PATTERNS OF PERCEPTUAL REORGANIZATION IN INFANCY: DECLINE, MAINTENANCE, AND U-SHAPED DEVELOPMENT

by **Antonia Götz**

Doctoral dissertation

submitted to the Faculty of Human Sciences at the University of Potsdam

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Cognitive Science

Under the supervision of

Prof. Barbara Höhle, University of Potsdam

Prof. Isabell Wartenburger, University of Potsdam

2020

Supervision: Prof. Dr. Barbara Höhle Wartenburger, Prof. Dr. Isabell Wartenburger
1st reviewer: Prof. Dr. Barbara Höhle
2nd reviewer: Prof. Dr. Bettina Braun
Day of submission: November 25, 2020
Thesis defense: Potsdam, April 12, 2021

https://doi.org/10.25932/publishup-53618
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-536185

Declaration of authorship

I hereby certify that the thesis I am submitting is entirely my own original work except where otherwise indicated. I am aware of the University's regulations concerning plagiarism, including those regulations concerning disciplinary actions that may result from plagiarism. Any use of the works by any other author, in any form, is properly acknowledged at their point of use.

Potsdam, November 2020 Antonia Götz

Acknowledgments

First and foremost, I would like to thank Prof. Barbara Höhle for her amazing support and the opportunity to write this thesis in the Crossing the Borders project, thus enabling me to explore the many possibilities and insights of developmental research. Thank you for the insights into this super exciting field of research, the feedback, and guidance through the processes. Also, my sincere thanks to Prof. Isabell Wartenburger for her consistent, valuable, and constructive suggestions in planning and developing this research work. Many thanks to both of you for your time and advice.

Also, it would not have been possible to continue to think and develop ideas, to get on new paths and change perspectives without the exchange with the members of the Psycho-Ling and Neuroteam. The regular meetings were an enrichment on many levels, both professional and personal. It helped to see that we sometimes face the same challenges and can support each other in doing so. Thank you all for the great cooperation! Natalie, Annika, Elina, Jule, Mireia thank you for the valuable discussions (and also coffee breaks). Tom, for the support, explanations with everything that has to do with the organization of BabyLab or the Isoldes.

What would the Golm time have been without you the *Mensagang*, on our *Mensagang*, along the *Mensagang*, to take our *Mensagang*. The thoughts about the whimsicalities of the German language have been burned into my head forever. Tusen takk, dank je wel, Danke Anne, Clara and Marc for that! Hoch die Kaffeetassen und Gläser zum Appell. Marie and Sascha: thanks for the incredibly good living situation!

Mama und Papa, vielen lieben Dank für die Unterstützung und Vertrauen! Vor allem danke ich euch, dass ihr niemals die Frage gestellt hat: „Und was willst du damit machen?!“

And last but not least, thank you to all those without whom this work would never have been possible. Thanks to the parents and the little participants who came to the BabyLab and looked left and right with full energy, had caps put on their heads, and didn't let themselves be distracted by our sometimes desperate attempts to get the technology going.

Content

List of figures

List of tables

CHAPTER 1

1.1 General introduction

Nothing seems to happen more easily and automatically than children learning to speak. Over the last 50 years, there has been an enormous increase in the knowledge of how children acquire languages. Some of the research has focused on how infants perceive stimuli that are or are not part of their visual or linguistic environment. In the visual domain, studies have investigated infants' perception of faces, objects, and colors (e.g., Fantz, 1964; Johnson et al., 1991). In the linguistic domain, the first studies focused on infants' abilities to perceive native and non-native speech contrasts (e.g., Eimas et al., 1971; Werker & Tees, 1984). Later, other studies investigated infants' early speech segmentation abilities (e.g., Jusczyk & Aslin, 1995) as well as word learning mechanisms (e.g., Stager & Werker, 1997). Furthermore, the link between both domains – the integration and interaction of visual and linguistic cues of infants' perception abilities – has been investigated in several studies (e.g., Burnham & Dodd, 2004; Lewkowicz, 1996). The studies in the visual and linguistic domains demonstrate how perception changes over time and how (language-)specific knowledge influences infants' perceptual sensitivity.

Finding suitable methods is a particular challenge in the investigation of language acquisition in infants. Infants cannot be asked whether they perceive a difference between sounds or the meanings of words. Researchers need(ed) to be creative. Advances in research were also made possible by integrating different methodologies and experimental techniques. Initial research began with behavioral studies using infants' sucking behavior (e.g., high-amplitude-sucking paradigms; Eimas, 1975; Eimas et al., 1971) and listening behavior (e.g., conditioned head turn paradigms; Werker & Tees, 1984). Later, methods that use listening times as a measure of infants' abilities were modified with respect to the initial exposure phase by using either a fixed initial exposure phase for familiarization procedures (e.g., Jusczyk & Aslin, 1995) or a variable initial exposure phase in habituation procedures (e.g., Stager & Werker, 1997) instead of head turns. More recently, eye-tracking paradigms as well as neuroscientific methods like electroencephalography (EEG; Friederici & Thierry, 2008) and functional near-infrared spectroscopy (fNIRS; Lloyd-Fox et al., 2010) have contributed to advancements in knowledge. Notably, these methodologies and experimental techniques differ in terms of constancy and stability at the group level (Cristia et al., 2016). Over the last decades, extensive research has been conducted in the field of phonological development in infancy using the aforementioned experimental methods. Previous research has produced a diverse spectrum of results. Among other findings, studies revealed that young infants discriminate between native and non-native sounds (Eimas et al., 1971; Werker & Tees, 1984). While the ability to discriminate between non-native speech sounds decreases in the first year of life, the ability to discriminate between native speech sounds increases (Anderson et al., 2003; Kuhl et al., 1992; Kuhl et al., 2006; Tsuji & Cristia, 2014; Werker & Tees, 1984). However, these developmental patterns are not consistent across all speech contrasts. For some speech contrasts, the ability to discriminate between non-native speech sounds is apparent in both infants and adults (Best et al., 1988; Liu & Kager, 2014; Polka & Bohn, 1996; Sundara et al., 2006). Furthermore, prior studies indicated that the order of presentation of speech contrasts could influence the discrimination of infants and adults (e.g., Polka & Bohn, 2011). Accordingly, for some contrasts infants might demonstrate good discrimination ability if they are first introduced to a sound X of a speech contrast X-Y (either in the form of a familiarization or habituation procedure) and then tested on the sound Y. However, they might not show any discrimination ability when tested in the reverse order – that is, if they are first presented with sound Y and then tested on sound X. In this case, they do not show perceptual sensitivity to the X-Y contrast.

