measure participants' attentiveness. Each of these phases included both auditory as well as visual stimuli. During habituation we used tokens from four different female speakers. Speaker variability has been argued to enhance generalization of abstract features in the process of developing phonetic categories (e.g., Lively et al., 1993; Rost & McMurray, 2009).

**Stimuli**

***Visual and auditory stimuli pre-and posttest.*** During the pre-and posttest infants were presented with both auditory (beep sounds, 330 Hz, duration 250ms, ISI 1000ms) and visual stimuli. Auditory stimuli were played at ~65 dB. The visual stimuli were three cartoon pictures displayed for two seconds on a light blue screen. The three pictures were drawn randomly out of a set of 25 pictures. These pictures could appear in nine different spots within an invisible 3 x 3 grid, see top left picture in Figure 2. After two seconds, a series of three new pictures appeared at different locations. Pictures were presented in pseudorandomized order.

***Visual and auditory stimuli habituation and test.*** During the habituation and test phase pictures of six smiling female faces were used (see an example in Figure 2, top right picture). In each block of four trials four pictures with different female faces were used, one picture per trial. Pictures were presented in pseudorandomized order. Between habituation trials a visual attention getter was displayed: a movie of a cute laughing baby (see Figure 2, bottom left picture). In between test trials a movie of a toddler going down a slide was used as an attention getter (see a Figure 2, bottom right picture).

Figure 2

*Visual stimuli presented during the pre- and posttest, habituation and test phase*


Auditory stimuli were the Dutch vowels /a:/ and /e:/ for the native contrast and the English /ɛ/ and /æ/ for the non-native contrast. Vowels were embedded in CVC syllables: /fa:p/, /fe:p/, /sæn/ and /sɛn/. Recordings of the Dutch pseudowords /fa:p/ and /fe:p/ were made of four female Dutch speakers, aged between 25 and 35 years. They all spoke Standard Dutch and came from the *Randstad area*, a mostly urban area in central-western Netherlands. They were asked to read out loud a list of 52 words, containing the target pseudowords, as well as monosyllabic Dutch real words with the same vowels (e.g., *gaap* – yawn, *feest* - party). Recordings of the English pseudowords were recorded of four female native English speakers, aged between 25 and 35 years. They came from different regions: South-East London, Belfast, Preston (Lancashire) and Manchester. The pseudowords /sɛn/ and /sæn/ were read out loud from a list of 52 words containing the target words and real words (e.g., *have* and *pet*).

Each speaker produced four tokens of each target pseudoword (e.g., /fa:p/ and /fe:p/). From all four speakers, one token of each target pseudoword per contrast was selected. Additionally, from one speaker a second token per target word was selected, as this was necessary for the test phase (see *Procedure*, below). This resulted in five tokens of four different speakers for both contrasts. Four tokens were used during habituation and the fifth token (token 2 from speaker 1, see also Figure 4) was used in the test phase (see *Procedure*). All auditory stimuli were played at ~65 dB(A). Tokens selected were those that were most child-friendly in prosody and speech affect (see Chapter 2) for more details on acoustic properties). All auditory stimuli were recorded in a sound-attenuated booth of the phonetics lab of the Utrecht University, using a Sennheiser microphone (ME-64) and a digital audio tape recorder (Tascam DA-40).

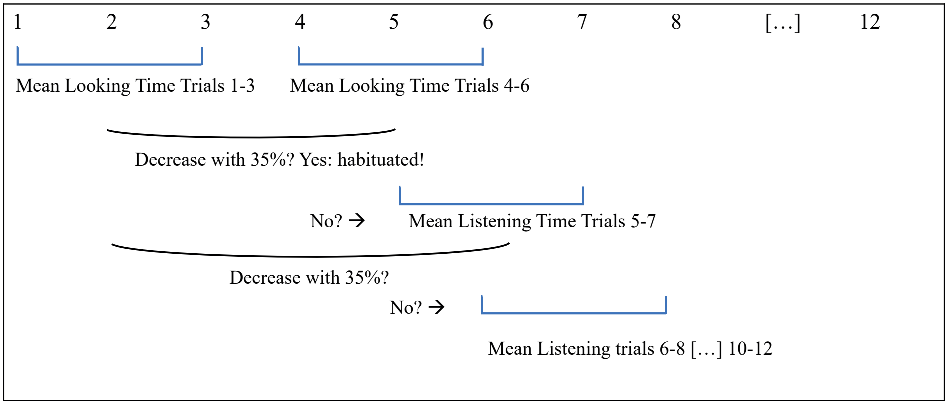
**Procedure**

***Pre- and posttest.*** The pre- and posttest had a fixed duration of ~24 seconds. The purpose of the pre- and posttest was to measure general attentiveness. Infants were excluded when total looking time of the posttest decreased with at least 50% compared to the total looking time of the pretest ($n$ = 1, see *Participants*).

***Habituation phase.*** The habituation phase consisted of a maximum of 12 trials, with a maximum number of 30 tokens (1 second inter-stimulus-interval) per trial, resulting in a total duration of approximately 48 seconds per trial. Participants were habituated on a repetition of one of the stimulus types (e.g., either /faːp/ or /feːp/ in the native condition and either /sɛn or /sæn/ in the non-native condition) with tokens from four female speakers. Within one trial, one token of one speaker was used. In each block of four trials the participant heard all four voices in randomized order within the blocks. Infants were considered to be habituated when they passed the habituation criterion, set at 65%: the mean of trials 1-3 was compared to the mean of trials 4-6. If looking time had not decreased with 35%, the mean of the first three trials was compared to the mean looking time of trials 5-7, then 6-8 up to 10-12, see Figure 3.

Figure 3

*Visual depiction of the assessment of the (65%) habituation criterion*



***Test phase.*** The test phase had a fixed number of 12 trials, with a maximum number of 30 tokens per trial (1 second inter-stimulus-interval), resulting in a maximum total duration of approximately 48 seconds per trial. Test trials consisted either of alternating pseudoword pairs (i.e. native /feːp/ - /faːp/) or non-alternating pairs (i.e. /faːp/ - /faːp/) see Figure 4. The alternating and non-alternating trials were presented in a semi-fixed order: the first trial could be either alternating or non-alternating, which was counterbalanced. The second trial was non-alternating if the first trial was alternating and alternating if the first trial was non-alternating. The three subsequent alternating trials occurred at positions 5, 8 and 12. The other trials were non-alternating. During the test phase a new token of a familiar speaker was introduced. This was done to ensure that the non-alternating trials (e.g., /faːp/ - /faːp/, *faap* - *faap*) had both a new token (*faap*-2 from speaker-1) and a familiar token (*faap*-1 from speaker-1), just like in the alternating trials a new token (*feep*-1 from speaker-1) and a familiar token (*faap*-1 from speaker-1) was used, see Figure 4. The exact same procedure was applied for the non-native contrast.

Figure 4

*Schematic overview of the experimental procedure with reference to the auditory stimuli only*

| Pretest | Habituation Phase | Test Phase | Posttest |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| Beep sounds | Trial 1 /fa:p/ (T1.S1) | Trial 1 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | Beep sounds |
| 330 Hz | Trial 2 /fa:p/ (T1.S3) | **Trial 2 /fe:p/-/fa:p/ (T1.S1 – T1.S1)** | 330 Hz |
| 250 ms | Trial 3 /fa:p/ (T1.S2) | Trial 3 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | 250 ms |
| ISI 1000 ms | Trial 4 /fa:p/ (T1.S4) | Trial 4 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | ISI 1000 ms |
| | Trial 5 /fa:p/ (T1.S3) | **Trial 5 /fe:p/-/fa:p/ (T1.S1 – T1.S1)** | |
| | Trial 6 /fa:p/ (T1.S2) | Trial 6 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 7 /fa:p/ (T1.S4) | Trial 7 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 8 /fa:p/ (T1.S1) | **Trial 8 /fe:p/-/fa:p/ (T1.S1 – T1.S1)** | |
| | Trial 9 /fa:p/ (T1.S1) | Trial 9 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 10 /fa:p/ (T1.S2) | Trial 10 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 11 /fa:p/ (T1.S4) | Trial 11 /fa:p/-/fa:p/ (T2.S1 – T1.S1) | |
| | Trial 12 /fa:p/ (T1.S3) | **Trial 12 /fe:p/-/fa:p/ (T1.S1 – T1.S1)** | |

*Note.* In this example, the contrast is native and the first test trial is non-alternating and consequently the second is alternating. The remaining three alternating trials have a fixed number, viz. the 5th, the 8th and 12th trial. Alternating trials are printed in bold. Token is abbreviated as 'T' and Speakers as 'S'.

**Offline coding**

A random subset (44% of the entire set) of the video recordings was recoded frame-by-frame (frame duration is 30 ms) using Psycode software (http://psy.ck.sissa. it/PsyCode/PsyCode.html), by 2 trained coders who were naive regarding the design and the purpose of the experiment. The results of the raw and recoded data correlated strongly, $r(105) = .99$, $p < .001$. We used the online coding data for analyses.

**Data analysis**

**Frequentist analyses.** In order to assess whether total looking time and number of trials needed to habituate change as a function of age and/or group (FR vs. no-FR), univariate ANOVAs and Kruskal-Wallis non-parametric tests were conducted. Random effects modeling (SPSS, version 23) was used to answer the questions 1) whether there was an effect of trial type (alternating (e.g., /faːp/ - /feːp/) versus non-alternating (e.g., /faːp/ - /faːp/); 2) whether there were differences between the age groups; 3) whether there were differences between the groups (no-FR and FR). The overall fit of the model was tested with a chi-square likelihood ratio test. Seven trials were not included for data analysis; One infant missed the last two trials because the experiment was terminated; five trials were excluded because (three different) infants were suspected of gazing and not looking at the screen. The missing data is not considered problematic, as a) very few trials are missing, and b) parameters can be estimated accurately with missing data using mixed modeling (Field, 2013). For all the frequentist analyses reported in this study, the alpha level was .05.

**Bayesian analysis.** In our Bayesian hierarchical regression model, we modelled the individual infant data in three age groups (six, eight and ten months), per group (FR and no-FR) and contrast (native and non-native), as we did in our previous study (Chapter 3). In that study we presented all details, notably priors, estimation and convergence, and posterior predictive checking, and we conducted a sensitivity analysis (see https://osf.io/xyh3g/). We used a regression model with an AR1 error structure, with Log10 transformed looking times as outcomes, and condition (alternating or non-alternating trial) as predictor. For all groups we obtained both group and individual estimates for the intercept (looking time alternating trials) and condition (difference in looking time between alternating and non-alternating trials).

## 4.3 Results

### Group Analyses: Random effect modeling

### Data screening

***Habituation phase.*** The mean of the total looking times to habituation trials as well as the number of trials required for habituation were assessed across ages. The looking time distributions were positively skewed. Log transformation (Log10) resulted in a distribution that approached a normal distribution (skewness = .026, *SE* = .039, kurtosis = .488, *SE* = .078). The mean number of trials needed to habituate did not approach a normal distribution after log transformation. Therefore, non-parametric tests were conducted on this measure.

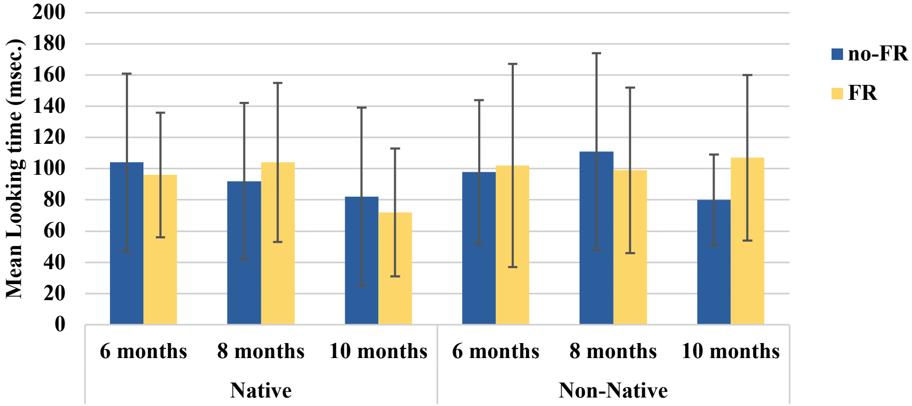
***Test phase.*** The raw looking times to alternating and non-alternating trials were not normally distributed; for this reason, a Log10 transformation was performed; after this transformation the skewness (.096, *SE* = .039) and kurtosis (.256, *SE* = .078) values were acceptable.

### The effect of contrast on discrimination

***Habituation phase.*** Mean looking times required for habituation are reported in Figure 5. Analyses yielded a significant main effect of age $F(2, 313) = 4.51$ $p = .012$. Post hoc analyses showed that the 10-month-olds had overall shorter looking times than the 6- and 8-month-olds. No other main effects and interactions between contrast and age or group were found. The total looking times to habituation trials did not differ between contrasts nor between FR and no-FR infants. The mean number of trials needed to habituate are presented in Table 4. A Kruskal-Wallis test revealed no differences on these mean numbers of trials between no-FR and FR infants in the native contrast, $H(1) = .23$, $p = .637$, nor in the non-native contrast, $H(1) = .24$, $p = .626$. As there were no significant differences between the no-FR and FR infants regarding habituation for both contrasts, we will not discuss habituation separately per contrast.

Figure 5

*Mean looking times (msec.) to habituation trials per contrast and group*




*Note.* The error bars represent SDs.

Table 4

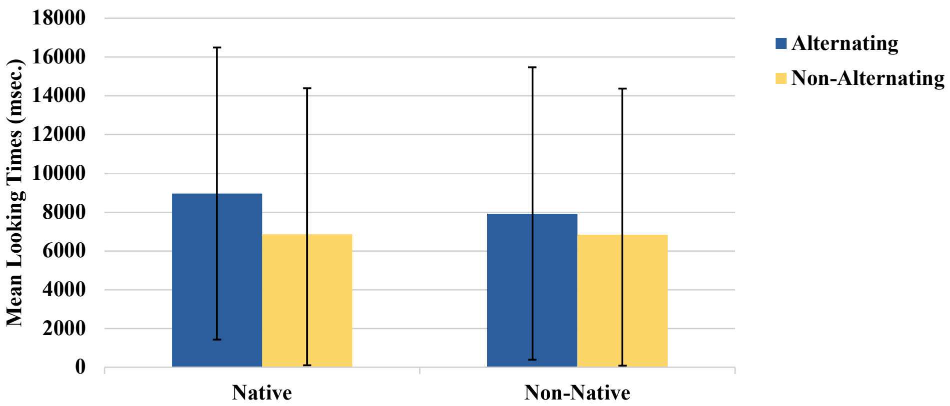
*Mean number of trials needed to habituate per contrast and group*

| Age group | Contrast | Group | |
|---|---|---|---|
| | | No-FR | FR |
| | *Native* | *M (SD)* | *M (SD)* |
| 6 | | 6.6 (1.3) | 6.5 (1.2) |
| 8 | | 6.7 (1.5) | 7.3 (1.8) |
| 10 | | 6.6 (1.1) | 6.8 (1.9) |
| **Total** | | **6.6 (1.3)** | **6.9 (1.6)** |
| | *Non-native* | | |
| 6 | | 6.7 (1.5) | 7.2 (1.8) |
| 8 | | 7.7 (2.0) | 7.1 (1.7) |
| 10 | | 7.1 (1.7) | 7.5 (2.0) |
| **Total** | | **7.1 (1.8)** | **7.3 (1.8)** |

***Test phase.*** We first investigated the effect of contrast (native, non-native), to find out whether trajectories differed between contrasts. Significant interactions of contrast with trial type and/or age would lead us to analyze the results per contrast separately. Looking times per trial type (alternating vs. non-alternating) are presented in Figure 6. A random effect modeling analysis included participant as random factor and trial number as a repeated effect (covariance structure AR1). The fixed factors were trial type (alternating and non-alternating trials), contrast (native and non-native) and age (six, eight and ten months). The model that best fitted the data included the fixed factors trial type (alternating and non-alternating trials), $F(1, 2729) = 88.26$, $p < .001$, contrast (native and non-native), $F(1, 338) = 1.16$, $p = .282$, age $F(1, 338) = 3.94$, $p = .020$, Trial Type*Contrast $F(1, 2729) = 8.24$, $p = .004$ and Trial Type*Contrast*Age $F(6, 906) = 2.78$, $p = .011$. The two-way and three-way interactions show that the effect of trial type on looking time varied across contrasts and ages. Therefore, separate analyses for each contrast are presented in the next sections. The main effect of trial type indicates that the infants looked longer to alternating trials than to non-alternating trials. Looking times decreased with age, as indicated by the main effect of age. No main effect of contrast was found, indicating the overall looking times were not significantly different for the two contrasts.

Figure 6

*Mean looking times to alternating and non-alternating trials for the native and non-native contrast*




*Note.* Error bars represent SDs.

### The native contrast

***Test phase.*** Looking times are reported in Table 5. A random effect modeling analysis included participant as random factor and trial number as a repeated effect (covariance structure AR1). The fixed factors were trial type (alternating and non-alternating trials), age (six, eight and ten months), group (no-FR and FR infants) and habituation stimulus (/faːp/ or /feːp/). A two-way interaction Trial Type*Group or a three-way interaction Trial Type*Age*Group would show that groups (no-FR and FR) responded differently to the alternating and non-alternating trial types and would indicate differences in discrimination performance. The model that best fitted the data included the fixed factors trial type (alternating and non-alternating trials) and age (six, eight and ten months). This model yielded significant effects of 1) trial type on looking time, $F(1, 1346) = 71.63$, $p < .001$, indicating that infants looked longer to alternating trials than to non-alternating trials, and 2) age, $F(2, 158) = 5.58$, $p = .005$, indicating that overall looking times decreased as age increased. As can be seen in Table 5, within the no-FR and FR groups all age groups discriminated the native vowel contrast. However, the effect sizes in Table 5 suggest that there are differences between the groups. The effect size (Cohen's *d*) per age group for the no-FR infants increases from a moderate (.49), to a large value (.76). This is not the case for the FR infants; these effect sizes drop from a moderate effect size of .53 at 6 months to a moderate-small effect size of .36 at 10 months. Large variations in looking times results in smaller effect sizes, implying a less robust effect of trial type.

In sum, both groups of infants were able to generalize over speaker variations during habituation and responded to those acoustic features that differentiate between Dutch /aː/ and /eː/, regardless of whether the habituation stimulus was /faːp/ or /feːp/. However, whereas for the no-FR infants there was an increase in the effect size of the mean difference between alternating and non-alternating trials across age, this was not seen for the FR infants, due to the 10-month-olds who showed more variance between infants.

**The non-native contrast**

***Test phase.*** Table 6 displays the results of the test phase. A random effect modeling analysis included participant as random factor and trial number as a repeated effect (Covariance structure AR1). The fixed factors were trial type (alternating and non-alternating trials), age (six, eight and ten months), group (no-FR and FR infants) and habituation stimulus (/sæn/ or /sɛn/). The model that best fitted the data included the fixed factors trial type (alternating and non-alternating trials), age (six, eight and ten months), group (no-FR and FR) and habituation stimulus (/sæn/ or sɛn/). Infants looked longer to alternating than to non-alternating trials, $F(1, 1321) = 12.63$, $p < .001$. The significant Trial Type*Age*Group interaction $F(7, 545) = 3.69$, $p = .001$, was explored by Bonferroni-adjusted pairwise comparisons. No-FR infants aged 8 months did not show a significant difference between alternating and non-alternating trials, whereas the other two age groups did; see Table 6. FR infants showed no evidence of discrimination in any of the age groups. The interaction between Trial Type*Habituation Stimulus*Group was also significant, $F(3, 435) = 5.59$, $p = .001$. Post hoc analyses showed that the no-FR infants discriminated the contrast regardless of habituation stimulus. FR infants did not discriminate the contrast, also regardless of habituation stimulus. The fixed factor habituation stimulus yielded no main effect, $F(1, 158) = 1.33$, $p = .250$, and no significant interaction was found between Trial Type*Habituation Stimulus, $F(1, 1251) = .09$, $p = .767$, nor between Trial Type*Habituation Stimulus*Age, $F(5, 356) = .67$, $p = .647$. There was no effect of age, $F(2, 157) = .07$, $p = .935$, meaning that no evidence was found for a difference in overall looking times between age groups. The main effect of group was not significant, $F(2, 157) = .36$, $p = .551$.

Unlike the results of the native contrast, performances of the no-FR and FR groups clearly differ. The 6- and 10-month-old no-FR infants showed evidence of discrimination of the non-native vowel, whereas the 8-month-olds did not. The FR infants did not show evidence of non-native discrimination at any of the ages.

Table 5

*Looking times to alternating and non-alternating trials of the native contrast per group*

| Group | Age | Alternating trials | | Non-alternating trials | | Statistics | | | Participants | Infants with longer looking times to alternating trials* | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *M* | *(SD)* | *M* | *(SD)* | *F* | *p* | *Cohen's d* | *N* | *n* | % |
| No-FR | 6 | 9.9 | (5.2) | 8.0 | (3.1) | 9.04 | .003 | .49 | 28 | 17 | 60 |
| | 8 | 9.2 | (6.0) | 6.4 | (3.0) | 13.32 | <.001 | .66 | 24 | 15 | 62 |
| | 10 | 7.6 | (3.7) | 5.4 | (2.4) | 21.33 | <.001 | .76 | 27 | 20 | 74 |
| | All | 8.9 | (5.1) | 6.6 | (3.0) | 42.07 | <.001 | .60 | 79 | 52 | 66 |
| FR | 6 | 9.7 | (6.0) | 7.5 | (2.8) | 4.66 | .031 | .53 | 28 | 19 | 68 |
| | 8 | 8.2 | (4.9) | 6.4 | (3.1) | 5.54 | .019 | .48 | 24 | 18 | 75 |
| | 10 | 8.9 | (4.5) | 7.3 | (4.4) | 14.13 | <.001 | .36 | 27 | 17 | 63 |
| | All | 9.0 | (5.1) | 7.1 | (3.5) | 23.47 | <.001 | .46 | 79 | 54 | 68 |

*Note.* Looking times are given in seconds. **Preference to alternating trials* refers to the number of infants who had on average longer looking times to alternating trials than non-alternating trials.

Table 6

*Looking times to alternating and non-alternating trials of both groups for the non-native contrast*

| Group | Age | Alternating trials | | Non-alternating trials | | Statistics | | | Participants | Infants with longer looking times to alternating trials* | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *M* | *(SD)* | *M* | *(SD)* | *F* | *p* | *Cohen's d* | *N* | *n* | % |
| No-FR | 6 | 8.6 | (4.9) | 6.4 | (3.6) | 11.37 | .001 | .39 | 31 | 25 | 80 |
| | 8 | 7.0 | (2.7) | 7.1 | (3.1) | .28 | .599 | -.03 | 27 | 14 | 51 |
| | 10 | 8.4 | (4.2) | 6.0 | (2.1) | 20.82 | <.001 | .44 | 26 | 22 | 85 |
| | All | 8.1 | (4.1) | 6.4 | (3.0) | | | .45 | 84 | 61 | 73 |
| FR | 6 | 8.5 | (6.1) | 7.6 | (2.3) | 1.40 | .237 | .23 | 31 | 17 | 55 |
| | 8 | 7.1 | (3.5) | 7.3 | (2.7) | .63 | .427 | -.07 | 27 | 11 | 41 |
| | 10 | 8.4 | (5.2) | 7.7 | (5.1) | 3.01 | .081 | .14 | 26 | 18 | 69 |
| | All | 8.0 | (5.1) | 7.6 | (3.6) | 2.26 | .133 | .10 | 84 | 46 | 55 |

*Note.* Looking times are given in seconds. **Preference for alternating trials* refers to the number of infants who had on average longer looking times to alternating trials than non-alternating trials.

**Individual Analyses: Bayesian hierarchical modeling**

Findings of the Bayesian hierarchical regression model are presented in Table 7. The parameter of interest was trial type (alternating vs non-alternating trials), as this allowed us to establish whether the looking times differed between the alternating and non-alternating condition for the individual infants. The aim was to classify infants as discriminators or non-discriminators. Using a frequentist approach this would mean that an individual is classified as being able to discriminate a contrast if the mean difference between alternating and non-alternating trials differs significantly from zero. Here we follow a similar criterion. We checked which infants had the value zero in their 95% credibility interval (CI) for the trial type parameter: inclusion of value zero means that there is no evidence for a difference in looking times between the two trial types (alternating vs. non-alternating). See Table 7 for the percentages of infants of which the 95% CI did not cross zero and see Appendix D-G for the group and individual estimates and the variance per group per contrast.

Table 7

*Number and percentage of infants whose 95% credibility interval (CI) did not include the value zero*

| Bayesian hierarchical modeling | | | | |
|---|---|---|---|---|
| | | | Nr. of infants whose 95% CI did *not* include zero | |
| Contrast | Age group | Total nr. of participants | Group | |
| Native | | | No-FR | FR |
| | 6 | 55 | 3/28 (11%) | 1/27a (4%) |
| | 8 | 48 | 4/24 (17%) | 16/24 (67%) |
| | 10 | 54 | 15/27 (56%) | 6/27 (22%) |
| | Subtotal | 158 | 20/79 (25%) | 23/79 (29%) |
| Non-Native | | | | |
| | 6 | 61 | 19/30a (63%) | 0/31 (0%) |
| | 8 | 53 | 0/27 (0%) | 0/26a (0%) |
| | 10 | 51 | 21/25a (84%) | 0/26 (0%) |
| | Subtotal | 168 | 41/84 (49%) | 0/84 (0%) |
| | Total | 326 | 61/163 (37%) | 23/163 (14%) |

*Note.* Data of four participants were not included in the Bayesian analysis. This was due to missing data. The superscript (a) in the columns No-FR-FR indicates in which group data was missing.

The parameter of interest was trial type (alternating vs non-alternating trials), as this allowed us to establish whether the looking times differed between the alternating and non-alternating condition for the individual infants. The aim was to classify infants as discriminators or non-discriminators. Using a frequentist approach this would mean that an individual is classified as being able to discriminate a contrast if the mean difference between alternating and non-alternating trials differs significantly from zero. Here we follow a similar criterion. We checked which infants had the value zero in their 95% credibility interval (CI) for the trial type parameter: inclusion of value zero means that there is no evidence for a difference in looking times between the two trial types (alternating vs. non-alternating). See Table 7 for the percentages of infants of which the 95% CI did not cross zero and see Appendix D-G for the group and individual estimates and the variance per group per contrast.

***Native vowel contrast.*** Of the no-FR infants, 11% of the 6-month-old infants, 17% of the 8-month-olds, and 56% of the 10-month-olds did not have zero in their 95% CI (see Table 7). Hence, these values are taken to reflect the percentages of children that discriminate between the native vowels. Appendix D, Figure D1 shows the estimated medians and credibility intervals. The data indicate that the individual 95% CIs for the 8-month-olds show larger uncertainty than the individual 95% CIs of the other age groups. Figure D2 shows that the variance estimates are larger for the 8-month-olds than the other age groups: the 8-month-olds differ more from one another than the 6- and 10-month-old participants. Larger variance at the group level influences the individual estimates as these become more uncertain; hence, fewer infants can be classified as discriminators. Figures 1 and 2 in Appendix D show that the estimated effect of condition in the 10-month-olds is comparable to that of the 8-month-olds. The higher percentage of infants that do not have zero in their CIs is due to the smaller variance in the 10-month-olds as a group. They resemble one another more than do the 8-month-olds and therefore we are more confident about their estimated condition effects at the individual level. Hence, this might indicate that discrimination of the native /a:/ - /e:/ contrast becomes more robust with maturation.