1.2 The present dissertation

This dissertation contributes further to the understanding of the developmental trajectory of phonological acquisition in infancy. The aim is to examine the perceptual sensitivity of lexical tones and vowels in German-learning infants – an area in which previous research has produced especially diverging results. Of particular importance are the questions of how perceptual sensitivity in the developmental trajectory changes during infancy and how the various experimental procedures contribute to the understanding of infants' abilities. To shed light on the ability of German-learning infants to discriminate lexical tones and vowels, we conducted different studies in the context of this dissertation. Since it has already been shown that different procedures may influence the strength of discrimination abilities (Cristia et al., 2016), we used three experimental procedures to establish an elaborated understanding of the infants' overall perceptual sensitivity. In Studies 1 and 3 (Chapters 5 and 7), we used behavioral methods (habituation and familiarization procedures). In Study 2 (Chapter 6), we measured neural correlates. Having used the same tone contrast and varied the experimental method, we can draw profound conclusions about the infants' discrimination ability and the relevance of the experimental procedures for determining the infants' ability to discriminate certain speech sounds.

The thesis is structured as follows: Chapter 2 introduces empirical evidence of infants' ability to discriminate native and non-native speech sounds. The chapter is divided into different aspects, namely discrimination on the segmental and suprasegmental levels and neural correlates of speech processing. Additionally, this chapter includes evidence of the effects of experimental procedures in infants' speech sound discrimination. Chapter 3 introduces different theoretical frameworks of infants' phonological development. This chapter discusses the impact of theoretical models in explaining how infants develop native phonological categories, how theoretical models can explain the discrimination performance regarding non-native sounds at a later stage of phonological development, and how the models can explain the emergence and formation of asymmetric processing. Chapter 4 provides an overview of the detailed aims and research questions and concludes by describing the main results of this dissertation. In Study 1 (Chapter 5), I first investigated infants' and adults' abilities to discriminate between Cantonese lexical tones. For this purpose, we tested German-learning infants at 6, 9, and 18 months of age and German-speaking adults on their ability to discriminate lexical tone contrasts. This study was intended to investigate whether the U-shaped developmental pattern found in Dutch-learning infants (Liu & Kager, 2014) can be replicated using a different tone contrast (Cantonese instead of Mandarin) and another native language (German instead of Dutch). This study additionally contrasted the impact of habituation and familiarization procedures on the infants' ability to discriminate between lexical tones. Study 2 (Chapter 6) complements the comparison of different experimental procedures. Study 2 aimed to determine whether the behavioral findings at 6 and 9 months and in adults are also reflected in the neural correlates. In this study, we used the same Cantonese lexical tone contrast, except this time measured by neural correlates to assess perceptual sensitivity in infants and adults. Furthermore, the neural processing of the lexical tone contrast was compared to a Cantonese vowel contrast. Study 3 (Chapter 7) investigated the question of how the acquisition of the native phonological system affects the development of asymmetric vowel processing. In this study, we contrasted two theoretical models intended to explain asymmetrical vowel processing. Chapter 8 concludes the thesis.

CHAPTER 2

2.1 Empirical evidence of infants' speech perception

Overall, infants' ability to discriminate sounds of the world's languages might be one of the most investigated aspects of infants' language acquisition. Several studies have been conducted to investigate discrimination of consonants (e.g., Eimas et al., 1971; Werker & Tees, 1984), vowels (e.g., Polka & Bohn, 1996, 2011; Polka & Werker, 1994; Tsuji & Cristia, 2014), lexical stress (Höhle et al., 2009; Skoruppa et al., 2009), and more recently, lexical tone contrasts (e.g., Liu & Kager, 2014; Mattock et al., 2008; Mattock & Burnham, 2006; Yeung et al., 2013). In the subsequent sections, I describe the empirical evidence of different patterns of infants' speech development during the first year of life that are apparent in the literature. This description is divided into four parts. The first focuses on infants' language development on the segmental level (Section 2.2). The second focuses on the development on the suprasegmental level, including lexical tone perception (Section 2.3). The third describes evidence of neural correlates of the perceptual sensitivity in the first year of life (Section 2.4). Finally, the fourth part describes the relevance of experimental procedures in assessing perceptual sensitivity in infancy (Section 2.5).

2.2 Infants' phonological development: Segmental level

One of the first patterns that has been discovered is infants' general ability to discriminate native as well as non-native speech sounds from early on. Young infants (below 6 months) discriminate speech sounds independent of their native language and the tested language contrasts (e.g., Eimas et al., 1971; Kuhl et al., 2006; Polka & Werker, 1994; Sundara et al., 2006, 2018; Trehub, 1976; Tsuji & Cristia, 2014; Werker & Tees, 1984). In one of the first studies on infants' speech perception, Eimas et al. (1971) showed that consonants are perceived categorically from early on. They investigated infants' discrimination ability at the age of 1 to 4 months on differences in voice onset time (VOT) with the newly established high-amplitude-sucking paradigm. The infants' ability to perceive differences in sound pairs that crossed the English perceptual boundary was compared to their ability to discriminate between pairs of sounds that fall into the same category. The results indicate that infants only increased their sucking behavior when they heard a new stimulus that crossed the boundary but not for a stimulus within the category even though the acoustic distance between the two conditions was equal. These results suggest that infants have initial sensitivity in perceiving psychophysical differences. Based on these findings, several other studies have been conducted to assess infants’ abilities to discriminate speech sounds across age groups. Follow-up studies indicated that young infants discriminated native sounds as well as non-native ones. Seminal works by Trehub (1976) and Werker and Tees (1984) showed that young infants (age 6-8 months) discriminated non-native contrasts to which adults showed minimal perceptual sensitivity. Furthermore, the decrease in the ability to discriminate non-native speech sounds occurred during the first year of life (Werker & Tees, 1984). By using a conditioned head turn paradigm, Werker and Tees (1984) found that infants’ ability to detect differences between non-native Hindi ([ʈa] – [ta]) or Nthlakampx contrasts ([k’] – [q’] decreased significantly within the first year of life. In contrast, infants learning Hindi or Nthlakampx did not show a decrease in these discrimination abilities.

The decrease in perceptual sensitivity to speech sounds that are not part of infants’ linguistic environment is described in several other studies for a wide range of consonants and vowels: Nthlakampx velar and uvular ejectives and the Hindi dental and retroflex contrast in English-learning infants (ages 6.5 and 8.5 months) (Anderson et al., 2003), the language-specific category boundary of stop consonants in French- and English-learning infants (ages 6–8, 10–12, and 14–20 months) (Burns et al., 2007), the English [r] – [l] contrast in English- and Japanese-learning infants (Kuhl et al., 2006), the continuum from bilabial to dental to retroflex stop consonants in Spanish-learning infants (Peña et al., 2012), the Catalan [e] – [ɛ] vowel contrast in Spanish-learning infants (Bosch & Sebastián-Gallés, 2003), the language-specific discrimination pattern of Swedish and English vowels in Swedish- and English-learning infants (ages 6–8 and 10–12 months) (Kuhl et al., 1992), the German [bu:k] – [by:k] contrast in Japanese-learning infants (ages 4.5 and 10 months) (Mazuka et al., 2014), and the German [dʏt] – [dʊt] and [dut] – [dyt] contrasts in English-learning infants (ages-8 and 10-12 months (Polka & Werker, 1994). In addition, these studies suggest that the decline in infants’ ability to discriminate non-native speech sounds is earlier for vowels than for consonants. For example, Polka and Werker (1994) examined the discrimination ability of infants between 6 to 8 and 10 to 12 months with a conditioned head turn paradigm as well as that of 4- and 6-month-old children in a habituation procedure with listening times. Both methods demonstrated that a decline in perceptual sensitivity occurs from 6 months on. Kuhl et al. (1992) and Bosch and Sebastián-Galles (2003) provide further evidence that the decline and language-specific processing for vowels occurs between 6 and 8 months of age. In the case of consonant processing, previous studies have shown that the decline occurs slightly later compared to vowel discrimination. For example, Werker and Tees (1984) found that English-learning infants discriminated the two non-native Hindi dental and Nthlakampx consonant contrasts between the ages of 6 and 8 months, whereas infants between 10 and 12 months did not show evidence of discrimination (Werker & Tees, 1984).

Additionally, it is not only the decrease of perceptual sensitivity to non-native speech contrasts but also an increase in sensitivity to native speech sounds that has been observed for consonants and vowels in the first year: the English native [r] – [l] contrast (Kuhl et al., 2006), the English native [m] – [n] contrast (Narayan et al., 2010), the English native [d] – [ð] contrast (Polka et al., 2001; Sundara et al., 2006), showing an increased sensitivity to consonants in 10- to 12-month-olds and a meta-analysis showing improved native vowel discrimination from 6 months on (Tsuji & Cristia, 2014).

Nevertheless, other results contradict these findings of increased perceptual sensitivity to native contrasts with diminished discrimination of non-native speech contrasts within the first year of life. Some studies found a reversed pattern that indicated no discrimination abilities in younger (below 7 months of age) infants but increased sensitivity to contrast with increasing age (from 8 months on) for native and non-native contrasts: the native Filipino [na] – [ŋa] contrast (Narayan et al., 2010), native vowel length contrast in Japanese (Mugitani et al., 2009; Sato et al., 2010), native Dutch [i] – [ɪ] perception (Liu & Kager, 2016b), and non-native German vowel discrimination by Japanese-learning infants (Mazuka et al., 2014). The results demonstrated that for some contrasts, infants need more experience to develop full perceptual sensitivity to the specific contrast. This (linguistic) maturation supports not only the development of native sound discrimination ability but possibly also the processing of non-native sounds that have similarities to native sounds (e.g., Mazuka et al., 2014). With gained linguistic experience, infants could therefore use their native phonological knowledge to better discriminate between other (non-native) contrasts.

In addition to the previous findings, a general decline in perceptual sensitivity has not been found for all non-native speech contrasts. There is evidence that infants and adults maintain discrimination abilities for several non-native consonants and vowel contrasts, namely Zulu-clicks in English-learning infants and adults (Best et al., 1988), the English [d] – [ð] contrast in French- and English-learning infants (Polka et al., 2001), the English [d] – [ð] contrast in English- and French-learning children (Sundara et al., 2006), and the German [dut] – [dyt] and English [dɛt] – [dæt] contrasts in English- and German-learning infants and adults (Polka & Bohn, 1996). In these cases, maintained perceptual sensitivity could be the result of acoustically high salient contrasts, possible assimilation effects of non-native speech sounds to native phonological categories, or the interaction between salience and similarity to native categories.

A few studies also provide evidence of a U-shaped developmental pattern by demonstrating discrimination abilities in younger (below 6 months) and older age groups (above 10 months) but not in intermediate age groups (6-9 months) (e.g., Best & Faber, 2000; Bosch & Sebastián-Gallés, 2003; de Klerk et al., 2019). These three studies investigated infants’ vowel discrimination abilities at three different ages within language development. All three studies found that infants failed to perceive a difference between two non-native (vocalic) contrasts at 6 to 8 months (Best & Faber, 2000) and 8 months (Bosch & Sebastián-Gallés, 2003; de Klerk et al., 2019). Nevertheless, infants showed perceptual sensitivity at younger (3 to 5 months, Best & Faber, 2000; 4 months, Bosch & Sebastián-Gallés; 6 months, de Klerk et al., 2019) and older ages (10 to 12 months, Best & Faber, 2000; 12 months, Bosch & Sebastián-Gallés; 10 months, de Klerk et al., 2019). The relatively rare findings of the U-shaped pattern can be attributed to the fact that only a few studies have investigated infants’ and toddlers’ speech perception abilities beyond the initial decline of perceptual sensitivities.

Another pattern that has been observed in adults’ speech perception and infants’ language acquisition is asymmetrical perception. In experiments testing infants’ perception, better discrimination of a speech contrast X-Y was found when X was selected as the background stimulus and a response to the change to Y was measured than vice versa. This asymmetrical perception has been observed for consonants (/ɔmpa-ɔnta/ (e.g., Tsuji et al., 2015) and vowels (e.g., Polka & Bohn, 2011 for a review on a variety of vowel contrasts). Results of previous studies indicate that asymmetrical perception emerges from either language-universal (Polka & Bohn, 2003, 2011; Tsuji et al., 2015) or language-specific mechanisms (e.g., Kuhl, 1991; Kuhl & Iverson, 1995).