The percentages of the FR infants that do not have zero in their CIs, and are thus considered to discriminate the native contrast, are 4% for the 6-month-olds, 67% for the 8-month-olds, and 22% for the 10-month-olds; see Table 7. Figures E1 and E2 (Appendix E) show that the group estimates of the condition effect are similar for the 8- and 10-month-olds. The 10-month-old FR infants show more variance in the group estimates of the condition effect (Appendix E, Figure E2) and, hence, larger uncertainty in their individual 95% CIs, similarly to the no-FR 8-month-olds. The finding that, compared to the 8-month-old FR infants, so few 10-month-old FR infants can be classified as discriminators is due to large uncertainty in the individual estimates of the 95% CI. As in the 10-month-olds, very few 6-month-olds have CIs that do not include the value zero. The group variances of the 6- and 8-month-olds (right panel Figure E2) are comparable. So, variance at the group level cannot explain the difference in the percentages of infants (6 vs. 8 months old) that do not include the value zero in their CIs. However, the estimate of the condition effect and the credibility interval of this effect are closer to zero (left panel Figure E2) for the 6-month-olds. In this hierarchical model, the estimated effect of condition at group level functions as a prior for the individual condition effect estimates, and individual estimates are pooled towards these group estimates. The literature has shown that incorporating group structures in the analyses lead to fewer mistakes for the individual parameters that are estimated in terms of the magnitude and the sign (direction) of the effects. This issue is addressed more elaborately in the literature on Type S and Type M errors, mostly by Gelman and Tuerlinckx (2000). Thus, keeping the group level variance equal, a smaller group level estimate for the condition effect will pool the individuals towards smaller individual effects, as can be seen in the 6-month-olds. At the group level, we find evidence for discrimination and homogeneity of the group. As the group estimate and the CI are closer to zero, the individual estimates are too.

To summarize, the no-FR infants can discriminate the native contrast. The data suggest an enhancement effect, as the percentage of infants that discriminate the contrast at 6-months of age is low, but this increases with age. The FR infants show a different developmental pattern. The 6-month-olds show evidence of discrimination at group level, but very few individuals can be classified as discriminators. The 8- and 10-month-olds are able to discriminate the native contrast at group level. At the individual level, however, too much uncertainty remains to classify many of the 10-month-old infants as discriminators.

**Non-native vowel contrast.** The percentage of no-FR infants that do not have zero in their 95% CI is 63% for the 6-month-olds, 0% for the 8-month-olds and 84% of the 10-month-olds. In Appendix F, Figures F1 and F2, it can be seen that the low percentage of the 8-month-olds cannot be attributed to larger uncertainty of the CIs, nor to the larger variation of the group estimates. The individual estimates of the median condition effect are close to zero, and the 95% CIs convincingly cross zero.

The FR infants show different results: for all age groups, none of the FR infants discriminated between the non-native vowels, as they all have zero in their 95% CI. These results cannot be explained by larger variability or uncertainty: The individual estimates of the means are all at or close to zero; see Appendix G, Figures G1 and G2. However, although the percentage of the 10-month-old infants that do not have zero in their CIs is 0, the group estimate for the condition effect is not close to zero, which indicates some effect of condition at group level. Also, the individual mean estimates are not all close to zero, see right panel of Figure G1, and hence some CIs barely cross the zero, as was seen in the 6-month-old FR infants in the native condition. The 6- and 8-month-old FR infants did not discriminate the non-native contrast, not at group level, nor at the individual level. At 10 months of age we find discrimination at the group level but the uncertainty at the individual level means that we cannot classify any individual as discriminator.

Together, the cross-sectional data of the no-FR infants suggest a U-shaped pattern of discrimination of the non-native contrast (discrimination at 6 months, not at 8 months, but again at 10 months) similar to our previous findings (Chapter 2). The 6- and 8-month-old FR infants did not show evidence of discrimination at the individual level, nor at group level. However, as was seen with the 6-month-olds in the native condition, the 10-month-olds do show some evidence of discrimination at group level, not at the individual level.

## 4.4 Discussion

The aim of this study was twofold. The first was to evaluate whether a similar pattern of perceptual attunement would be attested for children with and without a family risk of dyslexia. The second was to assess whether this pattern was reflected in group findings as well as individual-based analysis. In order to look into these questions, discrimination of native and non-native phonemes in 6-, 8-, and 10-month-old infants was studied. Perceptual attunement would be attested if 1) the native contrast (/aː/ - /eː/) were discriminated at all ages or when discrimination improved with age, and 2) discrimination performance of the non-native contrast (English /ɛ/ - /æ/) declined with age. In light of the proposed speech perception deficit in children with (an FR of) dyslexia (e.g., Molfese 2000; Richardson et al., 2003; Werker & Tees, 1987) it was expected that the FR infants would show evidence of discriminating the salient native contrast, but that this discrimination would show a slower improvement than in the no-FR infants. With respect to the subtle non-native contrast, it could either be the case that the FR group would not discriminate this contrast at any timepoint (e.g., van Leeuwen et al., 2006; Richardson, et al., 2003), or that there would be continued discrimination, opposed to a decrease in the no-FR group (Noordenbos, et al., 2012).

There was no evidence for a difference between the no-FR and FR infants on the native speech contrast. The (frequentist) group findings of the native contrast showed that both the no-FR and FR infants discriminated salient Dutch /a:/ from /e:/. However, the effect sizes show subtle differences between the two groups: Whereas there was an increase in effect sizes over age for no-FR infants between alternating and non-alternating trials at the group level, this was not found for the FR infants. Hence, the variability reduces with age in no-FR infants. This could be indicative of an increasingly robust discrimination performance with age/maturation, aligning with theories of enhancement of native speech perception (Kuhl, et al., 2008; Tsuji & Cristia, 2014). The FR group does not show this increase, which may imply that there is a subtle delay. A longitudinal study extending to older age groups can be used to investigate whether an enhancement effect does surface for the FR group at a later age.

The Bayesian individual outcomes showed that the percentage of no-FR infants that discriminate the native contrast at 6-months of age is low and increases with age. The observation that the 8-month-olds, compared to the 10-month-olds, showed a relatively low percentage of discriminators is explained by the larger variance in the group estimates. The pattern of the FR infants at the individual level is different. The 6-month-olds showed (very) weak evidence of (native) discrimination, and a relatively low percentage of 10-month-olds could be classified as discriminators, compared to the 8-month-olds. The explanation is the same here as it is for the 8-month-old no-FR infants: The variance for the 10-month-old FR infants is larger than that of 8-month-olds. Larger variance at the group level influences the individual estimates, as these become more uncertain. Hence, fewer infants can be classified as discriminators. The 10-month-old FR infants as a group behave less coherently, as they differ more from one another, matching the findings from the frequentist analysis. This could be indicative of a subtle delay in speech perception development.

The outcomes of the non-native contrast showed a different picture, as there were pronounced differences between the no-FR and FR groups. Both the frequentist group analysis and the Bayesian individual analysis suggest a pattern of U-shaped-development for the no-FR group but not for the FR group. The 6- and 10-month-old no-FR infants showed evidence of discrimination, whereas the 8-month-olds did not. These findings seem to confirm the findings of perceptual attunement between 6 and 8 months (Kuhl, et al., 1992; Polka & Werker, 1994). In a previous study (Chapter 2), we proposed that the improved performance at 10 months is due to their being better equipped than the 6-and 8-month-olds to make use of the speaker variation presented in the habituation phase.

For the FR infants, the frequentist group analysis showed no evidence of discrimination in any of the age groups. This was mirrored in Bayesian hierarchical modeling, in which none of the FR infants could be classified as discriminators, at any age. These findings are difficult to relate to the classical view on perceptual attunement. However, the literature on perceptual attunement has shown that salience influences the ability to discriminate, as more subtle and less salient contrasts need to be acquired through language exposure (e.g Liu & Kager, 2016; Narayan et al., 2010). Hence, the finding that even the 6-month-old FR infants did not show evidence of discrimination could be due to a lack of initial sensitivity to subtle contrasts and, as infants are not exposed to this non-native contrast, their discrimination performance does not improve. The present findings indicate that it might be important to investigate the developmental trajectory of discrimination performance of subtle *native* contrasts. Based on the data presented in this study a delay in discrimination performance is expected.

FR infants showed evidence of discriminating the native, but not the non-native contrast. Hence, the data of the current study supports the notion that FR infants have a (subtle) speech perception deficit. This is in line with studies that investigated speech perception at an early age (e.g., Richardson et al., 2003) and also with studies that investigated the relation between early speech sound processing and later reading outcomes (e.g., Molfese, 2000; van Zuijen, et al., 2013). Outcomes of the current study do not support the hypothesis that infants remain sensitive to irrelevant non-native contrasts (Noordenbos, et al., 2012). Furthermore, the data of the present study suggest a subtle delay in the development of speech sound categories. There is evidence that distributional learning plays a critical role in the acquisition of native speech sound categories (Kuhl, et al., 1992, Maye et al., 2008). Notably, Maye et al. (2008) have shown that the frequency distribution of non-native speech sound tokens (unimodal or bimodal) that differ along an acoustic parameter determines whether 8-month-old infants assign them to one or two classes. We hypothesize that FR infants are less proficient in exploiting such distributional information (see Kerkhoff et al., 2013; Wijnen, 2013). This is a hypothesis we are currently investigating.

There are three findings in the current study that require clarification. The first is the finding that the 8-month-old no-FR infants have a lower percentage of discriminators of the native contrast compared to the 6- and 10-month-olds. This finding, suggestive of a U-shaped development of native vowel discrimination, is due to the larger variance at group level at 8 months. In line with Werker and colleagues (2004), we propose that the heterogeneous performance at this age reflects a developmental (reorganizational) shift in vowel perception, and that it does not reflect a loss of discrimination. The shift refers to a change in processes and strategies applied during speech perception. Younger infants react to all perceivable phonetic differences and thus discriminate all speech sounds. During the reorganizational phase, they begin to learn to categorize speech sounds in phonemic units (Kuhl, 2004). As not all infants begin their shift at exactly the same timepoint, this could explain the heterogeneity of group performance at 8 months. Although the reorganization in speech perception has thus far been observed in tasks assessing discrimination of non-native speech sounds that are assimilated to native speech sounds (Maurer & Werker, 2014), we submit that our individual-based analysis also captured this pattern for a salient native contrast. Hence, we argue that the U-shape pattern reveals the underlying process of speech perception, which is in line with other studies that interpret U-shape findings (Bjorklund et al., 1997; Kachel et al., 2020; Pauls et al., 2013; Siegler, 2004 and other references in that issue).

The second finding that requires further consideration is related to this first issue of a 'dip' in performance of the 8-month-olds in the native speech contrast. It concerns the finding that no enhancement effect was found for the 10-month-old FR infants in the native contrast. Instead, the analyses at the individual level showed a declining percentage of infants that can be classified as discriminators. We would expect more robust discrimination of the native contrast by the 10-month-old FR infants on the basis of the finding that 8-month-old FR infants are able to discriminate this contrast. The 10-month-olds have had more experience with their native language and the salient native contrast and would therefore show better discrimination. However, if FR infants indeed have a subtle delay in the development of speech sound categories, as our data suggests, the lower percentage of FR infants that discriminate the native contrast at 10 months of age could be an indicator of them being at the reorganizational phase, similar to the 8-month-old no-FR infants. Further research in the underlying processes of this reorganization shift is welcome.

The third finding that needs clarification is that the no-FR 10-month-olds could discriminate the non-native contrast, as this was not anticipated on the basis of perceptual attunement. This cannot be due to exposure, as infants are not exposed to this non-native contrast in real life. We propose that the 10-month-old no-FR infants could discriminate this contrast because they were better able to make use of the speaker variation presented during the habituation phase. Speaker variation stimulates phonetic learning as it demands abstraction of invariant features (Lively et al., 1993), but this effect is likely to become stronger as age increases (see also Chapter 2; Rost & McMurray, 2009). The fact that this effect was not seen in the FR infants might indicate that this phonetic distributional information caused by speaker variation is not helpful to the same extent for FR infants at this age as it is for their no-FR peers.

We have shown that individual analysis can be used to infer whether infants in two different groups are discriminators of native and non-native speech sound contrasts. Moreover, we found large differences between the two groups of infants, indicating that the method presented here could be used to study language(-related) development prospectively. Although previous studies have connected infant speech perception data to later language outcomes (e.g., Newman et al., 2006; Tsao et al., 2004), using Bayesian hierarchical modeling might render even more sensitive results, as it is able to produce both group and individual estimates and could be extended in a straightforward manner to address prospective research questions, relating early speech perception to later language outcomes. For example, the results of the individual estimates can be used in a prospective longitudinal design in which language outcomes, such as vocabulary size; mean length of utterance; sentence complexity or reading outcomes, are predicted by the (amount of) discrimination performance at this early age. Future studies can thus provide valuable input on the question whether discrimination skills at an early age are associated with reading problems or are instead a risk factor (endophenotype) for developing dyslexia (Moll et al., 2013).

Although the method presented here for identifying discriminators seems a fruitful avenue for gaining more understanding of development, group differences, as well as prospective relationships, some limitations need to be mentioned. In the current approach, individual and group estimates are influencing one another. It is desirable to obtain a sufficiently large sample size to estimate group level parameters with confidence. Another limitation is that we did not assess discrimination performance longitudinally and relied on cross-sectional data. Testing speech sound discrimination longitudinally would allow us to establish whether the U-shaped patterns are also attested in such a sample and would provide more insight in whether individual classification is as valuable as we take it to be.

To conclude, we hope to have shown that individual analysis in speech discrimination experiments with infants is a promising avenue for further research. There are still some challenges using the Bayesian hierarchical approach, but it provides us with a tool that allows us to better understand how speech perception develops at an individual level (preferably longitudinally) as well as looking prospectively at the relationship with other facets of language and literacy development.

**Ethics statement**

Informed consent was obtained from the parents before testing; consent and participation could be retracted at any time. The authors declare that the research was conducted in accordance with APA ethical standards as well as *The Netherlands Code of Conduct for Scientific Practice* issued in 2004 (revised in 2018) by the Association of Universities in the Netherlands (VSNU).

**Acknowledgements**

We thank the parents of our participants for their contributions to our study. We thank Sule Kurtçebe, Tinka Versteegh, Lorijn Zaadnoordijk en Joleen Zuidema for assistance in data collection and Annemarie Kerkhoff for assistance in the design of the study. This study was funded by The Netherlands Organization for Scientific Research (NWO, grant nr. 360-70-270).

# Chapter 5

# Distributional phonetic learning in infants at family risk of dyslexia

**Abstract**
In this study we evaluated whether infants with a family risk (FR) of dyslexia are sensitive to the frequency distributions of speech sounds by conducting a distributional learning experiment (Maye et al., 2008). During the familiarization phase, Dutch 8-month-old infants with and without an FR were exposed to speech sound tokens of four English /ɛ/ -/æ/ 8-step continua, either in a unimodal or bimodal frequency distribution. In the test phase, it was assessed whether infants were able to discriminate token 3 and token 6. Depending on the frequency distribution, tokens 3 and 6 belong to the same (unimodal condition), or to two different categories (bimodal condition). If infants were sensitive to frequency distributions of the speech sounds in the input, those in the bimodal condition would discriminate the contrast, whereas those in the unimodal condition would not. This was expected for the infants without an FR (no-FR) of dyslexia, but not for the FR infants. Confirmatory statistical analysis did not yield evidence for an effect of condition (uni- vs bimodal) on discrimination in either group. Further data exploration, however, provided subtle indications for learning by the no-FR infants in the bimodal condition compared to the unimodal condition. Such indications were absent for the FR infants in the bimodal condition. Together, the data (very) tentatively suggest that FR infants are less sensitive to distributional learning

de Klerk, M., de Bree, E., & Wijnen, F. (in revision). Distributional Phonetic Learning in Infants at Family Risk of Dyslexia.

EdB, MdK and FW designed the study. MdK tested the infants and conducted the analyses. MdK wrote and revised the paper, with input and feedback from EdB and FW.

## 5.1 Introduction

Dyslexia is a language-specific neurodevelopmental disorder characterized by severe and persistent word recognition and spelling difficulties despite adequate instruction and intact sensory abilities (Lyon et al., 2003). Dyslexia is proposed to be a multifactorial disorder, with multiple genetic, environmental and cognitive risk- and protective factors interacting with each other (Peterson & Pennington 2015). Dyslexia has been found to be highly heritable, as children with a dyslexic parent have an increased risk of developing dyslexia (Snowling & Melby-Lervåg, 2016).

A phonological deficit is generally assumed to be one of the primary cognitive risk factors of dyslexia (e.g., Ramus, 2003; Vellutino, et al., 2004). Individuals with dyslexia have weak phonological coding skills, resulting in poorly specified phonological representations, which, in turn, impede the integration of speech sounds (phonemes) and letters (graphemes) (e.g., Blomert, 2011; Mittag et al., 2013; Vellutino et al., 2004). The formation of phonological representations, or speech sound categories, begins during the first year of life, when infants tune in to their native language (Werker & Tees, 1984). It is hypothesized that speech sound categories are acquired with the support of distributional learning (Kuhl, 2004), and there is evidence for phonetic distributional learning in infants (e.g., Maye et al., 2008). The current study investigates whether infants with a family risk (FR) of dyslexia are less sensitive to the frequency distributions of speech sounds compared to their no-FR peers.

Generally, infants are sensitive to distributional (statistical) properties of their language(s) they are exposed to: They are capable of tracking regularities in language input, such as the number of phoneme and word (co-)occurrences and sequences, and use these regularities to detect word-boundaries (e.g., Saffran et al., 1996), syntactic categories (e.g., Gerken et al., 2005), and speech sound categories (e.g., Maye et al., 2008). This sensitivity to distributional properties of the language input has been proposed to account for the process of perceptual attunement (Kuhl, 2004), which is a shift in speech sound perception that occurs during the first year of life. Initially, infants are generally sensitive to all speech sound contrasts, whereas during the second half of the first year of life, they have more difficulty in discriminating non-native contrasts. For instance, while both English and Japanese infants can distinguish [l] and [r] in the first few months of life, Japanese (but not English) infants lose this sensitivity in the ensuing months. The [l] - [r] contrast is not relevant in the phonology of Japanese, which is signalled by the frequency distribution of speech sounds in the [l] - [r] range (Kuhl, 2004). This distribution in Japanese is very different to the distribution of speech sounds in the [l] - [r] range that English-learning infants are exposed to. In English, the distribution of the speech sounds is bimodal, the two modes corresponding to acoustic values that are most frequently produced, resembling [l] and [r]. Japanese infants, in contrast, are exposed to a unimodal distribution (Lotto et al., 2004). The hypothesis is that infants create speech sound categories by keeping track of this distributional information, and, consequently, attune to the speech sounds in their target language (Kuhl, 2004). Thus, although there is inter- and intraspeaker variation which leads to some overlap in the acoustic values (second and third formant syllable-onset frequency in a CV-syllable) of [r] and [l], the frequency distributions in Japanese and English differ, leading to different sensitivity to the [l] and [r] contrast.

Maye et al. (2008) were the first to provide evidence for distributional learning in the emergence of speech sound categories. They exposed English learning infants to the non-native Hindi prevoiced vs. short-lag (unaspirated) stop consonant contrast /da/ - /ta/ and /ga/ - /ka/ contrast. Both contrasts are difficult to discriminate for English learning infants (e.g., Aslin et al., 1981). During familiarization, infants were exposed to 8-step /da/ - /ta/ and /ga/ - /ka/ continua, of which tokens were distributed either unimodally (tokens in the middle presented most frequently) or bimodally (tokens near the endpoints presented most frequently). Only the group of infants exposed to the bimodal distribution were subsequently able to discriminate the non-native contrasts; those in the unimodal condition were not. The finding that infants can learn phonetic contrasts through exposure to a bimodal distribution of speech sounds has been replicated by others with a variety of consonantal contrasts (Capel et al., 2011; Yoshida et al., 2010), vowel contrasts (ter Schure et al., 2016; Wanrooij et al., 2014) and tonal contrasts (Liu & Kager, 2014, 2017). Furthermore, distributional learning has been found to be most effective around the period in which infants tune into the phonological system of their native language (Liu & Kager, 2017, Reh et al., 2020; Yoshida et al., 2010). These studies suggest that distributional learning contributes to the acquisition of speech sound categories.

On the basis of both the phonological deficit of dyslexia and findings of poorer distributional learning in dyslexia (Banai & Ahissar, 2018; Bonte et al., 2007; Kerkhoff et al., 2013; Vandermosten et al., 2019; van Witteloostuijn et al. 2017; Wijnen, 2013), we expect that distributional learning of speech sounds will be poorer in children with (an FR of) dyslexia. There is some evidence that children with dyslexia are less sensitive to frequency distributions of speech sounds. Vandermosten et al. (2019) assessed distributional learning of a non-native Hindi dental-retroflex /da/-/ɖa/contrast, using a 7-step continuum, in Dutch-speaking Grade 3 children. Children were administered a two-alternative forced choice identification task before and after the exposure phase, in which tokens from the speech sound continuum followed either a unimodal or bimodal distribution.

A comparison between groups of children that were exposed to the bimodal condition showed that non-dyslexic children improved their performance due to the statistical information provided. This was not found for the children with dyslexia, suggesting that they were less sensitive to variation in speech sound distributions. If such a diminished sensitivity can also be demonstrated at a much earlier age, this will contribute to our understanding of the phonological deficit in dyslexia.

**Current study**

The aim of the present study was to evaluate whether infants with an FR of dyslexia show poorer distributional learning of speech sounds. We are not aware of any earlier studies into distributional phonetic learning in FR infants. It has, however, been found that children with (an FR of) dyslexia perform more poorly on tasks assessing sensitivity to distributional learning in order to detect word boundaries and (grammatical) strings (e.g., Banai & Ahissar, 2018; Bonte et al., 2007; Kerkhoff, et al, 2013; Lum et al., 2013). Based on these findings, FR infants would be expected to show poorer outcomes on a task that measures distributional phonetic learning.

We used the design developed by Maye et al. (2008). The experiment consisted of a familiarization, habituation and test phase and a post-test. In the familiarization phase, 8-month-old infants were familiarized to tokens from four 8-step continua. The tokens were presented according to either a unimodal or a bimodal frequency distribution. In the unimodal condition, speech sounds in the middle of the continua were presented most frequently, and frequency of exposure declined towards the endpoints of the continuum. In the bimodal condition, the stimuli near the endpoints of the continua were presented more frequently than those in the middle (see *Procedure*). The tokens that were presented equally frequently in both conditions (tokens 3 and 6) were used during habituation and test. During habituation, infants were habituated to one of these tokens and tested with the other token in the test phase. If infants are sensitive to frequency distributions, they will be able to discriminate the English contrast when exposed to the bimodal, but not the unimodal condition.

We used the English /ɛ/ -/æ/ contrast, which is difficult for learners of Dutch to discriminate (Broersma & Cutler, 2011). In a study from our lab (Chapter 4) we found no evidence of discrimination of this contrast in no-FR and FR 8-month-old infants. In the present study, we predicted that FR infants would not show evidence of discriminating the English /ɛ/ -/æ/ contrast, regardless of the frequency distribution. For the no-FR infants we hypothesized that discrimination would be evidenced in the bimodal, but not in the unimodal condition.

## 5.2 Method

### Participants

Eight-month-old infants were recruited via an invitation letter sent out to parents in the area of Utrecht City. Addresses were obtained via the local municipality. Infants were included if (a) they were raised as monolingual speakers of Dutch; (b) their gestational age at birth was between 37 and 43 weeks; (c) their weight at birth was between 2500 and 4500 grams; (d) there were no severe complications during pregnancy and/or delivery; (e) the infant did not have a history of known hearing loss and/or reduced vision; and (f) the infant did not have any reported neurological problems. Background information on the participants and the educational level of the parents are presented in Table 1.

Table 1

*Background Information of Participants*

| Measure | no-FR | FR | *Statistics (two-sided)* |
|---|---|---|---|
| | *M (SD)* | *M (SD)* | |
| Birth weight (grams) | 3554 (366) | 3760 (447) | $t(72) = 2.34$, $p = .022$ a |
| Gestational age (weeks) | 40.3 (1.4) | 40.5 (1.1) | $U = 780.00$, $z = .82$, $p = .413$ b |
| Educational level mother | 5.66 (.49) | 9.0 (.5) | $U = 660.00$, $z = -.51$, $p = .612$ b |
| Educational level father | 5.47 (.66) | 5.0 (1.0) | $U = 557.00$, $z = -1.47$, $p = .141$ b |
| No. of siblings | .31 (.6) | .30 (.4) | $U = 816.00$, $z = .46$, $p = .647$ b |
| Birth order | .31 (.6) | .30 (.4) | $U = 816.00$, $z = .46$, $p = .647$ b |

*Note.* Educational level was measured on a scale from 1 (primary school) to 6 (university/PhD). The superscripts: (a) data was normally distributed, an independent T-test was conducted, (b) data was not normally distributed, non-parametric (Mann-Whitney) tests were conducted.

Infants were considered to be at family risk of dyslexia when the parent had reported a history of severe literacy problems. This was confirmed by tests (Kuijpers et al., 2003) administered during the parent's visit to the lab, consisting of a timed word reading test, the 'Een-minuut-test' (EMT; Brus & Voeten, 1972) a timed pseudoword reading test ('de Klepel'; van den Bos et al., 1994) and a verbal competence test (Analogies), a subtest of the Dutch version of the Wechsler Adult Intelligence Scale (WAIS; Uterwijk, 2000).

Infants were included in the FR group if parents met one of the following criteria:

1) the percentile score on one of the reading tests was ≤ 10 ($n$ = 37); 2) the percentile score was ≤ 20 on both reading tests ($n$ = 1); or 3) the discrepancy between one of the reading tests and the verbal competence test was 60 percentile points or more ($n$ = 3, Kuijpers et al., 2003). Two parents refused to take the reading test; they were classified as dyslexic on the basis of self-report and the fact that they had already been formally diagnosed with dyslexia.
A total of 50 FR-infants were tested for the current experiment, of which seven (14%) were excluded from the data analysis due to crying during the experiment ($n$ = 6) or the caregiver interfering by pointing and talking throughout the experiment ($n$ = 1). Fifty no-FR-infants were tested, of which eight (16%) were excluded based on crying ($n$ = 7) or not reaching the habituation criterion (see Procedure; $n$ = 1). Table 2 describes the age ranges, mean ages, drop-out rates, the numbers of infants assigned to each condition (unimodal or bimodal, see Procedure) as well as the gender distribution. There were no differences between the groups regarding gender distribution (no-FR vs. FR, $\chi^2$ ( 85) = 1.41, $p$ = .281 (two-sided), drop-out rates, $\chi^2$ = (100) = .78, $p$ = .779 and mean age, $t$(83) = 1.57, $p$ = .120.

Table 2
*Age range, mean age and number of participants per group and condition*

| Group | Age range | Mean age | Condition | Infants tested | Excluded | | Included |
|---|---|---|---|---|---|---|---|
| | *months.days* | *M (SD)* | | *N* | *n* = | (%) | *n* (female) |
| no-FR | 8.0 - 8.30 | 256 (8.7) | Unimodal | 25 | 3 | (12%) | 22 (15) |
| | | | Bimodal | 25 | 5 | (20%) | 20 (8) |
| | | | Subtotal | 50 | 8 | (16%) | 42 (23) |
| FR | 8.0 - 8.30 | 258 (8.4) | Unimodal | 26 | 4 | (15%) | 22 (10) |
| | | | Bimodal | 24 | 3 | (13%) | 21 (8) |
| | | | Subtotal | 50 | 7 | (14%) | 43 (18) |
| | | | Total | 100 | 15 | (15%) | 85 (41) |

*Note.* no-FR = no family risk; FR = family risk. Mean age is given in days.