2.3 Infants' phonological development: Suprasegmental level and lexical tones

As described in the preceding paragraphs (Section 2.2), much of the research on infants' discrimination abilities has focused on their perception of consonants and vowels. However, to fully acquire language, suprasegmental information, like stress and intonation patterns, is also highly relevant. Previous studies have shown that infants develop a language-specific listening preference for native stress patterns between 4 and 9 months (Friederici et al., 2007; Höhle et al., 2009; Skoruppa et al., 2009). By comparing two groups of infants who are learning different stress systems (French with fixed stress and Spanish with contrastive stress patterns), Skoruppa et al. (2009) have shown that French-learning infants had difficulties in encoding alternating stress patterns. In contrast, Spanish-learning infants discriminated between different stress patterns without any problem. However, French infants are not per se insensitive to suprasegmental features on the acoustic level. In a follow-up experiment, Skoruppa et al. (2009) demonstrated that French-learning infants discriminated between different stress patterns when they were presented with single pseudo-words instead of multiple ones. The authors suggest that the simpler version allowed infants to rely on acoustic rather than language-specific encoding strategies.

The perception of prosodic information to discriminate between different intonation patterns has so far only been investigated in a handful of studies (e.g., Butler et al., 2016; Frota et al., 2014; Soderstrom et al., 2011). In the past, it has been shown that infants at 5 to 6 months as well as 8 to 9 months can discriminate statements and Yes-No questions in European Portuguese (Frota et al., 2014). In contrast, an initial perceptual ability to discriminate between prosodic broad and contrastive focus does not seem to be present. Butler et al. (2016) have shown that infants learning European Portuguese can only discriminate broad and contrastive focus at 11 to 13 months but not at 6 to 7 months. This finding demonstrates that the heterogeneous results obtained in the studies on segmental discrimination abilities can also be found to some extent for perceptual sensitivities at the suprasegmental level.

Apart from suprasegmental features, 60 to 70% of languages also have lexical tones (Yip, 2002). Lexical tones are defined by the property that a change in pitch results in different meanings on the lexical level. Recently, studies have investigated lexical tone perception in infants learning a tone and a non-tone language. Infants learning a tone language encounter the reality that both segmental and pitch changes lead to different word meanings. In contrast, for infants learning a non-tone language, variations in pitch are not functional on a lexical rather than, for example, pragmatic level. One exception is contrastive lexical stress in some non-tone languages, where pitch differences are also used at the word level. However, in those cases, pitch is only one factor that characterizes contrastive lexical stress; it occurs in combination with duration and/or intensity cues (Cutler, 2005).

The first studies of lexical tone perception have shown a decrease in discrimination during the first year of life in infants learning a non-tone language (Mattock et al., 2008; Mattock & Burnham, 2006; Yeung et al., 2013). However, later studies also showed a more differentiated picture of perceptual abilities. These studies investigated infants learning different prosodic systems (e.g., Dutch, English, French) and the perception of various non-native tone languages (e.g., Mandarin, Thai, Cantonese) by using a wide range of different experimental procedures (e.g., habituation, familiarization). When summarizing previous findings on infants' ability to discriminate between non-native lexical tones during the first year of life, four different developmental patterns emerged:

(1) Decrease in perception abilities from 4 to 9 months of age (Mattock & Burnham, 2006; Mattock et al., 2008; Yeung et al., 2013)

(2) Increase of discrimination performance from 4 to 9 months age (e.g., Chen et al., 2017; Chen & Kager, 2016; Singh et al., 2018)

(3) No change of perceptual sensitivity from 4 to 12 months age (e.g., Ramachers et al., 2018; Shi & Gao et al., 2017)

(4) U-shaped developmental pattern (Liu & Kager, 2014)

The diverging findings might be related to the fact that even infants learning a non-tone language have exposure to suprasegmental information on the intonation level. In this sense, infants might not be fully naïve to pitch differences on the word level.

2.4 Infants' phonological development: Neural correlates

In addition to behavioral studies, researchers in the field of language acquisition have investigated the neural correlates of speech processing. These neural studies complement the behavioral results and investigate whether infants' brains respond differently to different speech sounds to the same extent that listening time measurements indicate. The mismatch negativity (MMN) component, which can also be a positive mismatch response (P-MMR) in infants, has been widely used to investigate neural responses to speech and non-speech material. This component is pre-attentive; it is independent of the participant's attention. Therefore, this component is most suitable when investigating the speech perception abilities of special populations like infants, children, and individuals with special health conditions (see Duncan et al., 2009). An MMN or P-MMR is elicited when the brain detects a difference between a frequent standard and one or more rare deviants. Behavioral discrimination abilities correlate with the amplitude of the component. The higher the amplitude is, the easier the discrimination of the tested contrast (e.g., Näätänen, 2001). Nevertheless, the MMN and P-MMR can also be elicited in the absence of behavioral discrimination (Rivera-Gaxiola et al., 2005).

One of the first studies using this paradigm with infants investigated the influence of language experience on infants' neural discrimination of native versus non-native vowels (Cheour et al., 1998). Cheour and colleagues (1998) tested Finnish-learning infants at 6 and 12 months and Estonian-learning infants at 12 months on their perception of an Estonian (non-native for Finnish, native for Estonian infants) and Finnish (native for Finnish and for Estonian infants) vowel contrast. In the 6-month-olds, Cheour et al. (1998) observed a P-MMR for both Finnish and Estonian vowels. In contrast, the 12-month-olds showed language-specific responses: Finnish infants demonstrated a diminished MMN to the non-native Estonian vowel, but Estonian infants, whose phonological system contains both vowels, showed equally large P-MMRs. Jansson-Verkasalo et al. (2010) obtained similar results. They tested Finnish-learning infants between 6 and 12 months old on their response to a native (/ø/ vs. /e/)

and a non-native vowel contrast (/ɤ/ vs. /e/]. The results revealed a decreased response between 6 and 12 months for the non-native vowel contrast, whereas the P-MMR increased or was maintained for the native vowel contrast. Nevertheless, it has also been observed that infants showed neural residuals for non-native speech sounds (Rivera-Gaxiola et al., 2005). Rivera-Gaxiola et al. (2005) tested English-learning infants at 7 and 11 months on three consonant contrasts differing in VOT. For the native contrast (English voiceless vs. voiceless aspirated alveolar stop), the authors showed that the ERP response increased from 7 to 11 months, while for the non-native contrast (Spanish voiced vs. voiceless alveolar stop), the ERP response decreased. However, infants at 11 months did not show a homogeneous pattern: some infants showed a P-MMR, whereas the contrast elicited an MMN for some infants, which indicates residual neural evidence of the ability to discriminate non-native speech contrasts.

2.5 Effect of experimental methods in infants' speech sound discrimination

A crucial factor in assessing infants' discrimination abilities is the choice of an appropriate experimental procedure. Developmental scientists can measure infants' discrimination abilities of speech sounds via looking times. Looking times are measured via either head turn or central fixation methods. Both methods mainly use two phases: a pre-exposure phase and a test phase. Two different procedures can be used for the pre-exposure phase: familiarization and habituation procedures. In familiarization experiments, infants are often required to accumulate a certain amount of looking times. The experimenter predetermines the amount of looking times the infant is required to accumulate; therefore, this value is equal for all infants. However, as Kavšek and Bornstein (2010) point out in their review, a null result from the test phase in fixed initial exposures cannot be attributed to whether infants have failed to encode the speech signal from the initial exposure phase or whether they cannot discriminate the contrasts at the test phase. Furthermore, the infant's response to stimuli in the test phase in familiarization procedures can yield different effects: novelty preference or familiarity preference. On the group level, this can produce heterogeneous results (e.g., Hunter & Ames, 1988). In contrast, habituation procedures are expected to generate novelty preferences because infants enter the test phase based on an individually controlled encoding status of the stimulus. In other words, the habituation procedure is terminated when infants meet a predefined criterion through a reduction of the average looking time. The amount of looking time may thus vary between infants depending on how long it takes for the infant to reach this criterion. In contrast to fixed initial exposure phases, habituation might cover the infant's specific need in terms of the exposure time required to encode the speech signal. This may explain why experiments using habituation procedures showed more robust results than experiments using familiarization procedures regardless of whether the paradigm was implemented via head turns or central fixation methods (Cristia et al., 2016). Furthermore, replication and adapting experimental procedures suggest that some null findings are related to the sensitivity of the experimental procedure rather than the infant's actual perceptual sensitivity (e.g., Bijeljac-Babic et al., 2012; Sebastián-Gallés & Bosch, 2009; Sundara et al., 2018).

2.6 Summary of the previous findings on infants' discrimination abilities

In summary, the patterns of non-native speech sound discrimination abilities in infants are heterogeneous. At both the segmental and suprasegmental levels, it has been found that infants can initially discriminate between non-native sounds. However, this ability decreases during the first year of life (e.g., Eimas et al., 1971; Kuhl et al., 2006; Polka & Werker, 1994; Sundara et al. 2006, 2018; Tsuji & Cristia, 2014; Trehub, 1976; Werker & Tees, 1984,). Nevertheless, the opposite pattern has also been found, namely that younger infants cannot discriminate sounds but this ability emerges at an older age (e.g., Mazuka et al., 2014; Mugitani et al., 2009; Narayan et al., 2010; Sato et al., 2010). The third observed pattern is the maintained discrimination of non-native contrasts during the first year of life and beyond (Best et al., 1988; Polka & Bohn, 1996; Polka et al., 2001; Sundara et al., 2006). Notably, the results are especially divergent in the discrimination ability of lexical tones (decline in perceptual sensitivity: Mattock & Burnham, 2006; Mattock et al., 2008; Yeung et al., 2013; increase in perceptual sensitivity: Chen & Kager, 2016; Chen et al., 2017; no change of perceptual sensitivity: Ramachers et al., 2018; Shi et al., 2017; U-shaped developmental pattern: Liu & Kager, 2014). The research is complemented by results from neural discrimination experiments that showed either a decrease in perceptual sensitivity (e.g., Cheour et al., 1998;

Jansson-Verksalo et al., 2010) or residual neural evidence for non-native language contrast (e.g., Rivera-Gaxiola et al., 2005). The next chapter relates the empirical evidence of infants' phonological development to theoretical frameworks of perceptual development. Of particular relevance for describing the models are the questions of how native sound categories are formed, why infants and adults can discriminate between non-native sounds, and how asymmetric processing in the development of perception can be explained.

CHAPTER 3

3.1 Models of perceptual development

Aslin and Pisoni (1980) discuss three different theoretical accounts to cover infants' development of speech perception abilities: the universal, the attunement, and the perceptual learning account (see Figure 1). The models essentially differ in their assumptions about the infant's developmental status at birth. The universal and perceptual learning accounts represent opposing views of the developmental status at birth. The universal account assumes that infants' perceptual sensitivity is fully developed at birth, whereas the perceptual learning account assumes that infants' perceptual sensitivity to speech contrasts is completely undeveloped at birth. In the latter case, discrimination abilities develop with increasing age. Hence, the most relevant factor for developing perceptual sensitivity is experience with stimuli present in the environment. In contrast to this, the universal account claims that the lack of experience, due to non-presence of stimuli in infants' environment, is responsible for a loss of initially present perceptual sensitivities.