**Stimuli**

***Auditory stimuli.*** The vowel contrast is a non-native English /æ/-/ɛ/ contrast. Adult Dutch native speakers perceive both sounds as Dutch /ɛ/ (Broersma & Cutler, 2011), and Dutch learning 8-month-olds do not discriminate these sounds (Chapter 2). Four speech sound continua of the word pair /sæn/ - /sɛn/, created by Heeren (2006), were used. Four male speakers (mean age 32 years; range 21-55 years) of Standard English were instructed to read out a list of 50 words, containing real words (e.g., *have* and *pet*) and the nonce words (*san* and *sen*). The speakers were instructed to read the words in the list equally fast and loud and with a flat intonation. Speaker variation was introduced here, in view of evidence that variation stimulates phonetic learning, presumably by promoting the extraction of invariant features (e.g., Lively et al., 1993). Stimuli were recorded on a DAT tape, with a sample frequency of 48000 kHz. An Audio-technica AT841a microphone was used. The signal was high pass filtered through a preamplifier using a 4th order filter at 75 Hz.

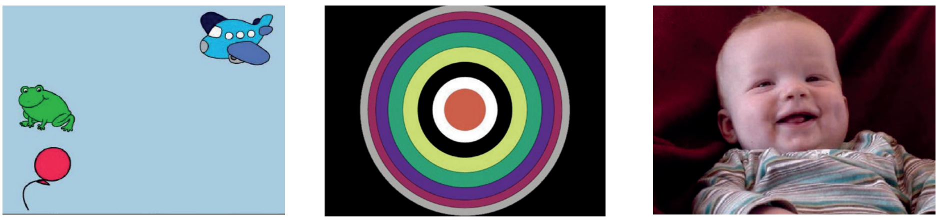
From each speaker one token of the nonce word /sæn/ was chosen by a phonetically trained listener on the basis of the quality of the recording. These four tokens were used to create the continua. A linear spectral interpolation was used to synthesize tokens, with help of the software program Provo (van Hessen, 1992, Chapter 3). The first and second formant frequencies of the speech sound tokens were manipulated. Other parameters, such as the pitch and duration of the vowel, and total duration of the stimulus were kept constant. These values did vary *between* continua, as the items were pronounced by different speakers, see Table 3. Intensity was set at 70db for all tokens. The first formant frequency (F1) decreased with steps of approximately 20Hz. The formant frequency of the second formant (F2) increased with steps of approximately 30Hz. All four continua were presented to 31 native listeners of British English in a classification task, in order to judge whether the endpoints were perceived as representatives of the English /ɛ/ and /æ/, see for details about the classification task Heeren (2006, p. 20-22). The endpoints of the continua were always classified as the intended endpoints (100%).

Table 3
*Acoustic characteristics of the endpoint stimuli*

| Stimulus | Total duration | Vowel duration | F1 | F2 | F0 |
|---|---|---|---|---|---|
| *Speaker-token* | *msec.* | *msec.* | *Hz* | | |
| Speaker1-01 | 627 | 237 | 672 | 1598 | 132 |
| Speaker1-08 | 627 | 237 | 546 | 1797 | 132 |
| Speaker2-01 | 668 | 256 | 680 | 1495 | 157 |
| Speaker2-08 | 668 | 256 | 533 | 1734 | 157 |
| Speaker3-01 | 589 | 169 | 672 | 1396 | 132 |
| Speaker3-08 | 589 | 169 | 530 | 1620 | 132 |
| Speaker4-01 | 531 | 138 | 688 | 1340 | 109 |
| Speaker4-08 | 531 | 138 | 578 | 1590 | 109 |

***Visual stimuli.*** Three different types of visual stimuli were used. During the familiarization phase infants looked at three cartoon pictures (see an example in left-hand panel of Figure 1) presented on the computer screen in one of the squares of a (3x3) grid. The grid lines were invisible; only pictures were shown. These three pictures were randomly drawn from a bank of 25 pictures. Every two seconds three new pictures appeared in one of the nine possible locations. Pictures were presented in pseudo randomized order. The visual stimulus during the habituation, test phase and the post test was a colourful bull's eye (middle picture in Figure 1). In between test trials a movie of a cute laughing baby was presented (a still is presented in the right-hand picture in Figure 1).

Figure 1
*Visual stimuli*



Visuals of familiarization phase

Visual of habituation-dishabituation phase

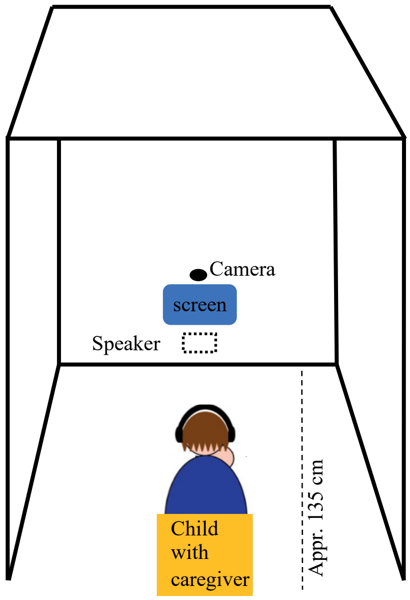
Attention grabber in between trials of the habituation-dishabituation phase

*Note.* Visual stimuli presented during each of the two phases and the attention grabber in between trials of the test phase (habituation-dishabituation).

**Procedure**

The infant was seated on the caregiver's lap in a three-walled off-white canvas test booth with a canvas ceiling in a sound-attenuated room adjacent to the experimenter's room. The computer monitor (Philips LCD 150P4) on which the visual stimuli were displayed was approximately 1.35 m from the child. Auditory stimuli were played through a loudspeaker (Tannoy i8) which was hidden behind the canvas of the booth and placed below the computer monitor that displayed the visual stimuli. The caregiver wore headphones (Telex, Echelon 20, over-ear headphones with passive noise attenuation of 20 dB), through which music was played in order to prevent the caregiver from hearing the stimuli and (potentially) influencing their child's behaviour. The experiment was recorded through a hidden video camera that was placed above the computer screen; see Figure 2.
The experimenter sat in the adjacent room and monitored the caregiver and infant through a closed-circuit TV. Looking times to the screen of the infant were recorded online with a button-box connected to an Asus P4PE computer. Looking times to the screen are interpreted as listening times (Aslin, 2007). A custom-made experiment control application (Veenker, 2008) was used for presentation of the auditory and visual stimuli and for data registration.

Figure 2
*Depiction of the test booth*



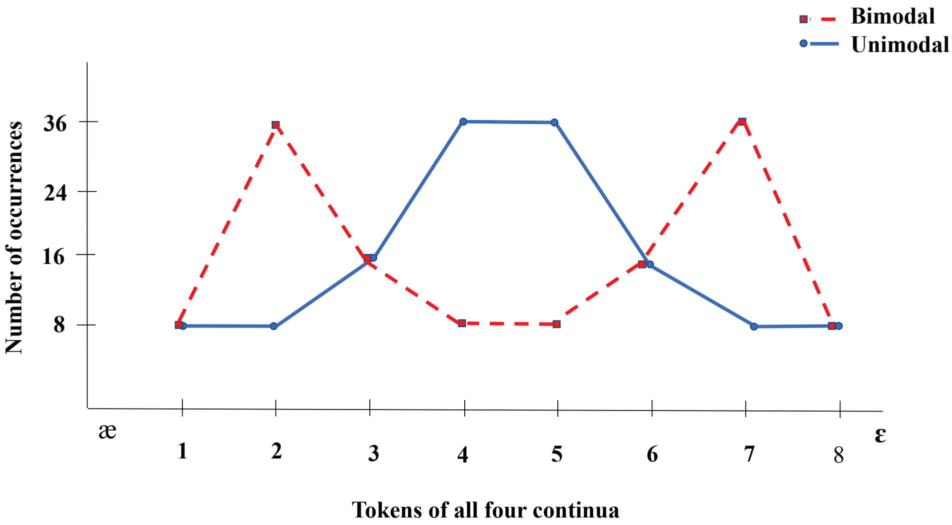
Prior to the experiment, the experimenter informed the caregiver about the procedure, without being too specific about the auditory stimuli and the design. The caregiver was instructed not to speak, but if the infant seemed uncomfortable, the caregiver was allowed to soothe the infant non-verbally. The experimenter explained that the session could be ended if the caregiver felt this was necessary and that the experimenter could do the same. The session was always terminated when the infant started to cry. The phases of the experiment, a familiarization phase, a habituation and test phase, and a post test, are explained below and are based on Maye et al. (2008).

***Familiarization phase.*** Infants were randomly assigned to one of the two conditions (unimodal or bimodal). All infants were familiarized to the 32 tokens of the four continua (8 tokens per continuum and 4 speech sound continua (see *Auditory stimuli*), is 32 tokens). The 32 tokens were presented randomly and presented with a stimulus interval of 1 second. The familiarization phase had a fixed duration of approximately 3 minutes and 30 seconds.

In both the unimodal and bimodal condition, tokens 3 and 6 were presented 16 times each. Hence, each infant heard the tokens 3 and 6 equally often, regardless of condition. These tokens were used in the habituation and test phases. Importantly, the frequency distribution of the other tokens differed between the two conditions. In the bimodal distribution, the near endpoints (tokens 2 and 7) of the continua were presented most frequently: 36 times, viz. from each of the four continua the tokens 2 and 7 were presented 9 times (9*4=36). Tokens 1, 4 and 5 and 8 of each continuum were presented twice, resulting in 8 (2*4) presentations. Tokens 3 and 6 were each presented 16 times in total, viz. 4 times per continuum. In the unimodal distribution, the number of occurrences for tokens 4 and 5 on the one hand and 2 and 7 on the other, were reversed: tokens 4 and 5 were presented most frequently and tokens 2 and 7 infrequently. The total number of tokens presented during the familiarization phase in both conditions was 136, see Figure 3.

Figure 3

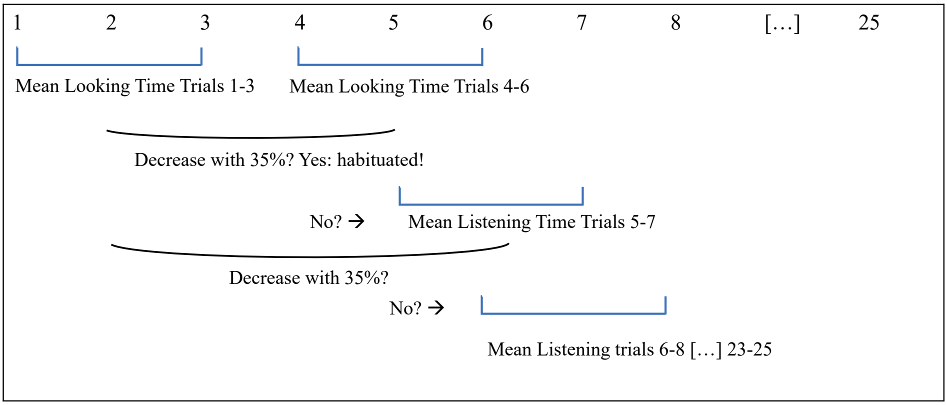
*The frequency distribution of the tokens of all four continua*




*Note.* Figure is adapted from Maye et al. (2008). Token 1 represents the first step, /sæn/, Token 8 the last step, /sɛn/.

***Habituation phase.*** The habituation phase started immediately after the familiarization phase. Infants were habituated to either tokens 3 or 6 (counterbalanced order) from all four speech sound continua. Tokens were presented randomly within the trial. Trials had a maximum duration of 45 seconds, with a maximum of 28 tokens per trial (ISI 1 second). Looking time to the screen was measured. The trial was terminated when an infant looked away for more than two seconds. Habituation criterion was reached if the infant's looking time to the screen had dropped by 35%, see Figure 4. The mean looking time of the first three trials was calculated and used to compare to three consecutive trials. If the mean looking time of the consecutive three trials (trial 4, 5 and 6) was less than 35% of the mean of trials 1-3, the habituation criterion was reached. If this was not the case, trial 1-3 was compared to trial 5-7, then 6-8, 7-9 and so on. Thus, habituation criterion was reached between 6-25 trials. Similar to the familiarization phase, infants heard the tokens (token 3 or token 6) of all the four continua within one trial, in random order. For instance, tokens number 3 from all four continua were presented randomly during each trial when the habituation token was 3.

Figure 4
*Visual depiction of the (65%) habituation criterion*




***Test phase.*** After the habituation criterion or the maximum of 25 habituation trials was reached, the test phase began and dishabituation (or change) trials were presented. These consisted of a new token (e.g., token 3 or 6, depending on the habituation stimulus). Stimulus repetition, ISI and maximum length were kept the same as in the habituation phase. The only change was the token: If an infant had been habituated on token 3, the dishabituation token was 6, and vice versa.

***Post-test.*** The visual stimulus was the same bull's eye as was displayed during the habituation and test phase. The auditory stimulus was a repetition of the pseudoword /bypoki/. Total duration of the post-test was also 45 seconds (ISI 1 second). The trial ended after the maximum duration was reached or when the infant looked away for two seconds or more. The post-test was introduced to measure general attentiveness. If looking times to the last two habituation trials and the dishabituation trials decreased and did not increase during the post-test, infant's data would not be used for analysis because it might indicate that infants are no longer attentive (Maye et al., 2008). This was not the case for any of the infants. Immediately after the post-test a Dutch commonly known children's song started playing. Looking times were not measured during the song.

## 5.3 Results

### Familiarization phase

The mean looking times for group and condition are presented in Table 4. Raw looking time data in the familiarization phase was normally distributed and was used as the dependent variable. There were no extreme (>2.5 *SD*) outliers. A univariate analysis of variance was used to analyze the data. No evidence was found for looking time differences between the groups (no-FR and FR), $F(1, 80) = .47$, $p = .494$, or between conditions (unimodal vs. bimodal), $F(1, 80) = .67$, $p = .416$, and there was no group by condition interaction effect $F(1, 80) = 1.3$, $p = .261$.

Table 4

*Mean looking times in the familiarization phase for each group (no-FR vs. FR) per condition (uni- vs bimodal)*

| Group | Condition | Participants | Mean looking time (seconds) familiarization phase | |
|---|---|---|---|---|
| | | *n* | *M* | *SD* |
| no-FR | Unimodal | 22 | 152 | 23.4 |
| FR | | 20 | 164 | 26.8 |
| no-FR | Bimodal | 21 | 155 | 27.3 |
| FR | | 21 | 153 | 34.1 |

**Habituation phase**

Mean total looking times during the habituation phase and number of trials infants needed to habituate are reported in Table 5. Looking times to habituation trials were not normally distributed; skewness = 2.575, *SE* = .261 and kurtosis = 10.823, *SE* = .517. The Log10 looking times approached a normal distribution, skewness = .402, *SE* = .263 and kurtosis = .803, *SE* = .520, and were entered in a univariate ANOVA to analyze habituation time per condition (unimodal, bimodal) and group (FR, no-FR). Although visual inspection of the data suggested that the FR-infants had longer mean looking times to the habituation phase in both conditions, there were no significant differences between the groups, $F(1, 80) = .49$, $p = .484$ and the conditions, $F(1, 80) = .90$, $p = .345$, and no significant interaction between group and condition, $F(1, 80) = .05$, $p = .823$.

As the number of trials that infants needed to habituate was not normally distributed; skewness = 2.469, *SE* = .261 and kurtosis = 6.868, *SE* = .517, a Kruskal-Wallis analysis was used to compare the outcomes. There were no differences between the groups $K(1, 84) = 1.07$, $p = .299$ and conditions $K(1, 84) = 2.99$ $p = 0.84$. There is no evidence that FR infants need more time or trials to habituate, and that condition had an effect on looking behaviour during habituation.

Table 5
*Number of habituation trials and, mean looking times to habituation trials*

| Group | Condition | Nr. of habituation trials | | Total habituation | | Last two habituation trials | | Dishabituation | |
|---|---|---|---|---|---|---|---|---|---|
| | | *M* | *SD* | *M* | *(SD)* | *M* | *(SD)* | *M* | *SD)* |
| no-FR | Unimodal | 7.1 | 1.5 | 74.6 | 26.4 | 5.5 | 2.6 | *5.7* | *2.8* |
| | Bimodal | 7.9 | 3.1 | 87.9 | 44.8 | 6.2 | 3.6 | *7.9* | *5.2* |
| FR | Unimodal | 6.3 | 0.8 | 80.6 | 30.3 | 6.7 | 4.3 | *7.2* | *4.1* |
| | Bimodal | 8.1 | 2.9 | 91.6 | 60.5 | 5.8 | 2.4 | *5.8* | *2.6* |

*Note.* The mean looking times to *Total habituation, Last two habituation trials* and the two *Dishabituation* trials are given in seconds.

**Test phase**
Three looking times to trials were not included in the analysis because they were 3 SDs above the mean looking times to habituation-dishabituation trials ($n$ = 1 in the unimodal condition, no-FR, $n$ = 2 in the bimodal condition, of which $n$ = 1 was FR and $n$ = 1 no-FR). Mean looking times to the last two habituation and the two dishabituation trials were used as the dependent variable. This measure was not normally distributed: skewness = 2.082, *SE* = .134; kurtosis = 6.848, *SE* = .268. After a log10 transformation looking times approached a normal distribution (skewness = .413, *SE* = .108 and kurtosis = .469, *SE* = .216). These transformed looking times were entered in a repeated measure ANOVA with factors trial type (habituation, dishabituation), condition (uni, bimodal), and group (FR, no-FR).
There was considerable looking time variation in both conditions and groups. Although the data suggest that the no-FR infants in the bimodal condition had the largest difference score between habituation and test (dishabituation) trials, the ANOVA did not yield a significant effect of trial type, $F(1, 81) = .99$, $p = .320$, $\eta^2 = .01$. No effects were found for condition (unimodal vs. bimodal) $F(1, 81) = .14$, $p = .714$, $\eta^2 = .002$ or group (no-FR vs. FR), $F(1, 81) = .02$, $p = .890$, $\eta^2 = .00$. There were also no significant interactions, Trial type *Condition, $F(1, 81) = .10$, $p = .756$, $\eta^2 = .01$, Trial type*Condition*Group, $F(1, 81) = .82$, $p = .369$, $\eta^2 = .01$. In summary, we found no evidence that 1) the frequency distribution of tokens during familiarization of the vowel continua (unimodal vs. bimodal) had an effect on subsequent discrimination performance and 2) that the groups performed differently.

**Data exploration**

The finding that the no-FR infants did not show significant differences between unimodal and bimodal learning is unexpected on the basis of previous studies on distributional learning in English-speaking infants (Maye et al., 2008; Yoshida et al., 2010) and Dutch-speaking infants (Capel, et al., 2011; Liu & Kager, 2014, 2017). As there were clear numerical differences between habituation-dishabituation trials in the no-FR group, and large individual differences, we looked into the data further.

We first studied the looking preference: the mean looking times of the last two habituation trials were subtracted from the mean looking times of the dishabituation trials. A mean difference above zero indicates a preference for dishabituation trials, which was the expected (novelty) preference. The findings in Table 5 show that the no-FR group had a mean difference of approximately 2 seconds in the bimodal condition and a mean difference approaching zero seconds in the unimodal condition. Table 6 shows that 75% of the no-FR infants in the bimodal condition looked longer to the dishabituation than the habituation trials, whereas this percentage is 50% in the unimodal condition. A Chi-squared test showed that the difference between conditions (unimodal vs. bimodal) in the numbers of infants with a preference for dishabituation trials approached significance ($\chi^2(1) = 2.78$, $p = .088$, one-sided). Taken together, these results are suggestive of an effect of condition.

No such tendency was found in the FR infants. In both conditions (unimodal and bimodal) the mean difference score is close to zero seconds (Table 5). Also, the percentage of FR infants in the bimodal condition looking longer to the mean dishabituation trials than the habituation trials is 57%, whereas the percentage is 55% in the unimodal condition. This difference is not significant, $\chi^2(1) = .029$, $p = .554$ (one-sided). These results suggest a subtle difference between the no-FR and FR infants: whereas the no-FR group shows a marginally significant difference between the percentages of infants that look longer to dishabituation than to habituation trials in the bimodal but not the unimodal condition, no such pattern is present for the FR infants.

Table 6
*Number of infants looking longer to the dishabituation trials*

| Group | Condition | Participants | Infants looking longer to dishabituation trials | |
|---|---|---|---|---|
| | | *N* | *N* | % |
| no-FR | Unimodal | 22 | 11 | 50 |
| no-FR | Bimodal | 20 | 15 | 75 |
| FR | Unimodal | 22 | 12 | 55 |
| FR | Bimodal | 21 | 12 | 57 |

Additionally, we evaluated the role of looking time during familiarization. The total time that infants spend looking during familiarization trials might affect performance during the test phase. Arguably, longer looking times could signify better learning, as looking time to screen is assumed to reflect listening (Aslin, 2007), which is a prerequisite for learning. As reported (Results, *Familiarization phase*), there were no significant differences between the groups (FR, no-FR) in mean looking times to familiarization phase. To explore the effect of looking time during the familiarization phase, we used the median familiarization looking time (~ 155 seconds, mean familiarization looking time was also ~ 155 seconds) as a cut-off value between short and long lookers (Colombo et al., 1991). Table 7 presents the number of long and short-looking participants per condition and group, and Figure 5 the mean looking times to the habituation and dishabituation trials for long and short lookers. A repeated measures ANOVA with trial type (habituation vs. dishabituation trials) as the within-subject factor, and condition (unimodal vs. bimodal), group (FR vs. no-FR) and looking time familiarization group (short vs. long lookers) as between-subject factors returned a main effect of looking time familiarization group, $F(1, 77) = 9.42$, $p = .003$, $\eta^2 = .11$. This reflects the median split: children identified as long lookers had significantly longer looking times. Of more interest is the finding that the interaction between trial type and looking time familiarization group was marginally significant, $F(1, 77) = 3.34$, $p = .071$, $\eta^2 = .04$, indicating that the effect of trial type on looking time is not entirely the same for short and long lookers. Post-hoc pairwise comparisons (Bonferroni corrected) showed that only the long looker no-FR infants in the bimodal condition discriminated the contrast when they were in the bimodal condition, not in the unimodal condition.

Table 7
*Mean looking time to familiarization phase for short and long looking infants*

| Looking time familiarization group | Group | Condition | Nr. of participants | Mean looking time to familiarization | |
|---|---|---|---|---|---|
| | | | *n* = | *M* | *SD* |
| Short Lookers | no-FR | Unimodal | 11 /22 | 134 | 15.4 |
| | FR | | 12 /22 | 137 | 13.1 |
| | no-FR | Bimodal | 8 /20 | 138 | 15.5 |
| | FR | | 11 /21 | 126 | 21.8 |
| Long Lookers | no-FR | Unimodal | 11 /22 | 171 | 12.9 |
| | FR | | 10 /22 | 180 | 19.5 |
| | no-FR | Bimodal | 12 /20 | 182 | 15.9 |
| | FR | | 10 /21 | 177 | 16.5 |

## Discussion

We investigated whether Dutch 8-month-old infants with a family risk (FR) of dyslexia and their no-FR peers discriminate the non-native English /ɛ/ - /æ/ vowel contrast after having been exposed to this vowel contrast in a unimodal or bimodal frequency distribution. The literature has indicated that no-FR infants show distributional learning: They will discriminate a non-native contrast after exposure to a stream of speech sounds spanning the relevant acoustic continuum that has a bimodal frequency distribution, such that sounds close to the continuum endpoints are more frequently presented than those in the middle. In contrast, they will not be sensitive to the non-native contrast after exposure to a unimodal distribution, in which sounds in the centre of the continuum are more frequently presented than those towards the endpoints (Capel et al., 2011; Liu & Kager, 2014, 2017; Maye et al., 2008; ter Schure et al., 2016; Wanrooij et al., 2014; Yoshida et al., 2010).

We expected that FR infants would not be able to discriminate the vowels, regardless of the condition (unimodal, bimodal), as there are indications that dyslexia is associated with a reduced capacity for distributional (statistical) learning (e.g., Lum et al., 2013). Reduced statistical learning may underlie the phonological deficit that is considered a core characteristic of dyslexia (e.g., Vellutino et al., 2004). Indeed, Vandermosten et al. (2019) found that Grade-3 children with dyslexia were not responsive to a bimodal speech sound distribution in a similar paradigm. The evidence reported here does not support (or refute) the prediction that no-FR infants discriminate the English /ɛ/ - /æ/ contrast after bimodal, but not unimodal, exposure. In fact, we did not find evidence for discrimination in either condition. Moreover, no differences were found between the no-FR and FR group.

In order to interpret the findings, it needs to be evaluated whether design issues could have affected the no-FR findings. This seems unlikely, as other studies from our lab, employing the same design, replicated the findings by Maye and colleagues (2008): significant effects of condition have been found for a consonantal (Capel, et al., 2011) and a tonal contrast (Liu & Kager, 2014, 2017). A second possibility is that a vowel contrast renders different findings than a consonant contrast. However, previous studies have found evidence for distributional learning in no-FR infants using the same vowel contrast as we did (ter Schure et al., 2016; Wanrooij et al., 2014). Moreover, in a meta-analysis on distributional learning, Cristià (2018) established that type of contrast (consonant, vowel or tone) was not a moderator that explained the variance in studies using Maye et al.'s (2008) design. Hence, we conclude that design and vowel contrast cannot account for the unexpected findings of the no-FR group.

Further exploration of the data provided subtle indications that performance of the no-FR group does point in the predicted direction, as no-FR infants, on average, looked longer to the dishabituation trials in the bimodal condition, but not in the unimodal condition. Another indication is that the percentage of (no-FR) infants who had a preference for the dishabituation trials (in agreement with our prediction), was higher in the bimodal condition (75%) than in the unimodal condition (50%). This observation mirrors findings by Capel et al. (2011), who investigated distributional learning of a non-native Hindi consonantal retroflex /ɖa/ - /ʈa/ contrast in Dutch 10-month-old infants. They also found a higher percentage of infants that looked longer to the dishabituation trials in the bimodal (63%) than unimodal (48%) condition. Hence, there are no statistically significant findings concerning the distributional learning outcomes, but the quantitative pattern in our data is similar to that of previous studies.

Following leads from Colombo et al. (1991), we explored whether looking time during familiarization had an effect on performance during test. When grouping infants as short or long lookers during familiarization, we found that the no-FR long lookers discriminated the contrast in the bimodal condition. In contrast, we did not see evidence of discrimination in the no-FR short lookers. Taken together, our exploratory results very tentatively suggest that there are subtle indications of distributional learning in no-FR infants. However, more research is needed to confirm this suggestion, taking looking times to familiarization phase into account.

The exploratory analyses did not provide any indication of a difference between discrimination outcomes after bimodal or unimodal exposure in the FR infants. The FR infants did not show longer average looking times to the dishabituation trials in either condition (mean difference approaching zero) and there was no difference between percentage of infants that looked longer to dishabituation trials in the bimodal (57%) and unimodal (55%) condition or in performance between short and long lookers. Thus, the data tentatively suggest that FR infants, in comparison with no-FR infants, are less responsive to the differences in distribution. However, as we did not find a significant difference for the no-FR infants, nor any significant interactions, no firm conclusions can be drawn.

Our findings show a mixed picture. On the one hand, there is an absence of a reliable group (FR, no-FR) by condition (unimodal, bimodal) interaction in our confirmatory statistical analysis. On the other hand, our exploratory results could point towards poorer distributional learning by FR infants. These inconclusive findings are in line with other studies on statistical learning in children and adults with dyslexia. For instance, van Witteloostuijn et al. (2019) found that children with dyslexia did not differ from those without dyslexia on distributional learning of grammar, and a meta-analysis by van Witteloostuijn et al. (2017) showed that distributional learning differences between children with and without dyslexia might be inflated because of publication bias (similar to Schmalz et al., 2017). Our findings may thus be taken to speak to the mixed findings in the literature on distributional learning in dyslexia.

In summary, our results do not permit a robust conclusion concerning FR infants' sensitivity to speech sound distributions and its effect on their discrimination. Nevertheless, our data provide some indications of a difference between the groups concerning the sensitivity to distributional learning, as the no-FR infants show better outcomes for the bimodal condition than the unimodal condition, while there are no such signs for the FR group. These findings can thus be used as a stepping stone to further investigations into the putative distributional learning deficit and its association with a phonological deficit in dyslexia.

**Acknowledgements**

This study was funded by The Netherlands Organization for Scientific Research (NWO, grant nr. 360-70-270). We thank the parents of our participants for their contributions to our study. We thank Charlotte Koevoets and Lisanne Geurts for assistance in data collection and Annemarie Kerkhoff for assistance in the design of the study. We also thank Willemijn Heeren for sharing her stimuli with us.

**Ethics statement**

Informed consent was obtained from the caregiver before testing and the caregiver was allowed to retract this consent and participation any time during testing. The authors declare that the research was conducted in accordance with APA ethical standards as well as *The Netherlands Code of Code of Conduct for Scientific Practice* issued in 2004 (revised in 2018) by the Association of Universities in the Netherlands (VSNU).

Chapter 6

# General discussion and conclusion

## 6.1 Introduction

The topic of this dissertation is the construction of speech sound categories - specifically vowels - in typically developing infants and infants at family risk of dyslexia (FR). According to a broadly supported hypothesis, a core feature of developmental dyslexia is a phonological deficit, which is argued to stem from poorly developed phonological representations (Ramus et al., 2003; Vellutino et al., 2004). The development of speech sound categories is required for constructing phonological representations. The work reported in this dissertation seeks to establish if and how the emergence of speech sound categories differs between typically developing and FR infants. This final Chapter presents the main findings of this dissertation and answers the research questions presented in Chapter 1.

The four experimental studies had different but related research questions. The main research question of **Chapter 2** was whether Dutch infants' vowel discrimination performance followed the developmental trajectory predicted by the theory of perceptual attunement. In **Chapter 3**, the main question was whether it could reliably be assessed which infants could and could not be classified as discriminators of a native vowel contrast. **Chapter 4** focused on the question whether FR infants show evidence of perceptual attunement and whether their developmental pattern differed from the pattern of their peers. Finally, **Chapter 5** evaluated whether infants with and without an FR of dyslexia were able to use phonetic distributional information in the input to acquire speech sound categories.

The present chapter is organized as follows. First, the rationale and findings for each of the four experimental studies are summarized (**6.2**). In **Section 6.3**, implications for the theory of perceptual attunement (6.3.1) and hypotheses of distributional learning (6.3.2) are discussed on the basis of the data of the typically developing infants. This is followed by a discussion of the findings concerning infants at family risk of dyslexia (**Section 6.4**), a section commencing with a discussion of the development of vowel categories in light of perceptual attunement and the phonological deficit hypothesis of dyslexia. (6.4.1). Next, the implications of the findings on distributional learning will be discussed in relation to the hypotheses of distributional learning and the phonological deficit (6.4.2). Future directions of research are provided at the end of each section. Limitations of the current dissertation are discussed in **Section 6.5**. The conclusion of this dissertation is offered in **Section 6.6**.

## 6.2 Rationale and summary of findings per chapter

In **Chapter 2** we investigated the developmental trajectory of native and non-native vowel discrimination in typically developing infants during the second half of the first year of life. Previous research demonstrated that infants develop from universal listeners to language-specific listeners. This perceptual transition is often referred to as *perceptual attunement* (Maurer & Werker, 2014) and is considered to be the first step into the formation of speech sound categories (Kuhl, 2004). Generally, it is found that infants are initially sensitive to all speech sound contrasts and that this sensitivity declines for those contrasts that are not relevant in their native language. Infants' ability to discriminate native contrasts remains throughout development and in some cases, discrimination of these contrasts even improves (Maurer & Werker, 2014). However, there is evidence that perceptual attunement is less straightforward than described above. For instance, some non-salient native contrasts are not discriminated from birth and not all studies find a decline in non-native consonantal and vowel contrasts (e.g., Best & Faber, 2000; Liu & Kager, 2016; Mazuka et al., 2014; Narayan et al., 2010; Polka & Bohn, 1996; Tyler, Best, Goldstein, & Antoniou, 2014). Hence, the literature is inconclusive regarding the development of non-salient native and non-native discrimination. Furthermore, so far, the developmental trajectory of vowel discrimination has received less attention (Tsuji & Cristia, 2014). This dissertation extends the literature by providing cross-linguistic evidence on perceptual attunement in vowels, as Dutch learning infants were assessed on native and non-native vowel discrimination during the second half of their first year of life.

To shed new light on this issue, Chapter 2 reports a study in which we investigated native and non-native vowel discrimination in 6-, 8-, and 10-month-old Dutch learning infants. The first main goal was to establish whether infants would show evidence of (improving) discrimination of the Dutch (/aː/ - /eː/) vowel contrast. The second question was whether typically developing infants would show decreasing discrimination performance of the non-native English /ɛ/ - /æ/ vowel contrast with increasing age, in line with predictions of perceptual attunement. The results confirmed the predictions for the native contrast, as infants of all age groups discriminated the native contrast. The increase in effect size was taken to refer to improved discrimination. However, the developmental pattern for the non-native contrast was not as expected: the 6- and 10-month-old infants discriminated the contrast, whereas the 8-month-olds did not. Thus, the decline in perception between the 6- and 8-month-old infants agrees with the perceptual attunement hypothesis, but the discrimination performance of the 10-month-olds does not. We discussed the possibility that the 10-month-olds were able to discriminate the non-native contrast because they were better able than the 8-month-olds to make use of the speaker variability that they were exposed to during the habituation phase. Speaker variability has been argued to enhance generalization of abstract features in the process of developing phonetic categories (Lively et al., 1993; Potter & Saffran, 2017; Rost & McMurray, 2009). However, there is evidence that the amount of variation needed in order to be helpful during a task differs between age groups (Estes & Lew-Williams, 2015; Singh et al., 2004; Singh, 2008; Vukatana et al., 2015). The main conclusion of Chapter 2 was that the ability to discriminate a non-salient non-native contrast is not lost after the process of perceptual attunement has emerged and that the outcomes depend on the applied experimental design, the age of the infants and the salience of the contrast.

In **Chapter 3** we investigated whether we could determine individual discrimination performance. Recently, there has been a growing interest in explaining individual differences in infants' speech sound discrimination skills (Cristia et al., 2014; Houston et al., 2007). This interest is sparked by attempts to connect early individual differences to later speech and language outcomes. A meta-analysis of Cristia et al. (2014) found evidence that early speech perception skills predict later language skills. A frequently used approach to (retro- or prospectively) predict infants' discrimination performance, is to use follow-up data, such as later vocabulary size, reading scores or other skills (e.g., Molfese, 2000; van Zuijen et al., 2013). However, these results are based on group-level analyses and cannot predict individual trajectories of language development. If individual discrimination trajectories of speech sound contrasts can be determined, this outcome could be used to assess speech sound development longitudinally and it can be used prospectively to predict other language and literacy outcomes. Houston and colleagues (2007) designed a discrimination paradigm to assess individual discrimination at the word level. The approach of Chapter 3 was twofold: we replicated the statistical model used by Houston and colleagues and we applied Bayesian hierarchical modeling to assess individual performance, as Bayesian hierarchical modeling is able to combine group information (hierarchical modeling) and autoregressive error structure (error of the preceding trial), without the need to correct for multiple testing.

In Chapter 3 we looked at individual performance of the same infants as reported on in Chapter 2 (aged 6, 8 and 10 months), but here we only looked at the data from the native contrast. The main goal was to establish whether we could reliably classify infants as discriminators of this native contrast. As the native contrast is a salient contrast, expectations were that we would find a high percentage of infants that could be classified as a discriminator, comparable to the 80% Houston and colleagues (2007) found in their study. However, the frequentist analysis approach established that 12% of the infants discriminated the contrast and after correction for multiple testing, merely 3%. This percentage seems at odds with previous findings and expectations. The Bayesian results showed that on average 50% of the infants discriminated the contrast: 53% of the 6-month-olds, 27% of the 8-month-olds and 77% of the 10-month-olds. Some explanations were provided for the lower percentage of discriminators in our sample compared to that of the 80% of the 9-month-olds in the study of Houston et al. (2007). The first was that the design with 12 test trials might have been less suitable for the younger infants, as it might have asked too much of their attention span. Second, Houston et al. used a word contrast, *boodup - seepug*, which differs markedly from the phonemic contrast /faːp - feːp/ we used. A word contrast is more salient than a phoneme contrast, which could explain the difference in percentages between the two studies. Third, we used 12 test trials whereas Houston et al., used 14. This might have caused a lower mean looking time to non-alternating trials, as infants' internal representation of the old (non-alternating) stimulus might become stronger during test, which is expected to result in a larger increase in looking time to new stimuli (Sokolov, 1963). Finally, Houston and colleagues showed video clips of females pronouncing the targets, whereas we used stills/images of females as visual support. The clips might have helped infants to robustly discriminate the contrast (see also ter Schure et al., 2016). The main conclusion of Chapter 3 was that Bayesian hierarchical modeling is preferred over a frequentist approach in assessing individual discrimination performance and that it can be used to classify infants as discriminators or non-discriminators.

Speech perception of infants at family risk of dyslexia (FR) was at the core of **Chapter 4**. We investigated whether FR infants and their no-risk peers (no-FR) differed in discrimination skills of a native and non-native vowel contrast. Numerous studies demonstrated that individuals with dyslexia have phonological difficulties (e.g., Ramus et al., 2003; van Bergen et al., 2012). The phonological deficit hypothesis asserts that a deficiency in forming representations of speech sounds is a causal factor in the emergence of dyslexia (e.g., Ramus et al., 2003; Vellutino et al., 2004). This deficiency may be founded on (innate) speech perception difficulties. There are roughly two accounts of how poorly developed phonological representations result from a speech perception deficit. The first is that individuals with (an FR of) dyslexia have poor speech discrimination skills. Adults and children with dyslexia perform worse on tasks assessing discrimination and classification of speech sounds (e.g., Maassen et al., 2001; Mody et al., 1997). This pattern has also been attested in FR children and infants (e.g., Boets et al., 2007; Guttorm, et al., 2005; van Leeuwen, et al., 2006; Richardson et al., 2003). Alternatively, the ‘allophonic speech perception’ hypothesis holds that adults and children with dyslexia and FR children show heightened and prolonged sensitivity to phonetic variation that is not relevant for native speech perception (e.g., Noordenbos et al., 2012; Serniclaes et al., 2004). The literature is thus inconclusive when it comes to predicting the developmental path of the construction of speech sound categories, as both poor discrimination and enhanced (allophonic) discrimination have been found.

The central question of Chapter 4 was whether FR infants have a delayed speech perception development. Similar to Chapter 2, the 6-, 8-, and 10-month-old infants were tested on a native Dutch (/a:/ - /e:/) vowel contrast a non-native English /ɛ/ - /æ/ contrast. On the basis of the speech perception deficit view, predictions for the FR infants were that they would discriminate the salient native contrast but would show poor perception at all ages for the non-native contrast, in line with a large body of research showing that infants have poor discrimination skills (e.g., Richardson, 2003; van Leeuwen et al., 2006). Based on the view of allophonic speech perception in dyslexia, it could be predicted that FR infants would show heightened sensitivity to the non-native contrast (e.g., Noordenbos et al., 2012). Results showed that no-FR and FR infants discriminated the native contrast at all ages. This was evidenced in both group and individual-based analyses. However, there was a subtle indication of a difference between the two groups: whereas improving performance was evidenced in the no-FR group, this was not the case for the FR infants. Findings of the *non-native* contrast showed a more pronounced difference between the no-FR and FR infants, as the FR infants showed no evidence of discrimination of the non-native contrast at any of the ages, whereas the no-FR 6- and 10-month-old infants did. Together, these findings indicate that FR infants do not (yet) show an increase in discrimination performance of the native contrast, which the no-FR infants did. Furthermore, the youngest group of FR infants did not show evidence of discriminating the non-native vowel contrast. The individual-based analysis even showed that zero percent of the FR infants discriminated the contrast. Based on these findings we concluded that FR infants have a subtle delay in their development speech sound categories. This interpretation is in line with the literature that found that individuals with dyslexia have poorer discrimination skills. Our findings do not support allophonic speech perception in our sample of FR infants.

**Chapter 5** approached the acquisition of speech sound categories from a different angle. The focus was not on the developmental pattern of vowel discrimination, but on a learning mechanism that might be responsible for the perceptual shift generally found in infants during their second half of the first year of life. Distributional phonetic learning has been put forward as an underlying learning mechanism that assists the perceptual reorganization or perceptual attunement in infants (e.g., Kuhl, 2004). There is evidence that typically developing infants are able to track the frequencies of occurrences of speech sound tokens (vowels, consonants and tones), in order to learn their native categories (Cristia, 2018; Maye et al., 2002, 2008). Whether FR infants show the same sensitivity to distributional information in the input is still an open question, as the literature on statistical and distributional learning is inconclusive, showing both poorer learning in samples with (a risk of) dyslexia (Banai & Ahissar, 2018; Bonte et al., 2007; Kerkhoff et al., 2013; Vandermosten et al., 2019) as well as absence of differences (e.g.; van Witteloostuijn et al., 2021) and indications of a publication bias for studies reporting group differences (Schmalz et al., 2017; Witteloostuijn et al., 2017). Assessing distributional learning of a phonetic contrast in FR infants is needed to establish whether indications of distributional learning can be found at an early age.

We investigated whether 8-month-old infants could discriminate an English non-native /ɛ/ - /æ/ vowel contrast after being exposed to these vowels in different distributions. During the familiarization phase infants were exposed to speech sound tokens taken from four English /ɛ/ -/æ/ 8-step continua, either in a unimodal distribution, which should lead to perceiving the continuum as one vowel category, or a bimodal frequency distribution, which should lead to perceiving the continuum as reflecting two vowel categories, namely /ɛ/ and /æ/. The main finding was that neither the no-FR nor the FR group showed evidence for discrimination in either condition (uni- vs. bimodal). Additional exploratory analyses yielded subtle indications that for no-FR infants, the time spent looking to the screen during the learning phase of the experiment was related to speech sound discrimination in the bimodal condition: Infants who looked longer were able to discriminate the speech sounds in this condition. Such results were not found for the no-FR longer lookers in the unimodal condition. These subtle findings agree with expectations of distributional learning. This pattern was not present in the data of the FR infants. Hence, our tentative conclusion was that there are differences between the groups.

### 6.3 Implications for theories on phonetic learning in typically developing infants

In this section, findings on vowel perception of the typically developing infants are positioned within the framework of perceptual attunement (6.3.1) and distributional learning (6.3.2). Each section features the implications of the findings as well as suggestions for future research.

*6.3.1 Perceptual attunement in typically developing infants*

The developmental trajectories of vowel discrimination performance of typically developing infants were assessed (Chapter 2 and 3). The findings of Chapter 2 showed that the ability to discriminate a non-salient non-native contrast is not lost after the process of perceptual attunement has emerged, and hence, they suggest that the developmental discrimination trajectories fluctuate and are not monotonic. These findings thus do not agree with the perceptual attunement framework. Our findings agree with other studies which do not show straightforward support for the perceptual attunement account. Some do not find evidence of a decline in sensitivity to non-native speech sounds (e.g., Best & Faber, 2000; Polka & Bohn, 1996; Mazuka et al., 2014; Tyler, Best, Goldstein and Antoniou, 2014). Others find that very young infants do not show initial sensitivity to (native) contrasts (Liu & Kager, 2016; Narayan et al., 2010).

Similar to our findings, results of other studies thus also disagree with perceptual attunement. For instance, Narayan, Werker and Beddor (2010) demonstrated that English 4-5-month-old infants were not able to distinguish a Filipino /na/ - /ŋa/ contrast in a visual habituation paradigm, whereas their 6-8-month-old peers were able to do so. Narayan et al. suggested that language experience was needed to acquire the specific contrast. Sundara et al. (2018) presented counterevidence to the findings of Narayan et al. (2010) as they showed that English learning 4- and 6-month-old infants could distinguish the Filipino /na/ - /ŋa/ contrast, when using a more sensitive design. Sundara et al.'s design used an infant-controlled paradigm, whereas Narayan et al. used continuous exposure. Additionally, Sundara et al. used longer (19 vs. 14 seconds) and more (15 vs. 9) habituation trials and they shortened the inter-stimulus-interval (ISI, 800 vs 1000 ms). The effect of shorter ISI duration poses fewer demands on these young infants' short-term memory (Pelphrey et al., 2004), which might enhance discrimination performance and might (partly) explain the difference in findings between the studies of Narayan et al. (2010) and Sundara et al. (2018). However, which specific design elements (infant controlled, ISI duration, longer and more habituation trials) contributed (most) to the positive finding of Sundara et al (2018) is unclear. For instance, Liu & Kager (2016) did not find initial sensitivity to a native non-salient Dutch /I/ - /i/ contrast, comparable to Narayan et al.'s 2010 study. Liu & Kager (2016) used an infant-controlled visual habituation paradigm, as Sundara et al. (2018) did. However, they used 1000 ms. ISI, as Narayan et al. (2010). What these examples illustrate is that it is difficult to make straightforward predictions on discrimination and attunement, as performance depends on the experimental design, the age of the infants and the salience of the contrast. Future research should focus on these factors and how they interact.

The findings of Chapter 2 showed that towards the end of the second half of the first year of life, the perceptual system of infants is still flexible. An assumption with regard to perceptual attunement is that the period during which infants tune into their native language is within the time window of the first year of life. Werker and colleagues framed this time window as a critical (or sensitive) period for phonetic perception (Werker & Hensch, 2015; Reh et al., 2020). During this critical period the brain is most sensitive to the (distributional) phonetic distinctions of the ambient speech. This sensitivity starts around 2 months of age, is highest around 6-8-months of age and ends around 12 months of age. While the result that the 10-month-old infants discriminated the non-native English /ɛ/ - /æ/ contrast (Chapter 2) is not counterevidence to this critical period per se, it does show that towards the end of this time window infants are still very flexible in their perception. It seems that some cues can activate the ability to distinguish a non-native contrast. One such cue could be speaker variability, as we argued was the case in our study (Chapter 2). Another could be social interaction, such as joint attention (Kuhl et al., 2003). Semantic cues (Sing & Tan, 2020; Yeung & Werker, 2009) could also impact on discrimination. For instance, Sing and Tan (2020) showed that 14-month-old English infants could not discriminate a Hindi dental/retroflex voiceless stop contrast. However, during a switch-paradigm they were able to map nonwords containing this same non-native contrast to new objects. Hence, they could use this non-native phonetic contrast to learn new words. Based on these findings, as well as all the studies that find sensitivity to contrasts where it was not expected (e.g., Best & Faber, 2000; Mazuka et al., 2014; Polka & Bohn, 1996; Tyler, Best, Goldstein & Antoniou, 2014), it could be argued that if there is a critical period for learning speech sound categories, this period exceeds the end of the first year of life and is more dynamic/flexible than sometimes assumed.

In Chapter 3 (and 4) the question at stake was whether we could assess individual discrimination performance in infants. Bayesian hierarchical modeling proved to be a promising avenue as it provided insights in the (cross-sectional) developmental trajectories that did not surface as clearly in the group-based analysis. For example, the 8-month-old no-FR infants showed an unexpected low percentage of infants that could be classified as a discriminator. This finding is due to the larger variance at group level at 8 months. We suggested that the heterogeneous performance at this age reflects a developmental (reorganizational) shift in vowel perception and does not reflect a loss of discrimination (Werker et al., 2004). During the development of speech perception, infants learn to listen from a new perspective. Whereas infants react to all perceivable phonetic differences at birth and can potentially discriminate many speech sounds, they begin to learn to categorize speech sounds in phonemic units during the reorganizational phase (Kuhl, 2004). As not all infants begin their shift at exactly the same timepoint; this could explain the heterogeneity of group performance at 8 months. Differences between infants may arise from differences in domain-general cognitive abilities such as inhibitory control (e.g., Conboy et al., 2008; Lalonde & Werker, 1995), or the ability to pick up distributional information (Maye et al., 2008).

**Taken together,** the findings of Chapters 2 and 3 showed that the developmental trajectory of speech sound category formation is less robust and monotonic than is often assumed. Monotonic predictions following from perceptual attunement are: An (increase) in the ability to discriminate native speech sounds as a result of language exposure and a decrease in the ability to discriminate speech sounds when there is no exposure to these sounds. However, the finding that the 10-month-olds distinguish a (subtle) non-native contrast (Chapter 2), indicates that it is difficult to make straightforward predictions, as the perceptual system is still very sensitive and flexible. Moreover, the individual discrimination patterns of the native contrast (Chapter 3) also suggest that the developmental trajectory is not monotonic, as the percentage of the 8-month-olds that discriminated the contrast was lower than expected. Finally, the findings of Chapter 2 raise a question about the proposed time window during which perceptual attunement takes place. This time window may not be closing around 12 months of age but somewhat later.

**Future research** on developmental trajectories in speech sound discrimination will have to investigate how experimental factors, such as the (sensitivity of the) design, the age of the infants and the acoustic properties of the contrast, interact. Additionally, future research should focus on child factors such as sustained attention, inhibitory control and how they relate to perceptual attunement, as this could shed light on how individual variation in the emergence of the developmental shift can be accounted for. Another important step for future research is to use longitudinal data to establish whether the fluctuating patterns of performance attested in this dissertation are evidenced as well. With regard to individual assessment of discrimination performance, future studies should focus on how to include group results in the estimation procedure for a single individual. The advantage of Bayesian hierarchical modeling is that prior information can be used to update the models. In case of the data presented in Chapter 3 and 4, data of new individuals that belong to the same group (e.g., 6-month-old FR infants) can be added to the analyses, thereby updating the information for all the individuals as the added data enriches the knowledge already at hand. For example, the data of a newly tested individual can be analyzed in the context of the population he/she belongs to, thereby adjusting the estimates. How to include group results in the estimation procedure for a single individual is the topic of a paper in progress by my colleague Duco Veen. The major advantage of the option to add new data to the knowledge (prior) already available, is that fewer infants well need to be tested in a future study and that the confidence in the parameter estimates (e.g., regarding a condition effect) will increase.

*6.3.2 Distributional learning of a vowel contrast*

In Chapter 5 we looked at differences between typically developing 8-month-old infants (no-FR) and FR infants. Implications for distributional learning in typically developing infants will be discussed here (see 6.4.2 for implications for FR infants). First and foremost, it should be emphasized that the finding that the no-FR infants do not show evidence of discriminating the English vowel contrast in either condition (uni- and bimodal), is a null result. The finding is therefore inconclusive about distributional learning taking place in the no-FR infants. Clearly, such a result is difficult to interpret. It could be that the finding is a true negative (distributional learning does not take place). This requires explanations as to why learning was not attested. Alternatively, the obtained null result is a false negative; the failure to find effects indicative of distributional learning may be due to a power issue. To start with the latter: the absence of significant effects in Chapter 5 could be due to the relatively large individual variance in the bimodal condition, in association with a small group size. In this view we would have had too little power to demonstrate a potential effect. The meta-analysis on distributional learning by Cristia (2018) demonstrated that each condition (uni- and bimodal) should have at least $n$ = 35 infants. Since our study included two different experimental groups (no-FR and FR), we should have included at least $n$ = 70 in each condition (e.g., $n$ = 35 in unimodal FR and $n$ = 35 unimodal no-FR), instead of our n = 44 (unimodal FR and no-FR together). Hence, considering the numerical findings, it is possible that our null result is a false negative due to a power issue, especially since we were looking for an interaction with group (FR and no-FR) and condition (uni- and bimodal).

If the result is a true null result, the question is why distributional learning did not take place. Previous research has found evidence of distributional learning in vowel perception (Wanrooij et al. 2014; ter Schure et al., 2016). Moreover, the meta-analysis by Cristia (2018) showed that overall, the evidence attests that infants' speech sound categorization is modulated by prior exposure to different distributions, and that type of contrast (consonant, vowel or tone) was not a moderator that explained the variance in studies using Maye et al.'s (2008) design. However, ter Schure et al (2016) showed that 8-month-old infants learned the English /ɛ/ - /æ/ vowel contrast only if the distribution of the speech sounds during familiarization was two-peaked (bimodal distribution) and if the sounds were supported by visual cues. In this auditory-visual condition, infants saw a screen with a female speaker pronouncing the nonwords. In the auditory-only condition, infants watched the same female speaker pronouncing the nonwords, but her mouth was hidden behind her hand and hence the articulatory movements were not visible. Infants in the auditory-only or visual-only condition (no auditory input), did not learn the contrast, regardless the distribution (uni- or bimodal) of speech sounds. The results of ter Schure et al (2016) indicate that auditory distributional information in the input alone might not be enough for Dutch 8-month-old infants to acquire (non-salient) vowel contrasts.

The hypothesis that infants need more information or (non-auditory) cues to discriminate a new (non-native) contrast is strengthened by the idea that the auditory distributional information that infants receive during the exposure, is manipulated only on two acoustic values that do not contain overlapping values. Real language exposure is much noisier: F1 and F2 show considerable overlap and variability, also in infant-directed speech (e.g., Swingley, 2009). The question how infants can use distributional information to acquire vowel contrast, based on F1 and F2 values that acoustically overlap, is therefore very relevant and has been the topic of investigation (e.g., Adriaans & Swingley, 2017, Feldman et al., 2013; Swingley, 2009). These studies investigate what other speech cues might facilitate distributional learning of vowel contrasts. For instance, computational models showed that exaggerated prosodic cues (a marker of infant-directed-speech) improved distributional learning of vowel categories (Adriaans & Swingley, 2017).

Assuming our result in Chapter 5 is truly a null result, one issue could be the participating infants' age in relation to perceptual attunement. There are indications that distributional learning is most effective during the period of perceptual attunement (Yoshida et al., 2010; Reh et al., 2021). For vowels, perceptual attunement is thought to emerge around 6-8 months of age (e.g., Kuhl et al., 1992). For consonantal contrasts this is around 10-12 months of age (e.g., Werker & Tees, 1984). Maye and colleagues (2008) tested 8-month-old infants on the consonantal unaspirated dental voicing /da/ - /ta/ contrast and found evidence for distributional learning as infants discriminated the contrast in the bimodal, but not in the unimodal condition. A study by Yoshida and colleagues (2010) showed that 10-11-month-olds failed to discriminate the same contrast, regardless of condition (uni- vs. bimodal). However, when the familiarization phase was doubled in duration infants were able to discriminate the contrast in the bimodal but not the unimodal condition. Yoshida and colleagues concluded that infants are most sensitive to phonetic distributions during the process of perceptual attunement (see also Reh et al., 2021). Apparently, Yoshida et al.'s participants had passed the phase of consonantal attunement, while Maye et al.'s had not.