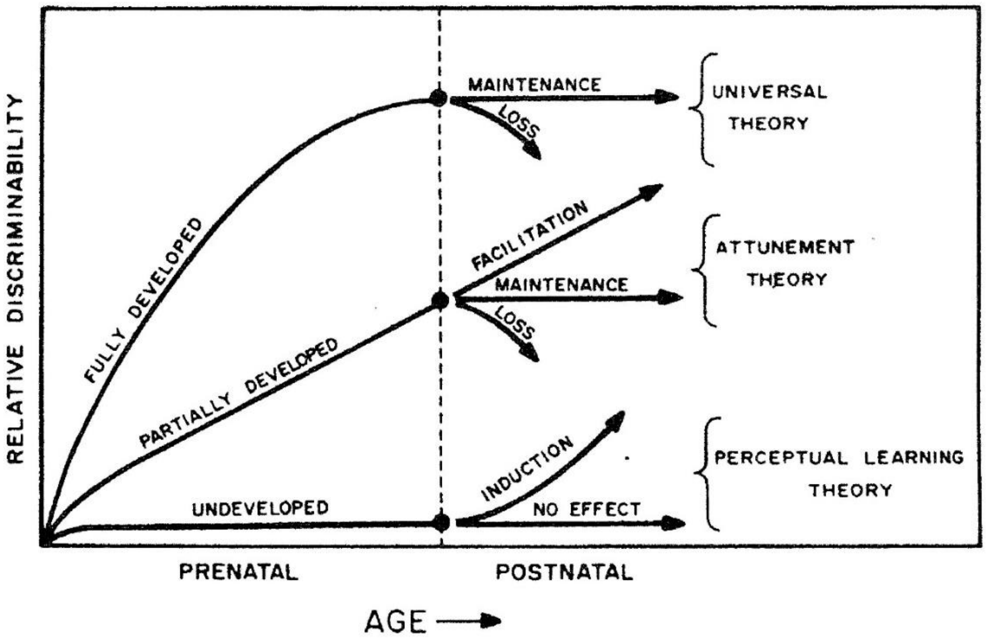

Figure 1. Effects of experience that can influence infants' perception of perceptual sensitivity according to Pisoni and Aslin (1980, p. 77).

The attunement account can be considered an intermediate account between the perceptual learning and universal accounts. The attunement account proposes that infants have already developed some prenatal perceptual sensitivity but that experience with stimuli present in their environment also shapes their perceptual sensitivity. The presence or absence of specific sounds in the infants' experienced input leads to facilitation, maintenance, or loss of discrimination abilities. An alternative term for specifying this process is *perceptual reorganization*. Likewise, the term *perceptual narrowing* is frequently used in the literature. However, *perceptual narrowing* refers only to the aspect of the loss of discrimination abilities at later ages but not to the entire range of developmental paths that have been observed. Since the term *perceptual narrowing* is relatively restricted to certain phenomena in the development of infants' perceptual abilities, we use the terms of *perceptual reorganization* and *perceptual attunement* to refer to the whole process of the development of infants' discrimination abilities.

Aslin and Pisoni (1980) state that a hybrid of all three accounts might provide the best description of infants' capacities to perceive different stimuli: "not only one of these classes of theories will uniquely account for the development of all speech contrasts, rather, but it may also be the case that some hybrid of the theories provides the best description of the development of specific classes of speech-sound discrimination" (Aslin & Pisoni, 1980, p. 80). Evidence for hybrids of these accounts comes from experimental findings. For instance, the first studies (e.g., Eimas et al., 1971) provided evidence for the universal account by demonstrating categorical perception at an early age (in 1- to 4-month-olds). However, later studies challenged the universal account (e.g., Kuhl et al., 2006; Narayan et al., 2010). Enhanced discrimination abilities for native contrasts have been shown for a variety of speech sounds (e.g., Kuhl et al., 2006; Narayan et al., 2010). Together with other results (see Chapter 2), it has been shown that the attunement account provides the best explanations for the diverging developmental patterns. Based on this account, several theoretical frameworks have been developed. These frameworks are intended to explain the influence of the native language system, the acoustic salience of a specific sound contrast, and the amount of language input on infants' speech sound discrimination. In the following section, I discuss central questions about infants' speech perception development and how different theoretical frameworks approach these questions.

3.2 Theoretical frameworks to explain phonological development

Several theoretical frameworks have been developed to explain phonological development. This section discusses the frameworks based on three central questions that are relevant for the empirical findings of the present dissertation:

(1) How are native speech categories formed?
(2) How and why can infants discriminate non-native speech sounds?
(3) How can asymmetrical speech perception in infancy be explained?

The models to be compared are the Processing Rich Information from Multidimensional Interactive Representations framework (PRIMIR; Curtin & Werker, 2018; Werker & Curtin, 2005), the Perceptual Assimilation Model (PAM; Best, 1994, 1995), the Native Language Magnet Model (NLM, or NLM-extended; Kuhl, 1991; Kuhl et al., 2006; Kuhl et al., 2008; Kuhl & Iverson, 1995), and the Natural Referent Vowel framework (NRV, Polka & Bohn, 2011). However, not all these theoretical frameworks can explain all three questions in the same way, so only those models that make corresponding assumptions about the specific questions are compared in the respective section.

3.2.1 Formation of native sound categories

As described in Chapter 2, young infants are able to discriminate between sounds independent of whether those sounds are part of their native language or not. The initial sensitivity to discriminate speech sounds changes with the experience infants have with their native language, which leads to the development of infants' phonological system. In this section, I discuss theoretical frameworks that emphasize partly overlapping and partly differing aspects of phonological development. The Native Language Magnet Model (NLM; Kuhl, 1991; Kuhl et al., 2006; Kuhl & Iverson, 1995), the Native Language Magnet Model extended (NLM-e; Kuhl et al., 2008), and the Processing Rich Information from Multidimensional Interactive Representations framework (PRIMIR; Curtin & Werker, 2018; Werker & Curtin, 2005) explicitly describe how infants form phonological categories in their native language. In the following, I first describe the developmental process of category formation described by the NLM and afterward by PRIMIR.

The NLM was developed with an initial focus on vowel perception. It is in line with the attunement account, and it assumes an enhancement of perceptual sensitivity for native speech sounds with increasing age (Kuhl et al., 2006; Tsao et al., 2006). A central component of this model is the assumption that speech perception is affected by the prototypicality of speech sounds. The most frequently activated exemplars of speech sounds are called *prototypes*. As a result, prototypes are good exemplars of phonetic categories, and those members are easier to remember. Based on cross-linguistic (Kuhl et al., 1992) and cross-species studies (Kuhl, 1991), the authors argued that these prototypes develop from accumulated experience with the native language, which then results in the development of phonological categories. These phonological categories then lead to warping of the perceptual space around the prototypes. Phonetic prototypes act as attractors and pull members of the same phonetic category and therefore diminish the perceived acoustic distances between the prototype and other exemplars of the category. In contrast, non-prototypes do not attract other members of the category. The underlying learning mechanism for the emergence of prototypes is statistical learning guided by the frequency of occurrence of specific sounds in the speech input to the child. The NLM-e distinguishes between four different phases through which children go in their development (Kuhl et al., 2008). In Phase 1, the infant has universal perceptual sensitivities, and the perceptual sensitivity toward speech contrasts is determined only by general auditory processing mechanisms. Therefore, the more acoustically salient the contrast is, the better infants’ ability to discriminate the contrast is. The distinction between native and non-native speech categories emerges in Phase 2. Infants’ speech perception is sharpened during the second phase and begins to be influenced by the linguistic environment. When Phase 3 begins, infants have already improved perceptual sensitivity toward native speech categories and diminished perceptual sensitivity to non-native speech contrasts. Furthermore, in Phase 3, infants begin to develop more language-specific processing and learning mechanisms (e.g., phonotactics, recognition of segments and syllables based on transitional probabilities and object-sound relations). Finally, Phase 4 describes the process in which neural networks are already developed and cannot be easily altered by new incoming speech signals.