If it is true that distributional learning of phonetic contrasts is strongest during the sensitive period, infants in our study (at 8 months of age) might have passed through the phase of attunement for vowels. These 8-month-olds might have needed more exposure during familiarization, similar to the 10-month-olds in Yoshida et al.'s study. However, Liu and Kager (2017) found that 11-month-olds, but not 5- and 14-month-olds were able to use distributional information to distinguish a tonal contrast. The 14-month-olds did not show evidence of discriminating the contrast in either condition. The opposite was seen for the 5-month-olds as they discriminated the contrast in both conditions. Perceptual attunement for tonal contrasts is found to emerge earlier than for vowel and consonantal contrasts. If sensitivity to distributional information is indeed strongest during the period of perceptual attunement, the 11-month-olds in Liu & Kager's study should not have shown evidence of discriminating the contrast in the bimodal condition and the 5-month-olds only in the bimodal but not the unimodal condition. Moreover, the meta-analysis performed by Cristia (2018), showed that while age is a moderator of distributional learning, the direction of the effect is the opposite of what is expected on the basis of the Yoshida et al.'s (2010) finding. The effect sizes increased as aged increased, which means infants behaved more coherently as they were older. This suggests that older infants are better at using the distributional information. Hence, the literature is inconclusive when it comes to the effect of age and the amount of exposure, on distributional learning (Cristia, 2018; Liu & Kager, 2017; Reh at al., 2021; Yoshida et al, 2010). Taken together, no firm conclusions can be drawn from the findings on distributional learning. Either the finding is a false negative, possibly due to power issues, or the effect is really absent and distributional learning did not take place. **Future studies** should investigate the interaction between age and distributional learning, as the literature is inconclusive about the interaction between age and the amount of exposure that is sufficient to learn a new contrast based on distributional information. Also, the questions which cues facilitate distributional learning in infants (e.g., exaggerated speech, visual input, speaker variability, consonantal context) and whether distributional learning is equally effective in vowel acquisition as it is evidenced in consonantal category acquisition should be the topic of future research.

## 6.4 Implications for phonetic learning in infants at family risk of dyslexia

In this section findings on vowel discrimination of FR infants are discussed within the framework of the phonological deficit and allophonic speech perception, followed by suggestions for future research (6.4.1). In section 6.4.2 results on distributional learning in FR infants are discussed.

*6.4.1 Perceptual attunement in infants at family risk of dyslexia*

Perceptual attunement was assessed in FR infants. As the native contrast was a salient contrast, it was predicted that FR infants would be able to discriminate this contrast, or discrimination performance was expected to increase with age. With respect to the developmental trajectory of discrimination of the non-native contrast, two paths were conceivable. Either FR infants would show no evidence of discriminating the non-native contrast at any of the ages, in line with the literature on poor speech sound discrimination skills (e.g., Leppänen et al., 1999; Richardson et al., 2003; van Leeuwen et al., 2006). Alternatively, infants would show heightened sensitivity to the non-native contrast at all ages, aligning with the theory of allophonic perception (e.g., Noordenbos et al., 2012).

The findings of Chapter 4 indicate that FR infants have a subtle speech perception deficit, as group-based analyses showed that infants discriminated the salient native contrast but not the non-salient non-native contrast. These results agree with other studies of poorer speech perception in FR infants (e.g., Richardson et al., 2003; van Leeuwen et al., 2006). Moreover, the results of the Bayesian hierarchical modeling also point towards a delay in the formation of speech sound categories. This interpretation is based on two findings: First, the percentage of infants that could be classified as a discriminator dropped in the 10-month-old infants compared to the percentage of the 8-month-old infants. Such a 'dip' in the percentage was not expected, as the contrast is native and hence the 10-month-olds have more experience with the contrast, rendering a decrease in sensitivity is unlikely. The 10-month-old FR infants behaved like the 8-month-old no-FR infants. Both groups showed more inter-individual variation. This heterogeneous performance may reflect the emergence of perceptual attunement, which is thus slightly delayed for the FR infants compared to the no-FR infants.

The second finding that supports the claim that FR infants have a delayed development stems from findings on the non-native contrast. The 6-month-old FR infants did not show evidence of discrimination, neither in the frequentist analysis nor in the Bayesian approach. This could be due to a lack of initial sensitivity to such subtle contrasts and, obviously, as infants are not exposed to non-native contrasts, experience cannot help improve their discrimination performance. Literature on perceptual attunement in typically developing infants has shown that salience influences the ability to discriminate, as more subtle and less salient contrasts need to be acquired through language exposure (e.g., Liu & Kager, 2016; Narayan et al., 2010, but see the discussion in 6.3.1). Hence, the lack of evidence for non-native discrimination in the 6-month-old FR infants discriminating the non-native contrast can be used to argue that FR infants have a subtle delay in their development of speech sound categories. This yields the expectation that the formation of non-salient native contrasts will develop more slowly compared to their typically developing peers.

The Bayesian hierarchical modeling approach provided insight in the developmental paths (based on a cross-sectional sample) which was not visible in the group level analyses. At group level no differences were found between the no-FR and FR infants regarding their native discrimination performance. In contrast, individual data revealed that very few of the 10-month-old FR infants could be classified as a discriminator, even though the percentage of discriminators at 8 months was high. The 10-month-old no-FR peers showed a similar 'dip' at 8-month-old (discussed above). Differences between groups were attested in the non-native contrast at group and individual level. Of interest is that the individual assessment revealed that none (0%) of the FR infants could be classified as a discriminator. An open question in the literature is whether speech perception deficits are associated with dyslexia or with a risk of dyslexia (e.g., Moll et al., 2013; Pennington, et al., 2012; van Viersen et al., 2018). The results of the Bayesian analysis suggest that it is the elevated risk as such that is associated with a speech discrimination deficit.

**Taken together**, findings of this dissertation with regard to perceptual attunement and discrimination abilities of FR infants show that they have poorer discrimination of the non-native contrast. Findings on the native and non-native contrasts together seem to point towards a subtle delayed development of speech sound categories in FR infants. A subtle delay in the formation of speech sound categories could contribute to poorly developed phonological representations and is consistent with the theory of a phonological deficit. In order to more fully understand the developmental trajectory of the formation of native speech sound categories in FR infants, **future studies** should investigate perceptual attunement longitudinally. Moreover, FR infants' ability to discriminate subtle native speech sounds should be investigated, as this could show whether the development of speech sound categories is delayed compared to their peers. Also, more research is needed to find out which factors have a facilitating effect on the development of robust speech sound categories. A question that could be asked is whether speaker variability can facilitate phonetic learning in FR infants to the same extent as in no-FR infants, and whether an age effect can be found. This question is relevant for no-FR as well as FR infants as so far, not much is known about the effect of speaker variability in the context of perceptual attunement.

*6.4.2 Distributional learning of a vowel contrast*

Findings reported in this dissertation are inconclusive as regards a distributional learning deficit in FR infants. We obtained a null result (see also 6.3.1); the FR infants did not differ significantly from the no-FR infants, and, overall no effect was found for condition (unimodal vs. bimodal). However, exploratory analyses suggested that learning may have occurred in (some of) the no-FR infants in the bimodal, as opposed to the unimodal condition. These indications were not found for the FR infants.

The results of this dissertation do not allow for firm conclusions: no clear differences were found between no-FR and FR infants. In the literature, findings on distributional learning are inconsistent. Some find evidence for differences in performance between adults and children with (a family risk of) dyslexia and their peers (Kerkhoff et al., 2013; Lum et al., 2013; Vandermosten et al., 2019) others report null findings (see van Witteloostuijn et al., 2021 for references). Moreover, there are indications that differences between adults and children with dyslexia and their peers are inflated because of publication bias (Schmalz et al., 2017; van Witteloostuijn et al., 2017). As literature is inconclusive about whether or not distributional learning is different for individuals with (a family risk of) dyslexia, more research is needed to provide an answer.

What can be concluded from the results is that FR infants did not discriminate the English vowel contrast, neither in de unimodal nor in the bimodal condition. Although not much can be said about their ability to pick up on the distributional information, it can be concluded that the results cannot be used in favor of an allophonic speech perception Noordenbos et al., 2012). If FR infants would have a heightened sensitivity to allophones, they would have shown evidence of discriminating the contrast in both conditions.

## 6.5 Limitations of this dissertation

This dissertation contains several strengths, such as the collection of data from three different age groups, the use of speaker variability and a habituation design which allowed for individual assessment. Also, the Bayesian hierarchical modeling applied here constitutes an innovation in this field of study. It was demonstrated to provide insights that the traditional group-based frequentist approach does not offer. However, there were also some limitations that need to be mentioned. First, infants were recruited by sending letters to the parents of newborn babies. Parents with a high socio-economic status have registered their child to participate in research (see Chapter 4 and 5). This bias in the sample is unintentional and has a (somewhat) limiting effect in that caution needs to be exercised in generalizing to infants of different socio-economic status (SES).

SES can explain individual variation in language development, with the most pronounced and reliable differences in vocabulary size (Hoff, 2006). With regard to speech perception, mothers from higher SES generally talk more to their children, use more (variable) words, read more books together and facilitate more communicative experiences for their children (see for references Hoff, 2006). Hence, these children are exposed to more and more variable speech input, which could in potential influence the rate at which native categories are acquired.

Second, the analyses are based on cross-sectional samples. Although this dataset does provide indications about the expected (group and individual) developmental discrimination patterns of speech-sound contrasts in infants with and without and FR, longitudinal data is needed to confirm these patterns.

Finally, in the Bayesian hierarchical modeling approach group parameters were used to estimate the individual parameters, hence the hierarchical structure. As a consequence, sample size influences the confidence in the estimates of the group parameters. This is illustrated in the reported percentages of no-FR infants that discriminated the native contrast (Chapters 3 and 4), as these differ between chapters, while the infant data came from the same sample. For example, in Chapter 3 fifty-three percent of the 6-month-olds discriminated the contrast, whereas in Chapter 4 a percentage of eleven percent was reported. These differences are a result of different sample sizes. The sample size in Chapter 4 was smaller, because we matched the no-FR infants to the FR infants. As there were fewer FR infants, not all data of the total no-FR sample were used. A larger sample size leads to more accurate estimates. Furthermore, the credibility intervals of many infants barely crossed the zero value. This means that if there is just a little less or more variance at group level, or a little higher or lower mean of the condition effect, percentages change, due to a relatively small sample size. Therefore, even when applying Bayesian hierarchical modeling, it is important to obtain a sufficiently large sample size to estimate group level parameters with confidence.

## 6.6 Conclusion

This dissertation aimed at investigating the development of speech sound categories in infants with and without a family risk of dyslexia. There are four main conclusions. The first is that the developmental trajectory of speech sound discrimination is less predictable than is often assumed. Whereas perceptual attunement predicts monotonic developmental trajectories, the findings of this thesis suggest that patterns of discrimination are rather fluctuating. Secondly, it appears that the time window during which perceptual attunement is assumed to take place is more flexible than is usually assumed; it may well extend beyond the first year of life. The data show that towards the end of this suggested time window infants are still very flexible in their perception; some cues can activate the ability to distinguish a non-native contrast. Thirdly, Bayesian hierarchical modeling can be used to classify infants' individual native and non-native discrimination performance. However, for the assessment of individual discrimination to be really fruitful, it should be applied in longitudinal studies that investigate the relation with later language and literacy skills. Fourth, FR infants have a (subtle) delayed development of speech sound categories. FR infants showed poor discrimination of a subtle contrast but did show evidence of discriminating a salient contrast. A subtle delay in the formation of speech sound categories could contribute to poorly developed phonological representations and is consistent with the theory of a phonological deficit.

# Samenvatting (Summary in Dutch)

Dyslexie is een hardnekkig probleem met de verwerving van lees- en/of spellingsvaardigheden. Kinderen bij wie dyslexie voorkomt in de (eerstelijns) familie hebben een (sterk) verhoogd risico op dyslexie. Deze kinderen hebben een zogeheten familiair risico (FR) op dyslexie. Volgens een breed gedragen hypothese is een kernkenmerk van dyslexie dat dyslectici een fonologisch tekort hebben. Een veronderstelling is dat dit fonologisch tekort voortkomt uit niet goed ontwikkelde fonologische representaties. Deze abstracte mentale representaties van spraakklanken (fonemen) zijn nodig om goed te leren lezen en spellen. Voor het goed leren lezen en spellen moet er immers een verbinding gemaakt worden tussen de letter op het papier (grafeem) en de spraakklank (foneem). De ontwikkeling van spraakklankcategorieën begint al in het eerste levensjaar en is het best waarneembaar in de tweede helft daarvan. Het werk dat in dit proefschrift wordt beschreven, probeert vast te stellen of en hoe de ontwikkeling van spraakklankcategorieën verschilt tussen baby's met en baby's zonder een familiair risico op dyslexie. Dit proefschrift rapporteert over vier experimentele studies (Hoofstukken 2-5). Hieronder volgen de aanleiding, de opzet en de belangrijkste bevindingen per studie. In de laatste alinea staan de hoofdconclusies (Hoofdstuk 6) beschreven.

In Hoofdstuk 2 onderzochten we hoe de ontwikkeling van moedertaalklanken en klanken die niet in de moedertaal voorkomen verloopt bij NFR-baby's in het eerste levensjaar. Eerder onderzoek heeft aangetoond dat baby's geboren worden als universele luisteraars en zich ontwikkelen naar taalspecifieke luisteraars. Deze overgang wordt vaak perceptuele afstemming genoemd en wordt beschouwd als de eerste stap in de ontwikkeling van spraakklankcategorieën. De literatuur laat zien dat baby's aanvankelijk gevoelig zijn voor alle mogelijke spraakklankcontrasten, ook voor die contrasten die niet relevant zijn voor het spraakverstaan van hun moedertaal. Deze gevoeligheid voor niet-relevante contrasten neemt af gedurende de tweede helft van het eerste levensjaar. Er zijn echter aanwijzingen dat deze perceptuele afstemming minder eenvoudig is dan hierboven beschreven. Sommige contrasten worden bijvoorbeeld niet onderscheiden vanaf de geboorte. Daarnaast vinden niet alle studies een afname van de gevoeligheid voor contrasten die niet in de moedertaal voorkomen, waarbij dat wél verwacht werd op basis van de akoestische eigenschappen. De literatuur geeft over het proces van perceptuele afstemming dus geen uitsluitsel. Bovendien hebben klinkers in de literatuur minder aandacht gekregen. Ook is er nog niet veel bekend over de ontwikkeling van spraakklanken van Nederlandse baby's.

Hoofdstuk 2 rapporteert over een studie waarin de discriminatievaardigheden van 6, 8 en 10 maanden oude Nederlandse baby's werden onderzocht. Het eerste doel was om vast te stellen of zij het Nederlandse (/a:/ - /e:/) klinkercontrast konden onderscheiden. Het tweede was om vast te stellen hoe de ontwikkeling eruitzag van een contrast dat niet in de moedertaal voorkomt, namelijk het Engelse /ɛ/ - /æ/ klinkercontrast. De resultaten bevestigden de voorspellingen voor het moedertaalcontrast, namelijk dat baby's van alle leeftijdsgroepen het Nederlandse (/a:/ - /e:/) klinkercontrast konden onderscheiden. Het ontwikkelingspatroon voor het Engelse /ɛ/ - /æ/ was echter niet zoals verwacht: de 6 en 10 maanden oude baby's discrimineerden het contrast, terwijl de 8 maanden oude baby's dat niet deden. De waargenomen afname tussen de 6 en 8 maanden oude baby's is in lijn met de theorie van perceptuele afstemming. De toename van het discriminatievermogen bij de 10 maanden oude baby's is echter onverwacht. Een mogelijke verklaring is dat de 10 maanden oude baby's beter in staat waren dan de 8 maanden oude baby's om gebruik te maken van de sprekervariatie waaraan ze werden blootgesteld tijdens de gewenningsfase. De belangrijkste conclusie van Hoofdstuk 2 is dat het vermogen om een subtiel niet-moedertaalcontrast (zoals het Engelse /ɛ/ - /æ/ contrast) te onderscheiden, niet verloren is nadat het proces van perceptuele afstemming in gang is gezet. Een andere hoofdbevinding is dat voorspellingen over het verloop van de ontwikkeling op basis van perceptuele afstemming moeilijk te formuleren zijn, omdat het ontwikkelingspatroon kan fluctueren en de uitkomsten afhankelijk zijn van het toegepaste experimentele ontwerp, de leeftijd van de baby's en de akoestische eigenschappen van het spraakklankcontrast.

In hoofdstuk 3 stond de vraag centraal of op individueel niveau kon worden vastgesteld of baby's in staat waren het Nederlandse (/a:/ - /e:/) contrast te onderscheiden. De laatste tijd is er een groeiende interesse in het verklaren van individuele verschillen in de spraakwaarnemingsvaardigheden van baby's. Zo vond een meta-analyse bewijs dat vroege spraakwaarnemingsvaardigheden de latere taalvaardigheden kunnen voorspellen. Studies die relaties tussen vroege spraakwaarneming en latere taal- en leesontwikkeling onderzoeken, zijn echter gebaseerd op analyses op groepsniveau en kunnen geen individuele trajecten van taalontwikkeling voorspellen. Wanneer bijvoorbeeld het discriminatievermogen van spraakklankcontrasten op individueel niveau kan worden bepaald, dan kan deze uitkomst worden gebruikt om de ontwikkeling van spraakklanken longitudinaal te beoordelen én kan deze prospectief worden gebruikt om andere taal- en leesvaardigheden te voorspellen. Houston en collega's (2007) ontwierpen een paradigma om individuele discriminatie op woordniveau te

beoordelen. In hoofdstuk 3 borduurden wij hierop voort, maar wij onderzochten discriminatievaardigheden op klankniveau. De benadering van Hoofdstuk 3 was tweeledig: we repliceerden het frequentistische model dat door Houston en collega's werd gebruikt en we pasten Bayesiaanse hiërarchische modellering toe om individuele prestaties te beoordelen. Doordat met deze Bayesiaanse (hiërarchische) aanpak rekening wordt gehouden met de samenstelling (de variatie) van de groep, zijn schattingen op individueel niveau betrouwbaarder. Hoe minder variatie er is in de groep, hoe zekerder we zijn van de schattingen van hun individuele discriminatievaardigheid.
Het belangrijkste doel van Hoofdstuk 3 was om vast te stellen of we baby's betrouwbaar konden classificeren op basis van hun vermogen om het Nederlandse (/a:/ - /e:/) klinkercontrast te onderscheiden. Aangezien het moedertaalcontrast een opvallend contrast is, was de verwachting dat we een hoog percentage baby's zouden vinden dat het contrast kon onderscheiden, vergelijkbaar met de 80% die Houston en collega's (2007) in hun onderzoek vonden. De frequentistische benadering stelde echter vast dat 12% van de baby's het contrast discrimineerde en na correctie voor meervoudig testen slechts 3%. Uit de Bayesiaanse analyse bleek dat gemiddeld 50% van de baby's het contrast discrimineerde: 53% van de 6- maand oude baby's, 27% van de 8- maand oude baby's en 77% van de 10-maand oude baby's. In Hoofdstuk 3 geven we enkele verklaringen voor de verschillen die zijn gevonden tussen onze studie en die van Houston et al. Ook wordt er dieper in gegaan op de verschillen die gevonden zijn tussen de verschillende leeftijdsgroepen. Er zijn twee grote voordelen in de Bayesiaanse hierarschische aanpak. Op de eerste plaats hoeft er niet gecorrigeerd te worden voor het veelvuldig testen. Daarnaast leidt de Bayesiaanse aanpak tot grotere betrouwbaarheid omdat de groepsstructuur wordt meegenomen in de schattingen van de individuele prestaties. De belangrijkste conclusie van hoofdstuk 3 is daarom dat Bayesiaanse hiërarchische modellering de voorkeur heeft boven een frequentistische benadering bij het beoordelen van individuele discriminatieprestaties.

De ontwikkeling van de spraakperceptie van baby's met een familiair risico (FR) op dyslexie stond centraal in Hoofdstuk 4. We onderzochten of FR-baby's en hun leeftijdsgenoten zonder dit verhoogde risico (niet-FR) verschilden in hun onderscheidingsvermogen van een moedertaal en niet-moedertaal klinkercontrast. Talrijke studies hebben aangetoond dat individuen met dyslexie fonologische problemen hebben. De fonologische tekorthypothese stelt dat een tekortkoming in het vormen van representaties van spraakklanken een oorzakelijke factor is bij het ontstaan van dyslexie. Dit tekort kan zijn oorsprong hebben in (aangeboren) problemen in de spraakperceptie. Er zijn grofweg twee theorieën over hoe minder ontwikkelde fonologische representaties het gevolg zijn van problemen met de spraakperceptie. De eerste is dat individuen met (een risico op) dyslexie zwakkere spraakdiscriminatievaardigheden hebben. Zo laten studies met volwassenen en kinderen met dyslexie zien dat zij minder goed presteren op taken die het discriminatievermogen meet. Dit patroon is ook aangetoond bij FR-kinderen en -baby's. Als alternatief stelt de allofonische spraakperceptiehypothese dat volwassenen en kinderen met dyslexie en FR-kinderen een verhoogde en langdurige gevoeligheid vertonen voor fonetische variatie die niet relevant is voor het spraakverstaan van de moedertaal. De literatuur is dus niet eenduidig als het gaat om het voorspellen van het ontwikkelingspatroon van de formatie van spraakklankcategorieën, aangezien in de literatuur zowel slechte discriminatie als verhoogde (allofonische) discriminatie zijn gevonden.

Net als in hoofdstuk 2 zijn de discriminatievaardigheden van 6, 8 en 10 maanden oude baby's onderzocht met een Nederlands (/aː/ - /eː/) klinkercontrast en een Engels /ɛ/ - /æ/ contrast. Op basis van de literatuur die heeft laten zien dat volwassen en kinderen met dyslexie en kinderen en baby's met een FR op dyslexie een verminderd spraakwaarnemingsvermogen hebben, waren de voorspellingen voor de FR-baby's dat zij het /aː/ - /eː/ contrast zouden onderscheiden, maar dat het Engelse, subtiele, contrast op geen van de leeftijden zou worden onderscheiden. Op basis van allofonische spraakperceptie theorie, werd voorspeld dat FR-baby's juist een verhoogde gevoeligheid zouden vertonen voor het subtiele Engelse contrast. De resultaten toonden aan dat de FR- en niet-FR baby's op alle leeftijden het moedertaalcontrast konden discrimineren. Dit werd aangetoond in zowel de groeps- als individuele analyses. Er was echter een subtiele indicatie van een verschil tussen de twee groepen: terwijl verbetering van de prestaties werd aangetoond in de groep zonder FR, was dit niet het geval voor de FR-baby's. De resultaten van het Engelse contrast lieten een meer uitgesproken verschil zien tussen de niet-FR- en FR-baby's, aangezien er geen evidentie gevonden is dat de FR-baby's het contrast kunnen discrimineren. Ook de 6 maanden oude FR baby's lieten geen gevoeligheid voor het Engelse contrast zien. De baby's zonder een verhoogd risico lieten het eerder besproken (Hoofdstuk 2) ontwikkelingspatroon zien. Zij konden het contrast met 6 en 10 maanden onderscheiden. De theorie van perceptuele afstemming voorspelt dat 6 maanden oude baby's niet-moedertaalklanken wél kunnen onderscheiden en dat deze gevoeligheid rond de 8 maanden afneemt. Dit is niet gevonden voor de FR-baby's. Op basis van de bevindingen van het moedertaalcontrast (geen verbetering) en op basis van het niet-moedertaalcontrast (geen vroege gevoeligheid) is de hoofdconclusie van dit hoofdstuk dat FR-baby's een subtiele vertraging hebben in de ontwikkeling van spraakklanken. Deze interpretatie is in lijn met de literatuur die vond dat personen met dyslexie slechtere discriminatievaardigheden hebben. Onze bevindingen ondersteunen niet de theorie van de allofonische spraakperceptie.

Hoofdstuk 5 benaderde de verwerving van spraakklankcategorieën vanuit een andere invalshoek. De focus lag niet op het ontwikkelingspatroon van het discriminatievermogen, maar op een leermechanisme dat verantwoordelijk zou kunnen zijn voor de perceptuele afstemming die doorgaans wordt aangetroffen bij baby's tijdens de tweede helft van hun eerste levensjaar. Distributioneel fonetisch leren is naar voren gebracht als een onderliggend leermechanisme dat de perceptuele afstemming bij baby's ondersteunt. Er zijn aanwijzingen dat typisch ontwikkelende baby's in staat zijn om de distributiepatronen van spraakklanken te herkennen wat hen in staat zou stellen om de spraakklankcategorieën van hun moedertaal te leren. Of FR-baby's dezelfde gevoeligheid vertonen voor distributionele informatie in de taalomgeving is nog een open vraag, aangezien de literatuur niet eenduidig is. Er zijn studies die laten zien dat individuen met (een risico op) dyslexie slechter zijn in het oppikken van distributionele informatie, maar ook studies die afwezigheid van verschillen rapporteren. Daarnaast zijn er aanwijzingen voor een publicatiebias voor studies die groepsverschillen rapporteren. Het beoordelen van distributioneel leren van een fonetisch contrast bij FR-baby's is nodig om vast te stellen of er op jonge leeftijd aanwijzingen zijn dat zij deze informatie kunnen gebruiken voor het ontwikkelen van spraakklankcategorieën.
We onderzochten of baby's van 8 maanden oud een Engels /ɛ/ - /æ/ klinkercontrast konden onderscheiden na blootstelling aan deze klinkers in verschillende frequenties van vóórkomen. Tijdens de gewenningsfase werden baby's blootgesteld aan spraakklanken uit vier Engelse /ɛ/ -/æ/ continua van 8 stappen. Deze continua werden aangeboden ofwel in een unimodale verdeling, wat zou moeten leiden tot het waarnemen van het continuüm als één klinkercategorie, omdat de meest voorkomende klanken in het midden van de continua zaten. In de andere conditie waren de klanken bimodaal verdeeld, dat wil zeggen dat de meest voorkomende klanken aan de uiteinden van de continua zaten. Deze verdeling zou moeten leiden tot het waarnemen van het continuüm als een weerspiegeling van twee klinkercategorieën, namelijk /ɛ/ en /æ/. We hebben geen evidentie gevonden voor discriminatie, noch in de unimodale, noch in de bimodale conditie. Verder zijn er geen verschillen gevonden zijn tussen de niet-FR en de FR-groep. Aanvullende verkennende analyses leverden subtiele aanwijzingen op dat baby's zonder FR die langer keken tijdens de leerfase van het experiment, het contrast in de bimodale conditie wel konden discrimineren. Dit werd niet gevonden voor de

andere conditie. In de FR-groep werd dit effect niet gevonden in beide condities (uni- en bimodaal). Onze voorlopige conclusie was dan ook heel voorzichtig dat er (subtiele) verschillen zijn tussen de groepen.

In Hoofdstuk 6 worden de resultaten van hoofdstuk 2 tot en met 5 uitgebreider besproken, worden implicaties gegeven en suggesties gedaan voor toekomstig onderzoek. Tot slot zijn er vier hoofdconclusies. De eerste is dat het ontwikkelingstraject van discriminatie van spraakklanken minder voorspelbaar is dan vaak wordt aangenomen. Terwijl perceptuele afstemming monotone ontwikkelingstrajecten voorspelt, suggereren de bevindingen van dit proefschrift dat discriminatiepatronen fluctueren. Ten tweede lijkt het tijdvenster, waarin perceptuele afstemming wordt verondersteld plaats te vinden, langer te duren dan het veronderstelde eerste levensjaar. Ten derde kan Bayesiaanse hiërarchische modellering worden gebruikt om de individuele discriminatieprestaties van baby's te classificeren. Ten vierde hebben FR-baby's een (subtiele) vertraagde ontwikkeling van spraakklankcategorieën. Een subtiele vertraging in de vorming van spraakklankcategorieën zou kunnen bijdragen aan slecht ontwikkelde fonologische representaties en is consistent met de theorie van een fonologisch tekort.

# References

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. https://doi.org/10.1121/1.1779271

Adriaans, F., & Swingley, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *The Journal of the Acoustical Society of America*, *141*(5), 3070-3078. https://doi.org/10.1121/1.4982246

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: new data from an anticipatory eye movement paradigm. *Developmental Science*, *14*(2), 395-401. https://doi.org/10.1111/j.1467-7687.2010.00989.x

Altvater-Mackensen, N., & Grossmann, T. (2015). Learning to match auditory and visual speech cues: social influences on acquisition of phonological categories. *Child Development*, *86*(2), 362-378. https://doi.org/10.1111/cdev.12320

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: a meta-analysis. *Journal of Educational Psychology*, *107*(3), 868. https://doi.org/10.1037/edu0000006

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48-53. https://doi.org/10.1111/j.1467-7687.2007.00563.x

Aslin, R. N., & Fiser, J. (2005). Methodological challenges for understanding cognitive development in infants. *Trends in Cognitive Sciences*, *9*(3), 92-98. https://doi.org/10.1016/j.tics.2005.01.003

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.

Banai, K., & Ahissar, M. (2018). Poor sensitivity to sound statistics impairs the acquisition of speech categories in dyslexia. *Language, Cognition and Neuroscience*, *33*(3), 321-332. https://doi.org/10.1080/23273798.2017.1408851

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society.* Series B (Methodological), *57*(1), 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *Haskins Laboratories Status & Report, Speech Research* 1991, SR·107/ 108, 1-30.

Best, C. T., & Faber, A. (2000). Developmental increase in infants' discrimination of nonnative vowels that adults assimilate to a single native vowel. *Paper presented at the International Conference on Infant Studies*, *16*, 19.

Best, C. T., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development*, *21*, 295. https://doi.org/10.1016/SO163-6383(98)91508-9

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development*, *18*(3), 339-350. https://doi.org/10.1177/00238309030460020701

Bjorklund, D. F., Miller, P. H., Coyle, T. R., & Slawinski, J. L. (1997). Instructing children to use memory strategies: Evidence of utilization deficiencies in memory training studies. *Developmental Review*, *17*(4), 411-441. https://doi.org/10.1006/drev.1997.0440

Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., & Blomert, L. (2009). Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Current Biology*, *19*(6), 503-508. https://doi.org/10.1016/j.cub.2009.01.065

Blomert, L. (2011). The neural signature of orthographic–phonological binding in successful and failing reading development. *Neuroimage*, *57*(3), 695-703. https://doi.org/10.1016/j.neuroimage.2010.11.003

Blomert, L., & Mitterer, H. (2004). The fragile nature of the speech- perception deficit in dyslexia: Natural vs. synthetic speech. *Brain and Language*, *89*(1), 21-26. https://doi.org/10.1016/j.bandl.2006.06.009

Boersma, P., & Weenink, D. (2015) *Praat: doing phonetics by computer* (version 5.4.06) [computer software]. https://www.fon.hum.uva.nl/praat

Boets, B., Ghesquière, P., Van Wieringen, A., & Wouters, J. (2007). Speech perception in preschoolers at family risk for dyslexia: Relations with low-level auditory processing and phonological ability. *Brain and Language*, *101*(1), 19-30. https://doi.org/10.1016/j.bandl.2006.06.009

Bonte, M. L., Poelmans, H., & Blomert, L. (2007). Deviant neurophysiological responses to phonological regularities in speech in dyslexic children. *Neuropsychologia*, *45*(7), 1427-1437. https://doi.org/10.1016/j.neuropsychologia.2006.11.009

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. https://doi.org/10.1177/00238309030460020801

Bradlow, A.R., Akahane-Yamada, R., Pisoni, D.B & Tokhura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*, 977–985. https://doi.org/10.3758/ BF03206911

Brandt, J., & Rosen, J. J. (1980). Auditory phonemic perception in dyslexia: Categorical identification and discrimination of stop consonants. *Brain and Language*, *9*(2), 324-337. https://doi.org/10.1016/0093-934X(80)90152-2

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *The Quarterly Journal of Experimental Psychology, 64*(1), 74-95. https://doi.org/10.1080/17470218.2010.499174

Brus, B. Th., & Voeten, M. J. M. (1972). Een-Minuut-Test. Vorm A en B [One minute Test]. Nijmegen: Berkhout Testmateriaal.

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. https://doi.org/10.1017/s0142716407070257

Bus, A. G., & Van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational psychology*, *91*(3), 403. https://doi.org/10.1037//0022-0663.91.3.403.

Caglar-Ryeng, Ø., Eklund, K., & Nergård-Nilssen, T. (2019). Lexical and grammatical development in children at family risk of dyslexia from early childhood to school entry: a cross-lagged analysis. *Journal of Child Language*, *46*(6), 1102-1126. https://doi.org/10.1017/S0305000919000333

Capel, D., de Bree, E., de Klerk, M., Kerkhoff, A., & Wijnen, F. (2011). Distributional cues affect phonetic discrimination in Dutch infants. *Sound and sounds*. Studies presented to MEH (Bert) Schouten on the occasion of his 65th birthday, pp.33-43. Utrecht: UiL-OTS, LOT Publications.

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. https://doi.org/10.1038/1561

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory*, *92*(2), 225-234. https://doi.org/10.1016/j.nlm.2008.06.002

Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseman, L. J. (1991). Individual differences in infant visual attention: Are short lookers faster processors or feature processors? *Child Development*, *62*(6), 1247-1257. https://doi.org/10.1111/j.1467-8624.1991.tb01603.x

Conboy, B. T., Sommerville, J. A., & Kuhl, P. K. (2008). Cognitive control factors in speech perception at 11 months. *Developmental Psychology*, *44*(5), 1505. https://doi.org/10.1037/a0012975

Cristia, A. (2011). Fine-grained variation in caregivers'/s/predicts their infants'/s/ category. *The Journal of the Acoustical Society of America*, *129*(5), 3271-3280. https://doi.org/10.1121/1.3562562

Cristia, A. (2018). Can infants learn phonology in the lab? A meta-analytic answer. *Cognition*, *170*, 312-327. https://doi.org/10.1016/j.cognition.2017.09.016

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Development*, *85*(4), 1330-1345. https://doi.org/10.1111/cdev.12193

Cristia, A., Seidl, A., Singh, L., & Houston, D. (2016). Test–retest reliability in infant speech perception tasks. *Infancy*, *21*(5), 648-667. https://doi.org/10.1111/infa.12127

de Bree, E., Wijnen, F., & Gerrits, E. (2010). Non-word repetition and literacy in Dutch children at risk of dyslexia and children with SLI: Results of the follow-up study. *Dyslexia*, *16*(1), 36-44. https://doi.org/10.1080/02699200701576892

de Klerk, M., de Bree, E., Kerkhoff, A., & Wijnen, F. (2019). Lost and found: decline and reemergence of non-native vowel discrimination in the first year of life. *Language Learning and Development*, *15*(1), 14-31. https://doi.org/10.1080/15475441.2018.1497490 de Klerk, M., de Bree, E., & Wijnen, F. (2021). Speech discrimination in infants at family risk of dyslexia: Group and individual-based analyses. *Journal of Experimental Child Psychology*, *206*, 105066. https://doi.org/10.1016/j.jecp.2020.105066

de Klerk, M., Veen, D., Wijnen, F., & de Bree, E. (2019). A step forward: Bayesian hierarchical modelling as a tool in assessment of individual discrimination performance. *Infant Behavior and Development*, *57*, 101345. https://doi.org/10.1016/j.infbeh.2019.101345

Deterding, D. (1997). The formants of monophthong vowels in standard Southern British English pronunciation. *Journal of the International Phonetic Association*, *27*(1/2), 47-55. http://www.jstor.org/stable/44516138

Dijkstra, C. G. J., & Fikkert, J. P. M. (2011). Universal Constraints on the Discrimination of Place of Articulation? Asymmetries in the Discrimination of 'paan'and 'taan'by 6-month-old Dutch Infants. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th Annual Boston University Conference on Language Development* (pp.170–182). Boston, MA: Cascadilla Press.

Dijkstra, N., & Fikkert, P. (2010). Do Dutch six-month-olds discriminate paan and taan and vice versa. *Poster presented at The Boston University Conference on Language Development*, Boston, MA.

Donker, M., Kroesbergen, E., Slot, E., Van Viersen, S., & De Bree, E. (2016). Alphanumeric and non-alphanumeric Rapid Automatized Naming in children with reading and/or spelling difficulties and mathematical difficulties. *Learning and Individual Differences*, *47*, 80-87. https://doi.org/10.1016/j.lindif.2015.12.011

Estes, K. G., & Lew-Williams, C. (2015). Listening through voices: Infant statistical word segmentation across multiple speakers. *Developmental Psychology*, *51*(11), 1517. https://doi.org/10.1037/a0039725

Eilers, R. E., & Minifie, F. D. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research*, *18*(1), 158-167. https://doi.org/10.1044/jshr.1801.158

Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review*, *120*(4), 751. https://doi.org/10.1037/a0034245

Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th edition). Sage Publication Ltd.

Fisher, H. B., & Linville, S. E. (1985). Acoustic characteristics of women's voices with advancing age. *Journal of Gerontology*, *40*(3), 324-330. https://doi.org/10.1093/geronj/40.3.324

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., ... & Lew-Williams, C. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*(4), 421-435. https://doi.org/10.1111/infa.12182

Gabay, Y., Thiessen, E. D., & Holt, L. L. (2015). Impaired statistical learning in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*, *58*(3), 934-945. https://doi.org/10.1044/2015_JSLHR-L-14-0324

Gelman, A. (2006). Multilevel (hierarchical) modeling: what it can and cannot do. Technometrics, *48*(3), 432-435. https://doi.org/10.1198/004017005000000661

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189-211. https://doi.org/10.1080/19345747.2011.618213

Gelman, A., & Tuerlinckx, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Computational Statistics*, *15*(3), 373-390. https://doi.org/10.1007/s001800000040

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis. (vol.3) Chapman and Hall/CRC Boca Raton, FL, USA.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, *32*(2), 249. https://doi.org/10.1017/S0305000904006786

Gerrits, E., & de Bree, E. (2009). Early language development of children at familial risk of dyslexia: Speech perception and production. *Journal of Communication Disorders*, *42*(3), 180-194. https://doi.org/10.1016/j.jcomdis.2008.10.004

Gervain, J. & Filippin, L. Psycode (version 1.1.10). http://psy.ck.sissa.it/PsyCode/PsyCode.html.

Gons, B. (2021). *doe het toch maar*. Atlas Contact.

Fraga González, G, Žarić, G., Tijms, J., Bonte, M., Blomert, L., & van der Molen, M.W. (2015). A Randomized Controlled Trial on The Beneficial Effects of Training Letter-Speech Sound Integration on Reading Fluency in Children with Dyslexia. *PLoS ONE, 10*(12): e0143914. https://doi.org/10.1371/journal.pone.0143914

Guttorm, T. K., Leppänen, P. H. T., Hämäläinen, J. A., Eklund, K. M., & Lyytinen, H. J. (2010). Newborn Event-Related Potentials Predict Poorer Pre-Reading Skills in Children at Risk for Dyslexia. *Journal of Learning Disabilities*, *43*(5), 391–401. https://doi.org/10.1177/0022219409345005

Guttorm, T. K., Leppänen, P. H., Poikkeus, A. M., Eklund, K. M., Lyytinen, P., & Lyytinen, H. (2005). Brain event-related potentials (ERPs) measured at birth predict later language development in children with and without familial risk for dyslexia. *Cortex*, *41*(3), 291-303. https://doi.org/10.1016/S0010-9452(08)70267-3

Guttorm, T. K., Leppänen, P. H., Richardson, U., & Lyytinen, H. (2001). Event-related potentials and consonant differentiation in newborns with familial risk for dyslexia. *Journal of Learning Disabilities*, *34*(6), 534-544. doi: https://doi.org/10.1177/002221940103400606

Hakvoort, B., de Bree, E., van der Leij, A., Maassen, B., van Setten, E., Maurits, N., & van Zuijen, T. L. (2016). The role of categorical speech perception and phonological processing in familial risk children with and without dyslexia. *Journal of Speech, Language, and Hearing Research*, *59*(6), 1448-1460. https://doi.org/10.1044/2016_JSLHR-L-15-0306

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., & Shakespeare, B. (2009). Speech perception abilities of adults with dyslexia: Is there any evidence for a true deficit? *Journal of Speech, Language, and Hearing Research*, *52*(6), 1510-1529. https://doi.org/10.1044/1092-4388(2009/08-0220)

Heeren, W. F. L. (2006). Perceptual development of phoneme contrasts in adults and children. [Unpublished doctoral dissertation]. Utrecht University. LOT Publications.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental review*, *26*(1), 55-88. https://doi.org/10.1016/j.dr.2005.11.002

Horn, D. L., Houston, D. M., & Miyamoto, R. T. (2007). Speech discrimination skills in deaf infants before and after cochlear implantation. *Audiological Medicine*, *5*(4), 232-241. https://doi.org/10.1080/16513860701659438

Houston, D. M., Horn, D. L., Qi, R., Ting, J. Y., & Gao, S. (2007). Assessing speech discrimination in individual infants. *Infancy*, *12*(2), 119-145. https://doi.org/10.1111/j.1532-7078.2007.tb00237.x

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*, *13*(4), 341-348. https://doi.org/10.1002/icd.364

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

Kachel, G., Hardecker, D. J. K., & Bohn, M. (2020). Young children's developing ability to integrate gestural and emotional cues. *Journal of Experimental Child Psychology*, *201*, 104984. https://doi.org/10.1016/j.jecp.2020.104984

Kerkhoff, A., de Bree, E., De Klerk, M. & Wijnen, F. (2013). Non-adjacent dependency learning in infants at familial risk of dyslexia. *Journal of Child Language*, *40*(1), 11-28. https://doi.org/10.1017/S0305000912000098

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General, 142*(2), 573.

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

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London*. *Series B, Biological Sciences*, *363*(1493), 979-1000. https://doi.org/10.1098/rstb.2007.2154

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. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096-9101. https://doi.org/10.1073/pnas.1532872100

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

Kuijpers, C. T. L., Leij, A., Been, P. H., van Leeuwen, T. G., Keurs, M. T., & Schreuder, R. (2003). Leesproblemen in het voortgezet onderwijs en de volwassenheid. *Pedagogische Studiën*, *80*(4), 272-287.

Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development*, *18*(4), 459-475. https://doi.org/10.1016/0163-6383(95)90035-7

Lee, M. D. (2018). Bayesian methods in cognitive modeling. In E. Wagenmakers (Ed.), Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Methodology (4th ed. Vol.5, pp 37-84). John Wiley & Sons.

Leppänen, P. H., Pihko, E., Eklund, K. M., & Lyytinen, H. (1999). Cortical responses of infants with and without a genetic risk for dyslexia: II. Group effects. *Neuroreport*, *10*(5), 969-973.

Leppänen, P. H., Richardson, U., Pihko, E., Eklund, K. M., Guttorm, T. K., Aro, M., & Lyytinen, H. (2002). Brain responses to changes in speech sound durations differ between infants with and without familial risk for dyslexia. *Developmental Neuropsychology*, *22*(1), 407-422. https://doi.org/10.1207/S15326942dn2201_4

Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, *133*(2), 385-394. https://doi.org/10.1016/j.cognition.2014.06.004

Liu, L., & Kager, R. (2015). Bilingual exposure influences infant VOT perception. Infant *Behavior and Development*, *38*, 27-36. https://doi.org/10.1016/j.infbeh.2014.12.004

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. https://doi.org/10.1177/1367006914566082

Liu, L., & Kager, R. (2017). Statistical learning of speech sounds is most robust during the period of perceptual attunement. *Journal of Experimental Child Psychology*, *164*, 192-208. https://doi.org/10.1016/j.jecp.2017.05.013

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242-1255. https://doi.org/10.1121/1.408177

Livingston, E. M., Siegel, L. S., & Ribary, U. (2018). Developmental dyslexia: Emotional impact and consequences. *Australian Journal of Learning Difficulties*, *23*(2), 107-135. https://doi.org/10.1080/19404158.2018.1479975

Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. In J. Slitka, S. Manuel, and M. Matthies (Eds.), *From sound to sense: 50 years of discoveries in speech communication* (C381-C386). Research Laboratory of Electronics at MIT, Electronic conference proceedings

Lum, J. A., Ullman, M. T., & Conti-Ramsden, G. (2013). Procedural learning is impaired in dyslexia: Evidence from a meta-analysis of serial reaction time studies. *Research in Developmental Disabilities*, *34*(10), 3460-3476. https://doi.org/10.1016/j.ridd.2013.07.017

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia*, *53*(1), 1-14. https://doi.org/10.1007/s11881-003-0001-9

Maassen, B., Groenen, P., Crul, T., Assman-Hulsmans, C., & Gabreëls, F. (2001). Identification and discrimination of voicing and place-of- articulation in developmental dyslexia. *Clinical Linguistics & Phonetics*, *15*(4), 319-339. https://doi.org/10.1080/02699200010026102

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154-178. https://doi.org/10.1002/dev.21177

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122-134. https://doi.org/10.1111/j.1467-7687.2007.00653.x

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111. https://doi.org/10.1016/S0010-0277(01)00157-3

Mazuka, R., Hasegawa, M., & Tsuji, S. (2014). Development of non-native vowel discrimination: Improvement without exposure. *Developmental Psychobiolog*y, *56*(2), 192-209. https://doi.org/10.1002/dev.21193

Melvin, S. A., Brito, N. H., Mack, L. J., Engelhardt, L. E., Fifer, W. P., Elliott, A. J., & Noble, K. G. (2017). Home environment, but not socioeconomic status, is linked to differences in early phonetic perception ability. *Infancy*, *22*(1), 42-55. https://doi.org/10.1111/infa.12145

Messaoud-Galusi, S., Hazan, V., & Rosen, S. (2011). Investigating speech perception in children with dyslexia: Is there evidence of a consistent deficit in individuals? *Journal of Speech, Language, and Hearing Research*, *54*(6), 1682-1701. https://doi.org/10.1044/1092-4388(2011/09-0261)

Mittag, M., Thesleff, P., Laasonen, M., & Kujala, T. (2013). The neurophysiological basis of the integration of written and heard syllables in dyslexic adults. *Clinical Neurophysiology*, *124*(2), 315-326. https://doi.org/10.1016/j.clinph.2012.08.003

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331-340. https://doi.org/10.3758/BF03211209

Mody, M., Studdert-Kennedy, M., & Brady, S. (1997). Speech perception deficits in poor readers: Auditory processing or phonological coding?. *Journal of Experimental Child Psychology*, *64*(2), 199-231. https://doi.org/10.1006/jecp.1996.2343

Molfese, D. L. (2000). Predicting dyslexia at 8 years of age using neonatal brain responses. *Brain and Language*, *72*(3), 238-245. https://doi.org/10.1006/brln.2000.2287

Moll, K., Loff, A., & Snowling, M. J. (2013). Cognitive endophenotypes of dyslexia. *Scientific Studies of Reading*, *17*(6), 385-397. https://doi.org/10.1080/10888438.2012.736439

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica*, *102*(2), 156-160. https://doi.org/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. https://doi.org/10.1037/a0014043

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. https://doi.org/10.1111/j.1467-7687.2009.00898.x

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

Nittrouer, S., Shune, S., Lowenstein, J. H. (2011). What is the deficit in phonological processing deficits: auditory sensitivity, masking, or category formation? *Journal of Experimental Child Psychology*, *108*(4), 762–785. https://doi.org/10.1016/j.jecp.2010.10.012

Noordenbos, M. W., Segers, E., Serniclaes, W., Mitterer, H., & Verhoeven, L. (2012). Allophonic mode of speech perception in Dutch children at risk for dyslexia: A longitudinal study. *Research in Developmental Disabilities*, *33*(5), 1469-1483. https://doi.org/10.1016/j.ridd.2012.03.021

Noordenbos, M. W., & Serniclaes, W. (2015). The categorical perception deficit in dyslexia: A meta-analysis. *Scientific Studies of Reading*, *19*(5), 340-359. https://doi.org/10.1080/10888438.2015.1052455

Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development*, *11*(3), 255-268. https://doi.org/10.1080/15248371003699977

Pauls, F., Macha, T., & Petermann, F. (2013). U-shaped development: an old but unsolved problem. *Frontiers in Psychology*, *4*, 301. https://doi.org/10.3389/fpsyg.2013.00301

Pavlidou, E. V., Louise Kelly, M., & Williams, J. M. (2010). Do children with developmental dyslexia have impairments in implicit learning? *Dyslexia*, *16*(2), 143-161. https://doi.org/10.1002/dys.400

Pelphrey, K. A., Reznick, J. S., Davis Goldman, B., Sasson, N., Morrow, J., Donahoe, A., & Hodgson, K. (2004). Development of visuospatial short-term memory in the second half of the 1st year. *Developmental Psychology*, *40*(5), 836. https://doi.org/10.1037/0012-1649.40.5.836

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385-413. https://doi.org/10.1016/j.cognition.2006.04.008

Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., … Olson, R. K. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of Abnormal Psychology*, *121*(1), 212-224. https://doi.org/10.1037/a0025823

Peterson, R. L., & Pennington, B. F. (2015). Developmental dyslexia. *Annual Review of Clinical Psychology*, *11*, 283-307. https://doi.org/10.1146/annurev-clinpsy-032814-112842

Pihko, E., Leppänen, P. H., Eklund, K. M., Cheour, M., Guttorm, T. K., & Lyytinen, H. (1999). Cortical responses of infants with and without a genetic risk for dyslexia: I. Age effects. *Neuroreport*, *10*(5), 901-905.

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

Polka, L., & Bohn, O. (2011). Natural referent vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467-478. https://doi.org/10.1016/j.wocn.2010.08.007

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of/d/–/ð/ perception: Evidence for a new developmental pattern. *The Journal of the Acoustical Society of America, 109*(5), 2190-2201. https://doi.org/10.1121/1.1362689.

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. https://doi.org/10.1037/0096-1523.20.2.421

Potter, C. E., & Saffran, J. R. (2017). Exposure to multiple accents supports infants' understanding of novel accents. *Cognition*, *166*, 67-72. https://doi.org/10.1016/j.cognition.2017.05.031

Protopapas, A. (2014). From temporal processing to developmental language disorders: mind the gap. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20130090. https://doi.org/10.1098/rstb.2013.009

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain*, *126*(4), 841-865. https://doi.org/10.1093/brain/awg076

Reh, R. K., Dias, B. G., Nelson, C. A., Kaufer, D., Werker, J. F., Kolb, B., Levine, J. D. & Hensch, T. K. (2020). Critical period regulation across multiple timescales. In Sokolowski, M.B. (Ed.), *Proceedings of the National Academy of Sciences*, 117(38), 23242-23251. https://doi.org/10.1073/pnas.1820836117

Reh, R. K., Hensch, T. K., & Werker, J. F. (2021). Distributional learning of speech sound categories is gated by sensitive periods. *Cognition*, 104653. https://doi.org/10.1016/j.cognition.2021.104653

Richardson, U., Leppänen, P. H., Leiwo, M., & Lyytinen, H. (2003). Speech perception of infants with high familial risk for dyslexia differ at the age of 6 months. *Developmental Neuropsychology*, *23*(3), 385-397. https://doi.org/10.1207/S15326942DN2303_5

Rosen, S., Manganari, E. (2001). Is there a relationship between speech and nonspeech auditory processing in children with dyslexia? *Journal of Speech, Language and Hearing Research*, *44*(4), 720–736. https://doi.org/10.1044/1092-4388(2001/057

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339-349. https://doi.org/10.1111/j.1467-7687.2008.00786.x

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606-621. https://doi.org/10.1006/jmla.1996.0032

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by japanese infants. *Developmental Psychology*, *46*(1), 106. https://doi.org/10.1111/j.1467-7687.2008.00786.x

Schmalz, X., Altoè, G., & Mulatti, C. (2017). Statistical learning and dyslexia: A systematic review. *Annals of Dyslexia*, *67*(2), 147-162. https://doi.org/10.1007/s11881-016-0136-0

Schouten, M. E. H. (1975). Native-language interference in the perception of second-language vowels: An investigation of certain aspects of the acquisition of a second language. [Unpublished doctoral dissertation]. Utrecht University. LOT Publications

Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (2001). Speech perception deficit in dyslexic adults as measured by mismatch negativity (MMN). *International Journal of Psychophysiology*, *40*(1), 77-87. https://doi.org/10.1016/S0167-8760(00)00152-5

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. https://doi.org/10.1111/j.1467-7687.2009.00829.x

Serniclaes, W., Van Heghe, S., Mousty, P., Carré, R., & Sprenger- Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology*, *87*(4), 336-361. https://doi.org/10.1016/j.jecp.2004.02.001

Siegler, R. S. (2004). U-shaped interest in U-shaped development - and what it means. *Journal of Cognition and Development*, *5*(1), 1-10. https://doi.org/10.1207/s15327647jcd0501_1

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

Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, *51*(2), 173-189. https://doi.org/10.1016/j.jml.2004.04.00

Singh, L., & Tan, A. R. (2020). Beyond perceptual narrowing: Monolingual and bilingual infants discriminate Hindi contrasts when learning words in the second year of life. *Developmental Psychology*, *57*(1), 19-32. https://doi.org/10.1037/dev0001137

Smits, R., Warner, N., McQueen, J. M., & Cutler, A. (2003). Unfolding of phonetic information over time: A database of Dutch diphone perception. *The Journal of the Acoustical Society of America*, *113*(1), 563-574. https://doi.org/10.1121/1.1525287

Snowling, M. J., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin*, *142*(5), 498. https://doi.org/10.1037/bul0000037

Sokolov, E. N. (1963). Perception and the conditioned reflex. New York: Macmillan

Steffens, M. L., Eilers, R. E., Gross-Glenn, K., & Jallad, B. (1992). Speech perception in adult subjects with familial dyslexia. *Journal of Speech, Language, and Hearing Research*, *35*(1), 192-200. https://doi.org/10.1044/jshr.3501.192

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. https://doi.org/10.1016/j.cognition.2018.05.009

Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*, *108*(1), 232-242. https://doi.org/10.1016/j.cognition.2007.12.013

Suwandi, G. R. F., Mustajab, M. A., Haekal, M., Khotimah, S. N., & Haryanto, F. (2020, March). *An Acoustic Analysis of Formants between Frequently Smoking Subjects and Non-smoking Subjects*. In Journal of Physics: Conference Series, Indonesia, 1505(1), 012053. https://doi.org/10.1088/1742-6596/1505/1/012053

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3617-3632. https://doi.org/10.1098/rstb.2009.0107

ter Schure, S., Junge, C., & Boersma, P. (2016). Discriminating non-native vowels on the basis of multimodal, auditory or visual information: effects on infants' looking patterns and discrimination. *Frontiers in Psychology*, *7*, 525, 1-11. https://doi.org/10.3389/fpsyg.2016.00525

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. https://doi.org/10.1016/j.clinph.2019.01.019

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, *75*(4), 1067-1084. https://doi.org/10.1111/j.1467-8624.2004.00726.x

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

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., ... & Best, C. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: Language-specific developmental changes in speech perception abilities. In *ICSLP-1994*, 1695-1698.

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, *71*(1), 4-21. https://doi.org/10.1159/000356237

Tyler, M. D., Best, C. T., Goldstein, L. M., & Antoniou, M. (2014). Investigating the role of articulatory organs and perceptual assimilation of native and non-native fricative place contrasts. *Developmental Psychobiology, 56*(2), 210-227. https://doi.org/10.1002/dev.21195

Uterwijk, J. (2000). WAIS-III Nederlandstalige bewerking. Technische handleiding [WAIS-III Dutch version. Technical manual]. Lisse: Swets & Zeitlinger.