In contrast, PRIMIR explains the formation of native speech categories as an integration process with continuous interaction of three different planes (the General

Perceptual, Word Form, and Phoneme planes), which are influenced by dynamic filters (Curtin & Werker, 2007; Werker & Curtin, 2005). In general, all the information about speakers, acquired phonetic categories, words, et cetera is organized in clusters in the respective planes. Phonetic properties and indexical information are stored in the General Perceptual plane. Additionally, all other information that is available in the speech signal is stored in this plane, including information about the identity, age, and gender of the different speakers. At later stages, phonetic categories are also stored in this plane. Clusters of first word-object relations are formed by learning and using the variability in phonetic and indexical information. Thus, the first word-object relations are based on phonetic information and only at later stages are based on phonemic information. The word-object relations are stored in the Word Form plane. Initially, word forms are stored separately from their referents. As soon as the infant has learned his or her first words, the process of identifying commonalities between different words begins. Phonemes emerge when infants generalize the common features between words. These phonemes are stored in the third plane (the Phoneme plane). Once the first phonemes are formed, they can anchor infants' attention. With the help of these newly established phonemes, further word-object relations can be formed, and the infant can focus on word learning. The increasing lexicon sharpens phonological categories until they reach a constant status. Consequently, in infants, the first phonemes and phonological categories do not necessarily represent their final status. Nevertheless, the entire process is also dependent on the acoustic salience of speech contrasts. Accordingly, it can be more difficult for infants to distinguish critical words when two sounds are close to each other in acoustic distance (e.g., Stager & Werker, 1997). In contrast to the General Perceptual plane, the Word Form and the Phoneme planes are language-specific. As previously mentioned, the different planes do not act independently of each other but rather as a network with a continuous exchange between the planes. Furthermore, dynamic filters sharpen the three different planes. For example, the universal, initial preference for infant directed speech (IDS) leads to the sharpening of language-specific units during the developmental process of the perceptual reorganization by guiding infants' attention directly to certain linguistic features. This guiding of attention by the dynamic filters then supports the infants' segmentation abilities of words from the speech stream. As soon as infants are able to extract words from the speech stream, the Word Form plane appears. After the first words are formed in the Word Form plane, native phonological categories begin to appear in the Phoneme plane. The underlying mechanism for extracting these native categories is statistical learning, such as in forms of frequencies of exposures via different distributions or transitional probabilities.

As already discussed in Chapter 2, previous research reported evidence of maintained or enhanced discrimination abilities for specific non-native phonological categories. In the next section, I discuss how theoretical frameworks are able to explain discrimination of non-native phonological categories.

3.2.2 Discrimination of non-native speech sounds

The attunement account predicts enhanced perceptual sensitivity to linguistic contrasts that are present in the infants' environment and decreased discrimination abilities for non-native speech sounds (Pisoni & Aslin, 1980). However, this decline has not been found for all non-native speech contrasts (e.g., Best et al., 1988). The Perceptual Assimilation Model (PAM; Best, 1994, 1995) and PRIMIR frameworks make different assumptions about infants' ability or non-ability to discriminate non-native speech sounds. PAM provides the most detailed predictions about if and when listeners can discriminate non-native speech sounds. The basic assumption of the model is that non-native speech sounds are assimilated to different patterns. This model considers not only differences but also similarities between native and non-native categories, which influence the ability to discriminate between those sounds. The similarity is defined based on articulatory phonology and the ecological approach to speech perception. The model defines six different assimilation patterns (see Best, 1995). The first one predicts good discrimination of two non-native speech sounds if the sounds are mapped onto two different phonological categories of the native language, the Two-Category type (TC). For this type, discrimination performance is near ceiling. In contrast, for the Single Category type (SC), if two non-native speech contrasts are assimilated to a single native speech category discrimination ability is low. The third pattern depends on the category goodness fit to the native speech category during the assimilation process (Category Goodness [CG] type). In this case, the discrimination ability differs according to how well the two sounds are assimilated to a native category. If both sounds correspond equally well to the same native category, the discrimination is poorer. However, if two sounds differ in how well they fit into the native phonological category, the discrimination ability is better. Therefore, the perception of CG contrasts is more scaled and less absolute compared to the other assimilation patterns. In the situation that both sounds cannot be mapped onto a native category (Uncategorized-Uncategorized [UU] type), discrimination varies depending on how close the non-native speech sounds are in acoustic space to native categories. Unlike the CG type, the two non-native speech sounds do not necessarily match native categories exactly but rather fall into the same phonetic space of the native categories. In the case where one speech sound is perceived as not belonging to the native phonological categories and the other sound is assigned to a native phonological category, this results in the Uncategorized-Categorized (UC) type. The model predicts good discrimination performance for the UC type. The last category (Non-Assimilable type [NA]) refers to non-native speech sounds that are not assimilated to native speech categories because their properties are too distinct from the native speech, leading to their perception as non-speech signals. This perception of sounds as non-speech can lead to a high degree of perceptual sensitivity to the contrast.

Regarding infants' perceptual ability, PAM predicts that young infants discriminate native and non-native speech sounds by detecting universal articulatory gestures (Best & McRoberts, 2003). Before completing the perceptual attunement process, infants do not yet have the phonological categories of their native language; thus, the assimilation processes do not affect perception, as described in PAM. As soon as infants develop native phonological categories the assimilation pattern, as previously described, can be applied.

PRIMIR allows for more flexibility regarding perceptual sensitivities to non-native speech sounds. It is of considerable relevance that the planes in the framework are organized non-hierarchically. The networking mechanisms lead the acoustic information in the General Perceptual plane still affecting speech perception at later stages of development – for example, after the completion of the perceptual reorganization process. Hence, discrimination depends on the acoustic salience of the contrasts. According to PRIMIR, the dynamic filters also have an impact on perceptual abilities. Specifically, using different experimental procedures can alter the detection of perceptual sensitivity. The more adapted the task is to specific needs, the more likely infants (and even adults) are to discriminate speech contrasts.