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

van Bergen, E., de Jong, P. F., Plakas, A., Maassen, B., & van der Leij, A. (2012). Child and parental literacy levels within families with a history of dyslexia. J*ournal of Child Psychology and Psychiatry*, *53*(1), 28-36. https://doi.org/10.1111/j.1469-7610.2011.02418.x

van den Bos, K. P., Spelberg, H., Scheepsma, A., & de Vries, J. (1994). De Klepel. Vorm A en B. Een test voor de leesvaardigheid van pseudowoorden. Verantwoording, handleiding, diagnostiek en behandeling. Berkhout, Nijmegen, The Netherlands.

van Hessen, A. (1992). Discrimination of familiar and unfamiliar speech sounds. [Unpublished doctoral dissertation]. Utrecht University. LOT Publications.

van Leeuwen, T., Been, P., Kuijpers, C., Zwarts, F., Maassen, B., & van der Leij, A. (2006). Mismatch response is absent in 2-month-old infants at risk for dyslexia. *Neuroreport*, *17*(4), 351-355. https://doi.org/10.1097/01.wnr.0000203624.02082.2dn van Viersen, S., de Bree, E. H., Zee, M., Maassen, B., van der Leij, A., & de Jong, P. F. (2018). Pathways into literacy: The role of early oral language abilities and family risk for dyslexia. *Psychological Science*, *29*(3), 418-428. https://doi.org/10.1177/0956797617736886

van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2017). Visual artificial grammar learning in dyslexia: A meta-analysis. *Research in Developmental Disabilities*, *70*, 126–137. https://doi.org/10.1016/j.ridd.2017.09.006

van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2019). Statistical learning abilities of children with dyslexia across three experimental paradigms. *PloS One*, *14*(8), e0220041. https://doi.org/10.1371/journal.pone.0220041

van Zuijen, T. L., Plakas, A., Maassen, B. A., Maurits, N. M., & van der Leij, A. (2013). Infant ERPs separate children at risk of dyslexia who become good readers from those who become poor readers. *Developmental Science*, *16*(4), 554-563. https://doi.org/10.1111/desc.12049

Vandermosten, M., Wouters, J., Ghesquière, P., & Golestani, N. (2019). Statistical learning of speech sounds in dyslexic and typical reading children. *Scientific Studies of Reading*, *23*(1), 116-127. https://doi.org/10.1080/10888438.2018.1473404

Veenker, T.J.G. (2008). The Zep Experiment Control Application (Version 1.4.1) [Computer software]. Utrecht Institute of Linguistics OTS, Utrecht University. Available from http://www.hum.uu.nl/uilots/lab/zep/

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*(1), 2- 40. https://doi.org/10.1046/j.0021-9630.2003.00305.x

Volkmer, S., & Schulte-Körne, G. (2018). Cortical responses to tone and phoneme mismatch as a predictor of dyslexia? A systematic review. *Schizophrenia Research*, *191*, 148-160. https://doi.org/10.1016/j.schres.2017.07.01

Vukatana, E., Graham, S. A., Curtin, S., & Zepeda, M. S. (2015). One is not enough: Multiple exemplars facilitate infants' generalizations of novel properties. Infancy, *20*(5), 548-575. https://doi.org/10.1111/infa.12092

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, 1-12. https://doi.org/10.3389/fpsyg.2014.00077

Werker, J. F. (1994) Cross-language speech perception: Development change does not involve loss. In: Goodman, J., & Nusbaum, H. C. (Eds.). *The development of speech perception: The transition from speech sounds to spoken words*. MIT Press

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197-234.: https://doi.org/10.1080/15475441.2005.9684216

Werker, J. F, Hall, D. G., & Fais, L. (2004). Reconstruing U-shaped functions. *Journal of Cognition and Development*, *5*(1), 147-151. https://doi.org/10.1207/s15327647jcd0501_15

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

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*(5), 672. https://doi.org/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 and Development*, *7*(1), 49-63. https://doi.org/10.1016/S0163-6383(84)80022-3

Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *41*(1), 48. https://doi.org/10.1037/h0084150

Wijnen, F.N.K. (2013). Acquisition of linguistic categories: cross-domain convergences. In J. Bolhuis & M. Everaert (Eds.), *Birdsong, Speech and Language: Converging Mechanisms* (pp. 157-178). MIT Press.

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. https://doi.org/10.1016/j.cognition.2009.08.010

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, *15*(4), 420-433. https://doi.org/10.1111/j.1532-7078.2009.00024.x

# Appendices

## Appendix A (Chapter 3)

*Mean looking times per condition (alternating and non-alternating), difference score and p value for condition for each infant*

| Participant | Age | Condition | | Difference | Statistics |
|---|---|---|---|---|---|
| | (months) | Alternating | Non-alternating | Alt minus non-alt | *p_adj* |
| **child 10** | **6** | **4,05** | **3,74** | **0,31** | **.012** |
| **child 38** | **6** | **3,71** | **3,59** | **0,12** | **.022** |
| child 31 | 6 | 3,92 | 3,69 | 0,22 | .055 |
| child 4 | 6 | 4,26 | 3,98 | 0,28 | .055 |
| child 18 | 6 | 4,43 | 4,08 | 0,35 | .062 |
| child 35 | 6 | 3,95 | 3,67 | 0,29 | .074 |
| child 15 | 6 | 4,22 | 3,98 | 0,24 | .100 |
| child 25 | 6 | 4,26 | 3,94 | 0,32 | .113 |
| child 29 | 6 | 4,06 | 3,95 | 0,11 | .128 |
| child 37 | 6 | 4,24 | 4,02 | 0,22 | .133 |
| child 17 | 6 | 3,74 | 3,58 | 0,16 | .134 |
| child 11 | 6 | 4,34 | 3,99 | 0,35 | .14 |
| child 26 | 6 | 4,20 | 4,06 | 0,14 | .211 |
| child 30 | 6 | 3,88 | 3,75 | 0,13 | .23 |
| child 14 | 6 | 3,61 | 3,76 | -0,16 | .258 |
| child 3 | 6 | 3,80 | 3,94 | -0,14 | .278 |
| child 28 | 6 | 4,16 | 3,90 | 0,26 | .293 |
| child 22 | 6 | 3,90 | 3,80 | 0,10 | .295 |
| child 7 | 6 | 3,82 | 3,91 | -0,10 | .335 |
| child 2 | 6 | 3,57 | 3,67 | -0,10 | .347 |
| child 33 | 6 | 3,87 | 3,94 | -0,07 | .406 |
| child 19 | 6 | 4,01 | 4,04 | -0,03 | .416 |
| child 27 | 6 | 4,05 | 3,99 | 0,06 | .46 |

| Participant | Age | Condition | | Difference | Statistics |
|---|---|---|---|---|---|
| | (months) | Alternating | Non-alternating | Alt minus non-alt | *p_adj* |
| child 8 | 6 | 3,77 | 3,78 | -0,01 | .524 |
| child 16 | 6 | 4,02 | 4,15 | -0,13 | .56 |
| child 1 | 6 | 3,97 | 3,87 | 0,10 | .603 |
| child 13 | 6 | 3,84 | 3,82 | 0,02 | .665 |
| child 20 | 6 | 3,96 | 3,78 | 0,19 | .675 |
| child 21 | 6 | 3,55 | 3,47 | 0,07 | .675 |
| child 32 | 6 | 3,72 | 3,66 | 0,06 | .723 |
| child 23 | 6 | 3,79 | 3,76 | 0,03 | .725 |
| child 6 | 6 | 4,09 | 4,05 | 0,03 | .748 |
| child 24 | 6 | 4,21 | 4,12 | 0,09 | .773 |
| child 36 | 6 | 3,99 | 3,91 | 0,08 | .847 |
| child 5 | 6 | 3,70 | 3,73 | -0,03 | .85 |
| child 12 | 6 | 4,19 | 4,23 | -0,05 | .857 |
| child 9 | 6 | 3,79 | 3,82 | -0,02 | .899 |
| child 34 | 6 | 3,88 | 3,92 | -0,04 | .905 |
| **child 9** | **8** | **4,42** | **3,70** | **0,72** | **.001** |
| **child 7** | **8** | **3,76** | **3,30** | **0,46** | **.001** |
| **child 20** | **8** | **3,97** | **3,49** | **0,48** | **.022** |
| **child 15** | **8** | **3,94** | **3,55** | **0,38** | **.031** |
| child 38 | 8 | 3,43 | 3,51 | -0,08 | .051 |
| child 19 | 8 | 3,95 | 3,74 | 0,21 | .053 |
| child 10 | 8 | 4,01 | 3,73 | 0,28 | .057 |
| child 27 | 8 | 4,20 | 4,00 | 0,20 | .062 |
| child 35 | 8 | 4,36 | 3,96 | 0,40 | .067 |
| child 17 | 8 | 4,30 | 4,11 | 0,19 | .092 |
| child 40 | 8 | 4,15 | 3,78 | 0,37 | .098 |
| child 29 | 8 | 4,24 | 4,02 | 0,22 | .142 |
| child 5 | 8 | 4,13 | 4,00 | 0,13 | .144 |

| Participant | Age | Condition | | Difference | Statistics |
|---|---|---|---|---|---|
| | (months) | Alternating | Non-alternating | Alt minus non-alt | *p_adj* |
| child 11 | 8 | 3,82 | 3,47 | 0,35 | .153 |
| child 25 | 8 | 3,88 | 3,72 | 0,16 | .160 |
| child 6 | 8 | 3,72 | 3,54 | 0,18 | .160 |
| child 12 | 8 | 3,85 | 3,70 | 0,15 | .202 |
| child 13 | 8 | 3,82 | 3,97 | -0,15 | .242 |
| child 41 | 8 | 3,68 | 3,95 | -0,27 | .254 |
| child 8 | 8 | 3,68 | 4,04 | -0,36 | .294 |
| child 16 | 8 | 4,25 | 4,00 | 0,25 | .319 |
| child 36 | 8 | 4,06 | 3,94 | 0,12 | .332 |
| child 3 | 8 | 3,92 | 3,80 | 0,12 | .354 |
| child 18 | 8 | 3,90 | 3,69 | 0,21 | .387 |
| child 23 | 8 | 4,04 | 3,85 | 0,19 | .397 |
| child 26 | 8 | 3,84 | 3,69 | 0,15 | .420 |
| child 39 | 8 | 3,79 | 3,59 | 0,20 | .440 |
| child 31 | 8 | 4,18 | 4,03 | 0,15 | .483 |
| child 4 | 8 | 4,12 | 3,98 | 0,13 | .499 |
| child 1 | 8 | 3,80 | 4,04 | -0,24 | .592 |
| child 33 | 8 | 3,88 | 3,71 | 0,17 | .612 |
| child 21 | 8 | 4,23 | 4,19 | 0,04 | .672 |
| child 2 | 8 | 3,70 | 3,70 | 0,01 | .692 |
| child 14 | 8 | 3,53 | 3,60 | -0,07 | .712 |
| child 22 | 8 | 3,87 | 3,87 | 0,00 | .716 |
| child 32 | 8 | 3,89 | 4,01 | -0,12 | .728 |
| child 44 | 8 | 3,81 | 3,88 | -0,07 | .745 |
| child 30 | 8 | 3,70 | 3,80 | -0,10 | .768 |
| child 43 | 8 | 3,48 | 3,54 | -0,05 | .786 |
| child 28 | 8 | 3,81 | 3,78 | 0,03 | .904 |
| child 37 | 8 | 4,13 | 4,22 | -0,08 | .909 |

| Participant | Age | Condition | | Difference | Statistics |
|---|---|---|---|---|---|
| | (months) | Alternating | Non-alternating | Alt minus non-alt | *p_adj* |
| child 34 | 8 | 3,55 | 3,66 | -0,11 | .925 |
| child 42 | 8 | 3,74 | 3,75 | -0,01 | .937 |
| child 24 | 8 | 4,12 | 3,87 | 0,25 | .947 |
| **child 20** | **10** | **4,14** | **3,52** | **0,62** | **.001** |
| **child 34** | **10** | **4,23** | **3,88** | **0,36** | **.003** |
| **child 22** | **10** | **4,15** | **3,66** | **0,49** | **.005** |
| **child 24** | **10** | **3,96** | **3,67** | **0,29** | **.014** |
| **child 30** | **10** | **3,85** | **3,53** | **0,32** | **.016** |
| **child 32** | **10** | **4,01** | **3,85** | **0,16** | **.018** |
| **child 31** | **10** | **4,04** | **3,54** | **0,50** | **.020** |
| **child 8** | **10** | **4,03** | **3,76** | **0,26** | **.043** |
| child 9 | 10 | 3,70 | 3,54 | 0,16 | .076 |
| child 25 | 10 | 3,98 | 3,66 | 0,32 | .096 |
| child 14 | 10 | 3,67 | 3,52 | 0,15 | .129 |
| child 28 | 10 | 3,97 | 3,75 | 0,22 | .155 |
| child 11 | 10 | 3,57 | 3,45 | 0,12 | .195 |
| child 10 | 10 | 3,93 | 3,82 | 0,11 | .197 |
| child 12 | 10 | 4,04 | 3,83 | 0,20 | .219 |
| child 19 | 10 | 3,57 | 3,81 | -0,23 | .262 |
| child 2 | 10 | 3,99 | 3,86 | 0,13 | .266 |
| child 4 | 10 | 4,03 | 3,79 | 0,25 | .29 |
| child 7 | 10 | 3,97 | 3,82 | 0,15 | .306 |
| child 16 | 10 | 3,81 | 3,97 | -0,15 | .327 |
| child 5 | 10 | 3,93 | 3,81 | 0,11 | .344 |
| child 3 | 10 | 3,84 | 3,93 | -0,09 | .395 |
| child 18 | 10 | 3,51 | 3,37 | 0,14 | .420 |
| child 35 | 10 | 4,15 | 4,13 | 0,02 | .520 |
| child 15 | 10 | 3,60 | 3,72 | -0,12 | .592 |

| Participant | Age | Condition | | Difference | Statistics |
|---|---|---|---|---|---|
| | (months) | Alternating | Non-alternating | Alt minus non-alt | *p_adj* |
| child 27 | 10 | 3,65 | 3,52 | 0,12 | .599 |
| child 29 | 10 | 3,67 | 3,77 | -0,10 | .601 |
| child 21 | 10 | 3,87 | 3,84 | 0,03 | .641 |
| child 23 | 10 | 3,99 | 3,85 | 0,14 | .734 |
| child 1 | 10 | 3,83 | 3,71 | 0,12 | .832 |
| child 17 | 10 | 3,90 | 3,93 | -0,03 | .891 |
| child 6 | 10 | 3,90 | 3,81 | 0,09 | .899 |
| child 33 | 10 | 3,73 | 3,61 | 0,12 | .902 |
| child 13 | 10 | 3,44 | 3,46 | -0,02 | .955 |
| child 26 | 10 | 3,59 | 3,61 | -0,02 | .996 |
| **768** | **9 (Houston)** | **25800** | **8380** | **17420** | **.000** |
| 929 | 9 (Houston) | 11614 | 7843 | 3771 | .056 |
| 668 | 9 (Houston) | 12425 | 13060 | -635 | .336 |
| 762 | 9 (Houston) | 8671 | 6743 | 1928 | .529 |

*Note.* In the column *p_adj* the *p* values are reported for condition (alternating vs. non-alternating) in the autoregressive analyses of each infant. Houston (rows at the bottom) reports on raw looking time data received from Derek Houston (personal communication) which we were able to replicate with our model. Numbers in **bold** are significant (alpha level .05).

## Appendix B (Chapter 4)

*Number of participants, divided by habituation stimulus and the contrast order in the native condition*

| Age Group | Group | Participants | Habituation stimulus /fa:p/ | Native first |
|---|---|---|---|---|
| | *no-FR* | | | |
| 6 | | 28 | 13/28 | 16/28 |
| 8 | | 24 | 9/24 | 15/24 |
| 10 | | 27 | 20/27 | 15/27 |
| Total | | 79 | 42/79 | 46/79 |
| | *FR* | | | |
| 6 | | 28 | 13/28 | 16/28 |
| 8 | | 24 | 6/24 | 14/24 |
| 10 | | 27 | 13/27 | 8/27 |
| Total | | 79 | 32/79 | 38/79 |

*Note. no-FR* refers to no family risk of dyslexia, *FR* to infants at family risk of dyslexia. *Native first* refers to the number of participants who received the native contrast first during the test session.

## Appendix C (Chapter 4)

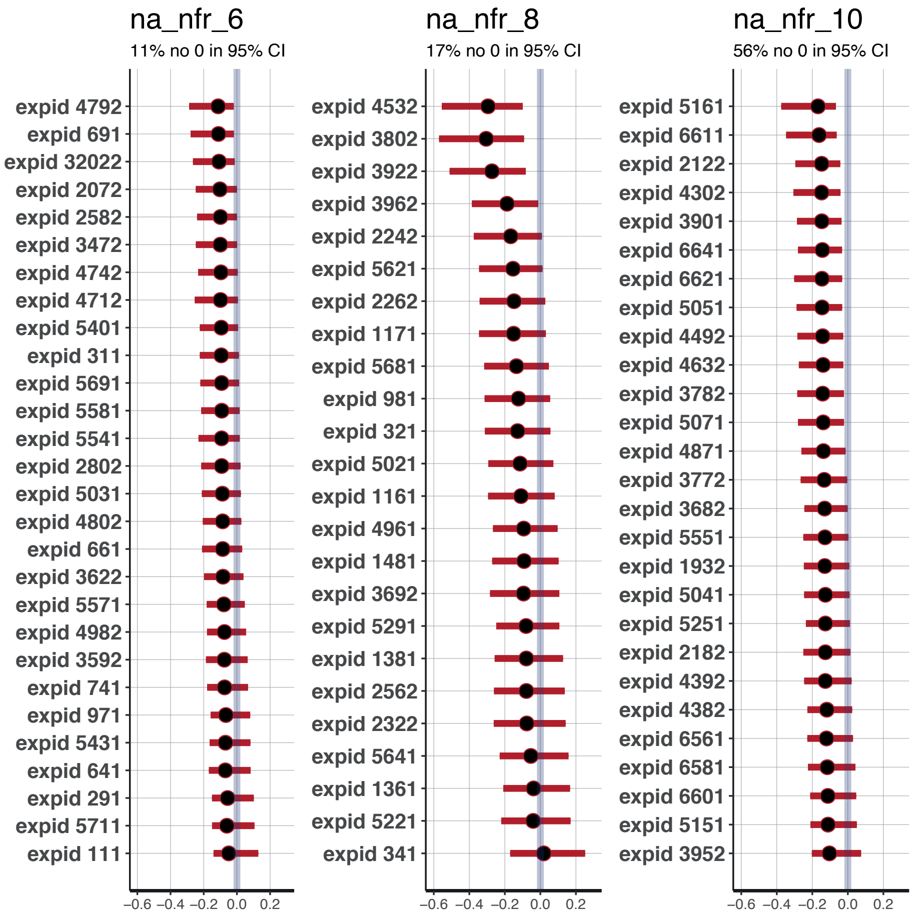
*Number of participants, divided by habituation stimulus and the contrast order in the non-native condition*

| Age Group | Group | Participants | Habituation stimulus /sæn/ | Non-native first |
|---|---|---|---|---|
| | *no-FR* | | | |
| 6 | | 31 | 18/31 | 18/31 |
| 8 | | 27 | 18/27 | 16/27 |
| 10 | | 26 | 14/26 | 17/26 |
| Total | | 84 | 50/84 | 51/84 |
| | *FR* | | | |
| 6 | | 31 | 17/31 | 17/31 |
| 8 | | 27 | 19/27 | 15/27 |
| 10 | | 26 | 11/26 | 24/26 |
| Total | | 84 | 47/84 | 56/84 |

*Note. no-FR* refers to no family risk of dyslexia, *FR* to infants at family risk of dyslexia. *Non-native first* refers to the number of participants who received the non-native contrast first during the session.

**Appendix D (Chapter 4)**
Individual and group estimates for the **native** (na) contrast and the **no-FR** (nfr) infants

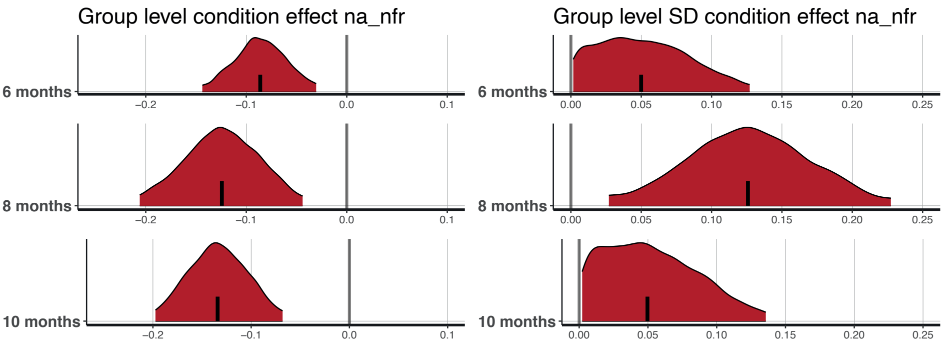
Figure D1
*Results of the hierarchical model for each individual per age group.*




*Note.* The black dots represent the mean; the red bars represent the 95% credibility intervals.

Figure D2

*Group estimates for condition effects and variation per age group*



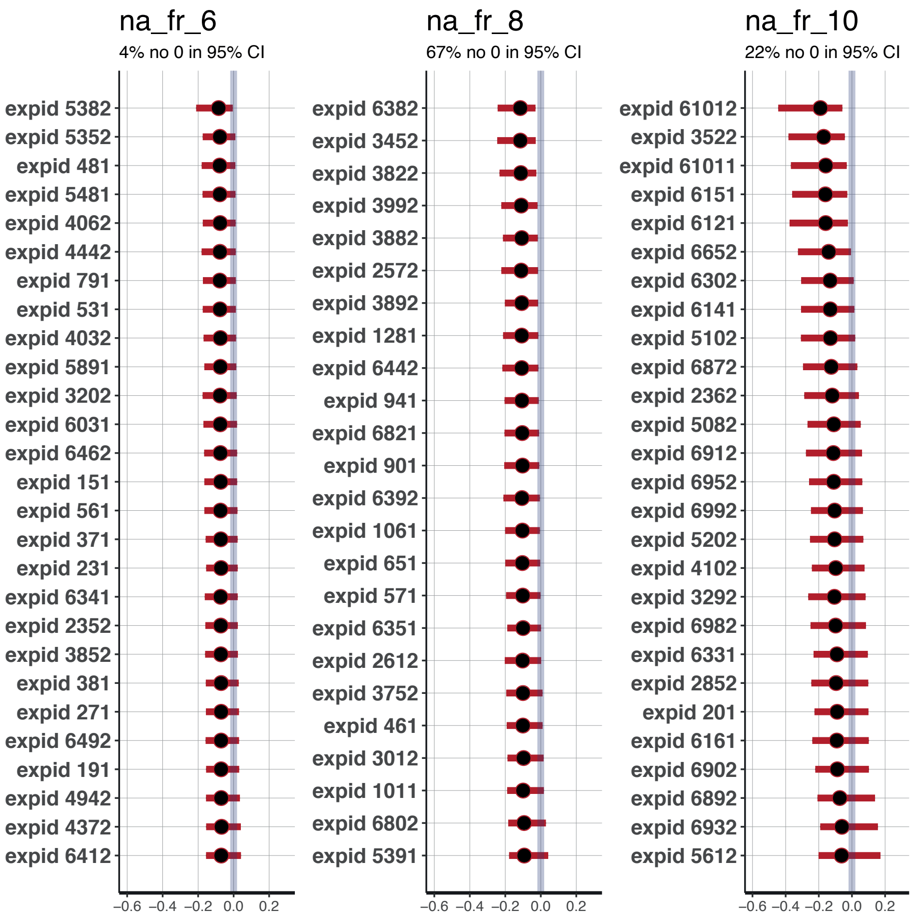

*Note.* The left panel shows the group estimates for condition effects. The right panel shows the standard deviation of the condition effect per age group. The densities, presented in red, represent the 95% credibility interval.

## Appendix E (Chapter 4)

Individual and group estimates for the **native** (na) contrast and **FR** (fr) infants

Figure E1

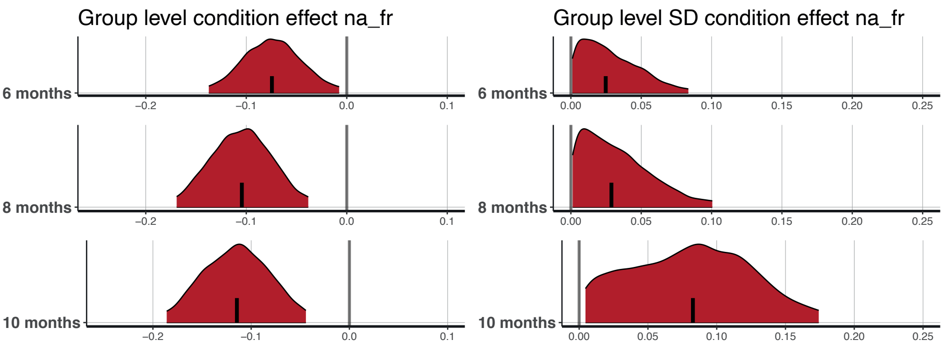
*Results of the hierarchical model for each individual per age group.*




*Note.* The black dots represent the mean; the red bars represent the 95% credibility intervals.

Figure E2

*Group estimates for condition effects and variation per age group*



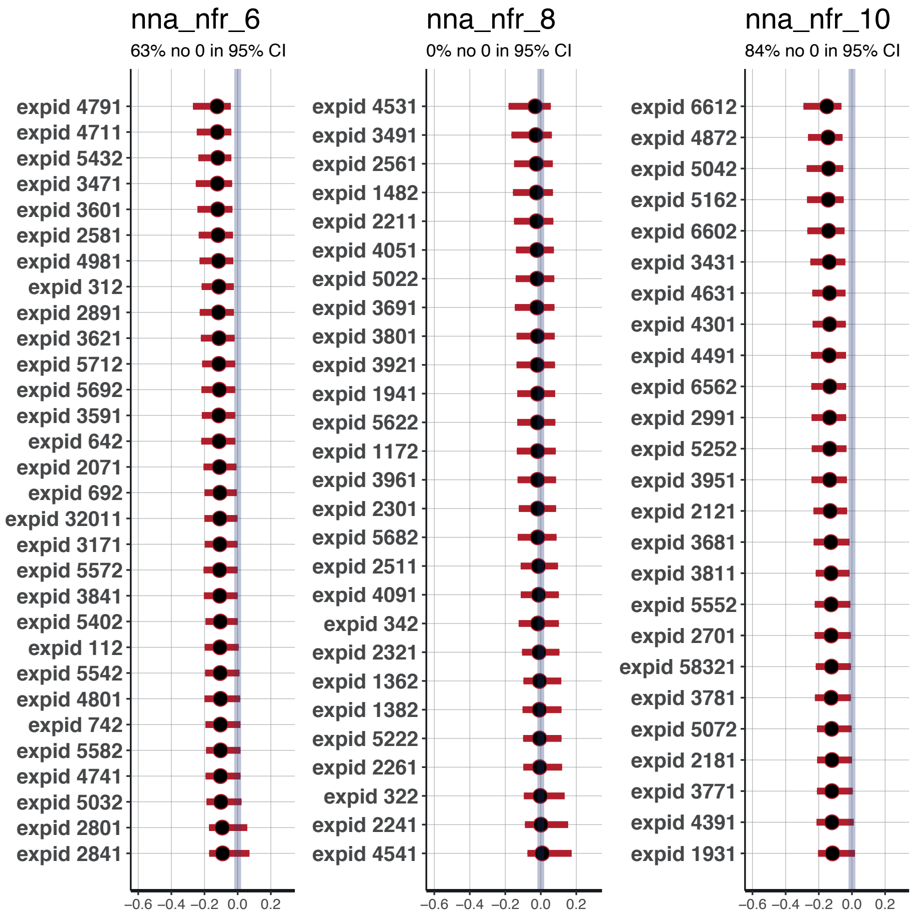

*Note.* The left panel shows the group estimates for condition effects. The right panel shows the standard deviation of the condition effect per age group. The densities, presented in red, represent the 95% credibility interval.

**Appendix F (Chapter 4)**

Individual and group estimates for the **non-native** (nna) contrast and the **no-FR** (nfr) infants

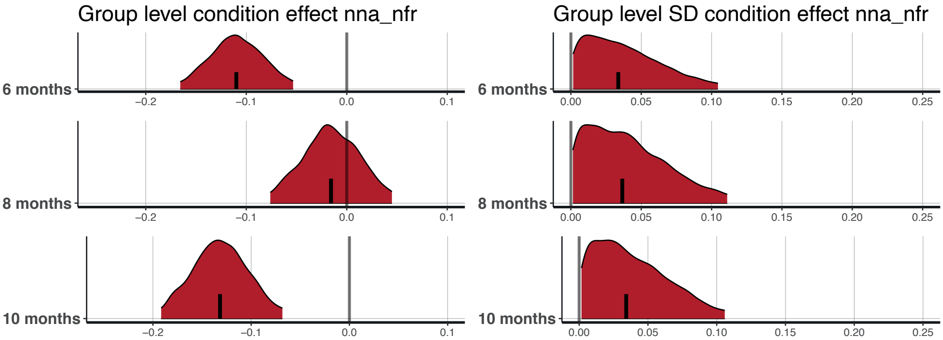
Figure F1

*Results of the hierarchical model for each individual per age group.*




*Note.* The black dots represent the mean; the red bars represent the 95% credibility intervals.

Figure F2
*Group estimates for condition effects and variation per age group*



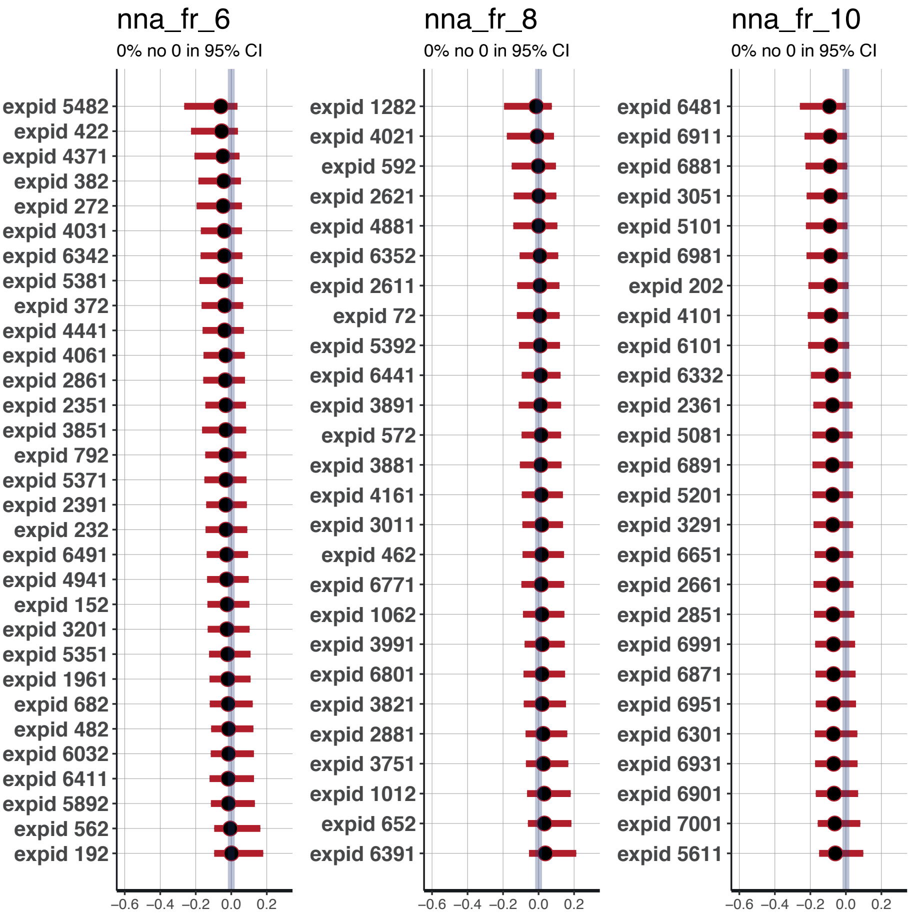

*Note.* The left panel shows the group estimates for condition effects. The right panel shows the standard deviation of the condition effect per age group. The densities, presented in red, represent the 95% credibility interval.

## Appendix G (Chapter 4)

Individual and group estimates for the **non-native** (nna) contrast and **FR** (fr) infants

Figure G1

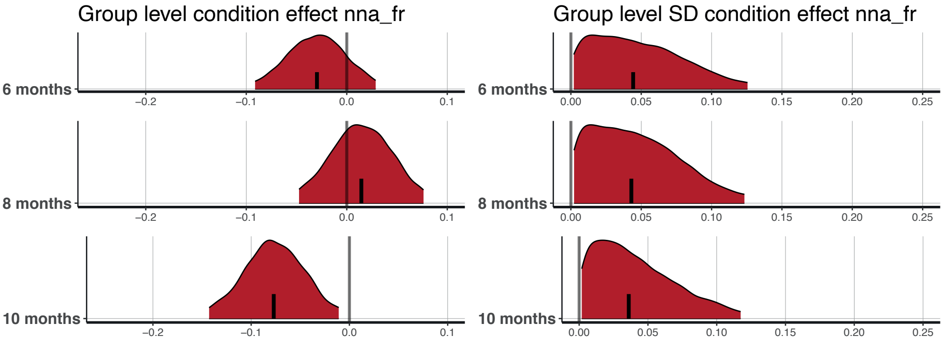
*Results of the hierarchical model for each individual per age group.*




*Note.* The black dots represent the mean; the red bars represent the 95% credibility intervals.

Figure G2

*Group estimates for condition effects and variation per age group*




*Note.* The left panel shows the group estimates for condition effects. The right panel shows the standard deviation of the condition effect per age group. The densities, presented in red, represent the 95% credibility interval.

# Dankwoord (acknowledgements)

Gedreven door passie en onderhouden door doorzettingsvermogen; het schrijven van een proefschrift is een mentale marathon. In mijn geval zonder een van tevoren vastgestelde eindtijd. Het is (ongeveer) tien jaar geworden…

Een groot doel behaal je niet alleen maar met hulp van mensen die de voorwaarden scheppen om eraan te beginnen maar vooral ook om het vol te houden. Voor mij zijn er een aantal mensen heel belangrijk geweest bij het nastreven van mijn doel. Op de eerste plaats zijn dat mijn beide promotoren, Frank Wijnen en Elise de Bree. Zonder hen was ik er nooit aan begonnen en zonder hen was het boek (en niet te vergeten de drie publicaties) nooit zo goed geworden. Frank, *liesten very carefoellie I shall say dies only once*: Ik wil jou enorm en met heel mijn hart bedanken voor je begeleiding, waarin je mij de ruimte hebt gegeven om mijn eigen weg en stem te vinden. Ik wil je danken voor het vertrouwen dat je mij hebt gegeven en voor de duwtjes in de rug om te doen wat ik heel leuk, maar ook heel spannend vond. Dank voor de keren dat je mij het bos weer hebt laten zien, wanneer ik alleen nog maar de bomen zag. Heel veel dank ook voor al je geduld, je wijsheid en je grapjes(!). Dank voor je liefde voor de taal en voor de psycholinguïstiek in het bijzonder. Ze was erg aanstekelijk.
Elise, jou wil ik bedanken voor je tomeloze optimisme, scherpzinnigheid en humor. Je bent de allerliefste, slimste, grappigste, vrolijkste en energiekste (co) promotor die ik mij kan voorstellen. Ook jij weet als geen ander hoe weer tot de kern van het verhaal te komen, om van daaruit weer verder te gaan. Je gaf richting, zonder te wijzen. Je motiveerde wanneer nodig en wanneer het vuurtje brandde, wakkerde jij het verder aan. Nooit te moe om een stuk nog een keer te lezen of om over data/resultaten te praten. Ik kan mij geen betere (co)promotor wensen. Aan het begin van mijn traject was jij, Eliza, mijn copromotor, maar nu ben je BIJZONDER HOOGLERAAR en dus ook mijn promotor. Trots op jou, fantastische vrouw en lieve vriendin!
Annemarie Kerkhoff, ik wil je bedanken voor jouw bijdrage tijdens de eerste, maar o zo belangrijke fase, van het project. Het eerste artikel – en dus eigenlijk ook de twee die daarop volgden – was zonder jouw kritische vragen niet geworden wat het nu is. We hebben een tijdje niet meer samengewerkt, maar dat is gelukkig nu weer anders! Hoe mooi kan het soms lopen :-). Kunnen we weer lekker koffieleuten en kletsen over werk (enzo…).

Na mijn afstuderen in 2008 vroeg Frank of ik interesse had om als coördinator van het Babylab bij het UiL OTS (Utrecht Institute of Linguistics, Onderzoeksschool Taal en Spraak) aan de slag te gaan. Het Babylab was in 2008 nog een kleinschalig lab. Het bestond voornamelijk uit experimentele studies die voortkwamen uit het project *Category Formation*[1] van Frank waarbij Elise en Annemarie betrokken waren als postdocs. Dit project werd het uitgangspunt van mijn eigen traject.
Ik heb geen traditioneel PhD-pad bewandeld. Mijn onderzoek combineerde ik met mijn werk voor het Babylab en de ethische toetsingscommissie van UiL OTS/Faculteit GW. Ik had dus ook niet echt mede-PhD'ers. Behalve Ao Chen en Liquan Liu. Dit kwam vooral omdat zij als enige PhD's altijd in het lab aan het werk waren, waar anderen op de Trans werkten en alleen naar het lab op het Janskerkhof kwamen wanneer zij een experiment moesten draaien. Ao en Liquan wilden samen met Iris Mulders en *me-myself-and-I* in de diepe, koele en donkere (soms naar gracht meurende) kelders van het lab werken. Iris en ik deelden jaren een kamer, maar wanneer ik aan mijn onderzoeksproject werkte, zat ik bij Ao en Liquan op de kamer. Daar zaten wij úren te werken, verdiept in de statistiek, met soms zachtjes een Chinees muziekje op de achtergrond. Liquan die altijd met een of ander gezond proteïnerijk drankje in de weer was en Ao die tussendoor aan het chatten was via een Hello Kitty-kleurig platform. Ao en Liquan, jullie wonen niet meer om de hoek, maar wel nog in mijn hart!

Samen met Iris en de technische jongens (eerst Theo, later Martijn, Jan, Chris, Maarten, Jacco en Ty) ondersteunde ik de onderzoekers die gebruik wilden maken van de labs van het Uil. Ik hield mij hoofdzakelijk bezig met het Babylab en Iris vooral níet met het Babylab. Die vaak niet-meewerkende proefpersoontjes waren niet haar favoriete doelgroep. Iris, ik mis je nog steeds. Hopelijk vinden we elkaar straks weer één dag in de week op de Drift: de beste werkplek ever!
Het Babylab begon klein, maar breidde snel uit, waardoor de werkwijze moest veranderen. Waar ik eerst álle baby's testte, kon dat niet langer zo doorgaan. Het Babylab moest een plek worden waar studentes echte onderzoekservaring op konden doen. Waar ze zouden meewerken aan alle facetten die komen kijken bij het verzamelen van data. Een plek waar ze een waardevolle bijdrage zouden leveren. Die studentes, ja echt alleen van het vrouwelijke geslacht, kwamen er en o(!) wat maakten zij van mijn werk een feestje. Ik heb genoten van het begeleiden van stages en scripties en al het coördinerende werk dat erbij kwam kijken. Het is dankzij deze studentes dat mijn project niet nog véél langer geduurd heeft. Ik ga ze niet allemaal opnoemen, maar ze staan wel in Figuur 1 ;-).
Een aantal van die toen-nog-meisjes die wat langer zijn blijven hangen en waar ik een bijzonder leuke tijd mee heb gehad, wil ik wel even apart noemen. Lisanne

Geurts, jij hebt een tijdje een gedeelte van mijn werkzaamheden als labmanager écht overgenomen waardoor ik eindelijk de tijd en ruimte had om meer dan één doordeweekse dag te werken aan mijn project!! In dit kader wil ik ook Willemijn Doedens, Lorijn Zaadnoordijk en Cora Pots noemen. Deze drie zijn trouwens eerder gepromoveerd dan ik, terwijl ze als student bij mij gewerkt hebben, hahaha! Goed gedaan, meiden! Ook verdient Sule Kurtçebe hier een ereplekje. Zij heeft ontelbare baby's getest en daarnaast o.a. ook nog monnikenwerk verricht met het digitaliseren van de NCDI's! Niet te doen was dat, maar goed. Dankjewel daarvoor! Tot slot wil ik ook Charlotte Koevoets bedanken voor haar inzet en gezelligheid(!). Hopelijk werken we straks weer even samen aan de Maye-studie. Héél misschien…
Wat zeker ook bijdroeg aan de feestvreugde tijdens het werk waren natuurlijk de vrijdagse Babylab-meetings met (niet allemaal tegelijk): Frank, Elise, Annemarie, Rob Zwitserlood, Lizet van Ewijk, Liquan, Ao, Desiree Capel, Brigitta Keij, Mengru Han, Carolien van den Hazelkamp, Ileana Grama, Silvia Radulescu, Caroline Junge en alle studentes die hierboven en in Figuur 1 staan en alle anderen die ik wellicht vergeten ben te noemen. Dank-jullie-wel voor jullie wijsheid, kritische vragen en al het andere, maar zeker ook de humor. Het was heel fijn om de toch vaak wel ingewikkelde stof te kunnen afwisselen met luchtige, onzinnige grapjes. Frank en ik konden erg lachen om onze spraakimitaties (vooral tooncontrasten vielen bij óns in de smaak) of non-woordverzinsels.
Het groeien van het Babylab en het voltooien van mijn proefschrift, was natuurlijk nooit mogelijk geweest zonder de inzet van al die ouders met hun prachtige baby's: ontzettend bedankt voor het meedoen!

Brigitta en Britt, mijn paranimfen, mijn oud-collega's, maar ook mijn huidige collega's. Veel dank voor het regelen van alle praktische zaken rondom de promotie en de mentale steun in de aanloop naar de verdediging! Ook wil ik jullie, en Desiree, Annemarie en Elise, bedanken voor het voorbereiden van en het meedoen aan mijn proefpromotie.

Mijn thuisbaken. Mijn man, Anton, en mijn drie mooie jongens, Jonne, Louis en Benjamin zijn diegenen die – al dan niet bewust – het meest hebben geleden onder dit krankzinnig langdurende project. Dankbaar ben ik dat ik hen mijn liefde mag geven en ik zo veel liefde terugkrijg. Er zijn in die afgelopen tien jaar best wat momenten geweest dat ik er niet was. Dat ik op de Drift aan het werk was, totaal verdiept in mijn werk zonder besef van tijd. Dat Anton dan belde en begreep dat ik daar nog even bleef. Dat ik in het weekend des morgens vroeg weg ging en zij mij pas aan het einde van de dag weer zagen. Anton, *min stora kärlek,* *tack från hjärtat!* Jou wil ik bedanken voor je eindeloze steun, je optimisme en voor ál die keren dat je mij weer aan de praat hebt gekregen. Voor je meevoelen en je opbeurende woorden. Je woorden landdden niet altijd, dat lag niet aan jou, maar aan mij. De tijd, of iets anders ongrijpbaars, loste de donderwolk weer op. Dat wist je, na een tijdje. Gewoon even laten, dan staat ze zo wel weer rechtop. Andere keren, wanneer ik een dag werk kwijt was (dit snappen alleen mensen die bekend zijn met het digitale werken) omdat ik het toch niet goed had opgeslagen, of wanneer Surfdrive niet gesynchroniseerd (!!) had, dan bleef jij altijd kalm en geduldig aan de lijn, mij leidend (en lijdend) door het systeem. Je zou denken, dit gebeurt je eens maar nooit weer…:

‘We zullen doorgaan, als niemand meer verwacht
Dat we weer doorgaan, in een sprakeloze nacht
We zullen doorgaan
We zullen doorgaan
We zullen doorgaan
Tot we samen zijn’[2]

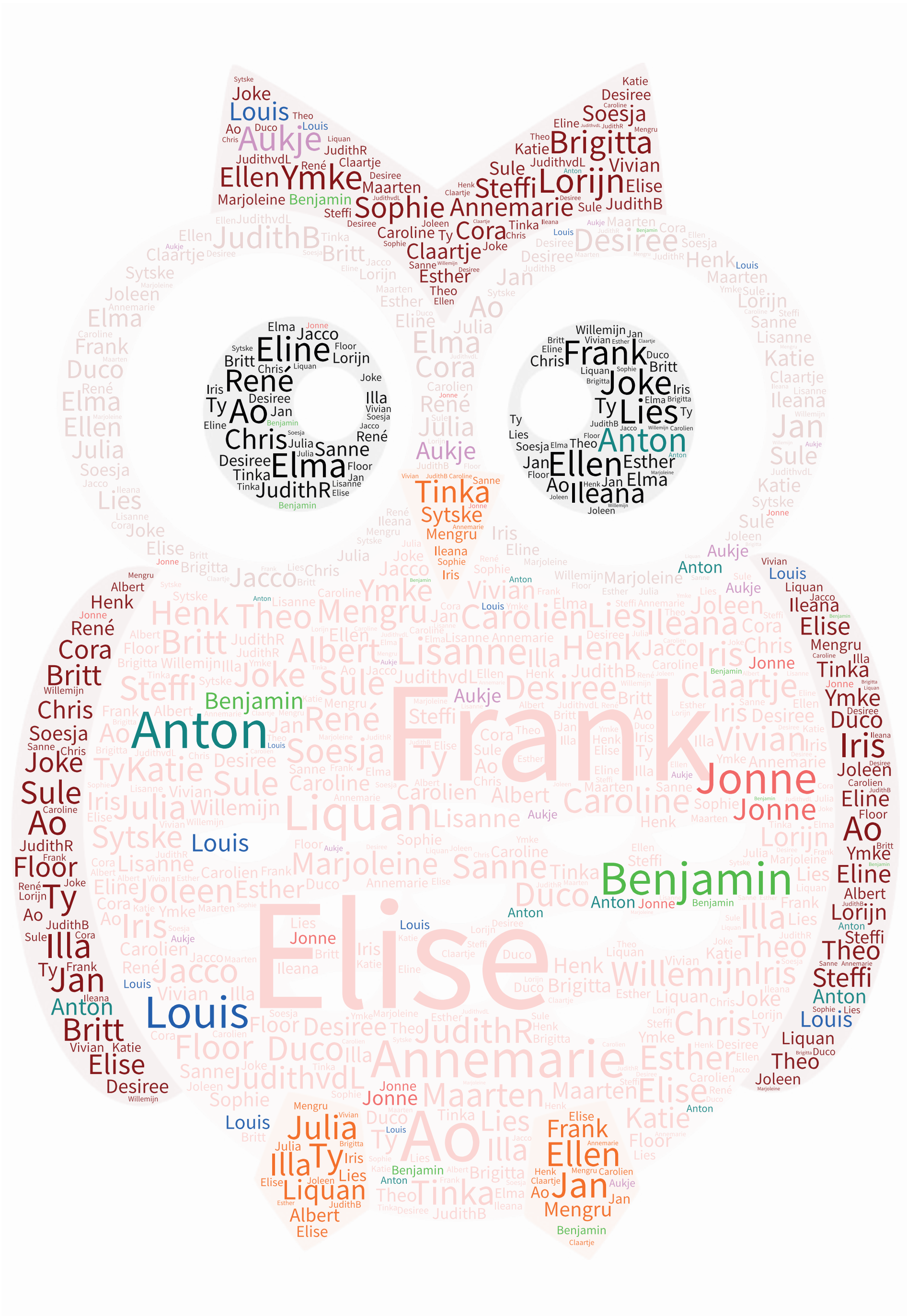
Jonne, Louis en Benjamin. Mijn allerliefste en mooiste mannen. Ik hou van jullie. Dat weten jullie wel, want dat zeggen we best vaak tegen elkaar, maar nu staat het ook lekker in een boek :-). Ik hoop dat jullie hebben meegekregen, of ooit nog zullen meekrijgen, dat je met doorzettingsvermogen heel ver kunt komen. Natuurlijk moet je ook een beetje ‘iets’ kunnen, máár: uitkomst = inzet + talent. Deze wijsheid komt van juffie Anna, maar misschien ook wel van een tegeltje, of anders kan het er een worden. Kijk naar mij: veel inzet en ook een beetje talent, met als uitkomst? Dit boek. Laat je nooit vertellen dat je iets niet kunt, tenzij je (eigenlijk) weet dat ze gelijk hebben. Blijf zelfkritisch en nieuwsgierig. En vooral, als je iets écht wilt, geef dan vooral niet op! (Benjamin, dat komt bij jou wel goed; iets minder mag ook wel af en toe, hahaha). Lieve Jonne, Louis en Benjamin, voorin het boek staat een stukje uit een gedicht van Babs Gons, op blz. 5, lees het maar even. Het zijn een paar regels uit een veel langer gedicht, dat in de kern gaat over het volgen van je hart. Doe dat maar gewoon, omdat je uiteindelijk toch niet anders kan. Of ongelukkig wordt. Dus luister naar je hart. En naar je moeder.
Bij het thuisfront, mijn achterban, horen ook Joke en Henk, mijn mamma en pappa. Jullie hebben mij altijd mijn eigen keuzes laten maken. Ook wanneer jullie het er niet mee eens waren. Dankjewel voor het rotsvaste vertrouwen in mijn kunnen en voor het vertrouwen dat jullie er altijd voor mij zullen zijn. En ik voor jullie.
Tot slot, Sytske en Albert, jullie wil ik bedanken voor het beschikbaar stellen van

het fijne schrijfhuisje in jullie tuin aan het bos. Het is inmiddels een welbekend huisje onder mijn collega's, gezien over het algemeen de onlinevergaderingen daar plaatsvinden. Het schrijfhuisje is nu toe aan een nieuw boek. Een kinderboek? Een dichtbundel? Een dichtbundel voor kinderen? Wie weet. Ook mijn lieve vriendin Lies wil ik bedanken. Met heel veel geduld en liefde heeft ze samen met mij dit boek mooi gemaakt. Daar zijn heel wat heen-en-weertjes voor nodig geweest. Toch nog kleine foutjes, andere kleurtjes, plaatjes in een hogere resolutie, kleine tekstuele aanpassingen. Toch niet ieder inhoudswoord van de titel met een hoofletter: f*ck you APA-stijl, zo is het mooier! En wat dies meer zij (genoeg, het is een *never-ending story* ben ik bang). Maar BOEM. Daar is-ie. Het is af.

De mentale race is gelopen. De finish gehaald. Ik zou wel in het bekende gat vallen, zei men nadat het manuscript was goedgekeurd. Weg doel. Nou, nee hoor. Er is alweer een nieuwe in het vizier. En niet zomaar één: de halve marathon van Egmond. Okay, misschien niet die van januari 2022. Maar dan toch zeker wel het jaar daarop!

[1] Category formation in phonology and grammar: distributional learning in children with and without a developmental language delay (2007-2011; NWO Humanities Open Competition)

[2] Ramses Shaffy - We zullen doorgaan


Elise
Frank
Anton
Benjamin
Jonne
Louis
Ao
Ellen
Eline
Elma
Chris
René
Joke
Lies
Ileana
Anton
Tinka
Sytske
Mengru
Iris
Aukje
Ymke
Brigitta
Lorijn
Steffi
Sophie
Annemarie
Cora
Claartje
Esther
Theo
Soesja
Desiree
Katie
Vivian
Sule
JudithB
JudithR
JudithvdL
Marjoleine
Maarten
Caroline
Carolien
Liquan
Jacco
Henk
Albert
Britt
Floor
Duco
Illa
Ty
Jan
Julia
Joleen
Lisanne
Sanne
Willemijn
Ellen
Theo
Iris

# List of publications

**This Dissertation**

**de Klerk, M.**, de Bree, E., Kerkhoff, A., & Wijnen, F. (2019). Lost and found: decline and reemergence of non-native vowel discrimination in the first year of life. *Language Learning and Development*, *15*(1), 14-31. https://doi.org/10.1080/15475441.2018.1497490

**de Klerk, M.**, de Bree, E., Veen, D., & Wijnen, F. (2021). Speech discrimination in infants at family risk of dyslexia: Group and individual-based analyses. *Journal of Experimental Child Psychology, 206,* 105066. https://doi.org/10.1016/j.jecp.2020.105066

**de Klerk, M.**, de Bree, E., & Wijnen, F. (in revision). Distributional Phonetic Learning in Infants at Family Risk of Dyslexia.

**de Klerk, M.**, Veen, D., Wijnen, F., & de Bree, E. (2019). A step forward: Bayesian hierarchical modelling as a tool in assessment of individual discrimination performance. *Infant Behavior and Development*, *57*, 101345. https://doi.org/10.1016/j.infbeh.2019.101345

**Other collaborations**

Capel, D., de Bree, E., **de Klerk, M.**, Kerkhoff, A., & Wijnen, F. (2011). Distributional cues affect phonetic discrimination in Dutch infants. *Sound and sounds. Studies presented to MEH (Bert) Schouten on the occasion of his 65th birthday*, pp.33-43. Utrecht: UiL-OTS, LOT Publications.

Junge, C., Everaert, E., Porto, L., Fikkert, P., **de Klerk, M.**, Keij, B., & Benders, T. (2020). Contrasting behavioral looking procedures: a case study on infant speech segmentation. *Infant Behavior and Development*, *60*, 101448. https://doi.org/10.1016/j.infbeh.2020.101448

Kerkhoff, A., de Bree, E., **de Klerk, M.** & Wijnen, F. (2013). Non-adjacent dependency learning in infants at familial risk of dyslexia. *Journal of Child Language*, *40*(1), 11-28. https://doi.org/10.1017/S0305000912000098

The ManyBabies Consortium (2020). Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52. https://doi.org/10.1177/2515245919900809

# About the author

Maartje de Klerk was born on May 13, 1976, in De Bilt, the Netherlands. She completed her secondary school at De Werkplaats Kindergemeenschap (Kees Boekeschool), Bilthoven, in 1994. After obtaining her bachelor's degree in *Nederlandse Taal & Cultuur* (Dutch Language and Culture) at Utrecht University, she started the master *Taal & Spraak: Verwerking en Stoornissen* (Langauge & Speech: Processing and Disorders) at the same university in 2008. After graduating she worked as the babylab manager at Babylab Utrecht (2008-2019) and as the secretary of the ethics assessment committee of the Uil OTS of the Utrecht University (2014-2019). In 2012 she also started her PhD at the department of *Talen, Literatuur en Communicatie* (Languages, Literature and Communication) of the Faculty of Humanities, Utrecht University. Currently Maartje works as a senior researcher at Koninklijke Auris Group. At Koninklijke Auris Group she is involved in several research projects aimed at stimulating language and communication skills in children with (presumed) developmental language disorders.

This dissertation investigates the development of speech sound categories in typically developing infants (NFR) and infants at family risk of dyslexia (FR). According to a broadly supported hypothesis, a core feature of developmental dyslexia is a phonological deficit. This deficit is argued to stem from poorly developed phonological representations. Speech sound categories are required for constructing phonological representations. The work reported in this dissertation seeks to establish if and how the acquisition of speech sound categories differs between NFR and FR infants.

Within the first year of life infants' perception changes from universal to language specific. This process of perceptual attunement is the first step into the formation of speech sound categories and is the focus of this dissertation. Next to group- and age-based comparisons, individual discrimination performance was assessed. If successful, this could be useful for longitudinal studies that aim to investigate the relation between early speech perception and later language and literacy skills.

The results of this dissertation indicate that the process of perceptual attunement is less predictable than often assumed and may well extend beyond the first year of life. Furthermore, FR infants have a (subtle) delayed development of speech sound categories compared to their peers. This could contribute to poorly developed phonological representations and is consistent with the theory of a phonological deficit. Finally, Bayesian hierarchical modeling can be used to classify infants' individual discrimination performance.

UiL OTS
UTRECHT INSTITUTE OF LINGUISTICS