3.2.3 Asymmetrical vowel perception

A common phenomenon in speech processing is asymmetrical vowel perception in infants and adults (e.g., Kuhl & Iverson, 1995; Masapollo et al., 2017; Polka & Bohn, 2011). Asymmetrical perception can occur when a speech contrast X-Y is better distinguishable in one direction (e.g., when Y is presented in a stream of X) than in the other direction (e.g., X in a stream of Y). The asymmetries are observable in experimental procedures, such as when infants are habituated with one speech sound and tested with the other speech sound, resulting in better discrimination than vice versa. With respect to asymmetrical speech perception, the Native Language Magnet Model (NLM; Kuhl, 1991; Kuhl et al., 2006; Kuhl & Iverson, 1995), NLM-e (Kuhl et al., 2008), and the Natural Referent Vowel framework (NRV; Polka & Bohn, 2011) make explicit predictions. Both models' theoretical foundations are based on the evidence of asymmetrical vowel perception.

The NLM assumes that asymmetries in vowel perception emerge due to the language-specific internal structure of a vowel category. According to this approach, native language sound categories have prototypical exemplars that act as perceptual magnets. These magnets attract other members of the category. Based on this attraction, better performance is predicted in detecting the change from a non-prototypical to the prototypical exemplar of a vowel category compared to the change in the opposite direction. The asymmetries in vowel perception are related to the internal organization of the native language vowel categories. Hence, the asymmetries only emerge when these vowel categories are established as an effect of language acquisition. In contrast, the NRV proposes that asymmetrical vowel perception is based on specific acoustic and phonetic factors (Polka & Bohn, 2011). According to this framework, perception in the direction of a less focal to a more focal vowel is easier compared to from a more focal to a less focal vowel. Vowel focalization refers to the formant convergence of two adjacent formants (for example F1-F2, F2-F3)[1]. Thereby, the spectral energy between the two formants is reinforced and the acoustic energy is concentrated on a narrower field. The focal vowel acts as an anchor point, and the more focal vowel is easier to detect when tested against a less focal vowel (Polka & Bohn, 2011; Schwartz et al., 2005). Irrespective of the specific acoustic basis for the asymmetrical discrimination, the asymmetries in vowel perception predicted by the NRV are considered universal and thereby initially language independent since they are grounded in general acoustic

---

[1] The equation to calculate formant convergence is the following:

$$E_F = \alpha \sum_{i=1}^{N} \left(\frac{-1}{(\mathrm{F1}_i - \mathrm{F2}_i)^2} + \frac{-1}{(\mathrm{F2}_i - \mathrm{F3}_i)^2} + \frac{-1}{(\mathrm{F3}_i - \mathrm{F4}_i)^2} \right)$$

The lower E_F reflects the more focal vowel (Sanders & Padgett, 2008).

and phonetic properties of the vowels. However, these perceptual asymmetries may be modulated by native language experience (Masapollo et al., 2017; Polka & Bohn, 2011). Thus, Polka and Bohn (2011) argue that the influences of the native language phonological system may override the perceptual asymmetry since asymmetries were only evident for non-native but not for native vowel contrasts in adults and 12-month-old infants (Polka & Bohn, 2011). The NRV was adjusted to these findings by proposing that asymmetrical perception can disappear for native speech sounds but be maintained for non-native speech sounds.

3.2.4 Summary of the theoretical frameworks

All models have in common that they assume that the underlying process in infants' speech development is the attunement account, as suggested by Aslin and Pisoni (1980). Table 1 provides an overview of the different frameworks. Overall, PRIMIR, PAM, and the NLM predict enhanced perceptual sensitivity to linguistic contrasts that are present in infants' environments and decreased discrimination abilities for non-native speech sounds. However, there is also evidence of perceptual sensitivity to specific non-native speech contrasts after the first year of life (e.g., Liu & Kager, 2014). Not all the presented frameworks can explain this late discrimination to the same extent. PAM and PRIMIR provide the most detailed predictions and explanations. In contrast, the strength of the NRV lies in the description of asymmetric vowel processing and the framework can describe the asymmetries very well in terms of the acoustic properties of the vowels. The predictions of asymmetrical perception apply to both native and non-native speech sounds. According to the NLM model, the most activated exemplars of a speech category, which correspond to the prototypes of the phonological category, are developed based on accumulated experience to the native language. Prototypes act as magnets and warp the perceptual space by pulling other members of the category, which results in decreased perceptual sensitivity of speech contrasts close to the prototype. In contrast, no magnet effect can occur for non-native contrast since the formation of prototypes is dependent on experience.

Furthermore, the frameworks differ with respect to how native categories are formed during the first year of life. PAM is primarily intended to explain the change of discrimination performance during infancy and does not consider how phonological categories are built. In contrast, PRIMIR and the NLM include the process of native language phonetic category formation. Since the NRV describes rather general perceptual phenomena, it makes no assumptions about how phonetic categories are formed.

To explain asymmetrical (vowel) perception, only the NRV and NLM predict a concrete pattern. These two frameworks differ especially concerning language specificity. Whereas the NRV anticipates universal and therefore language-independent asymmetrical perception, the direction of perception in the NLM is language specific.

	PRIMIR	PAM	NLM	NRV
Underlying account	Attunement account	Attunement account	Attunement account	Framework is not specialized for explanation of developmental pattern
Explanation of category formation in infancy	Yes	No	Yes	No information
Explanation of non-native speech sound discrimination	Discrimination is task dependent and accessible over the General Perceptual plane	Yes, via different assimilation pattern	No, formation of prototype is dependent on language experience	Asymmetrical discrimination for non-native vowel sounds
Explanation of asymmetrical (vowel) perception	No information	No information	Yes, divided between universal (early stage) and language-specific direction perception (later stage)	Yes, asymmetries are language independent and based on general perceptual abilities.

Table 1. Overview and comparison of the different theoretical frameworks regarding infants’ phonological development.

CHAPTER 4

Aims, research questions, and main results

In this dissertation, I deepen the insights into the developmental trajectory of infants' phonological acquisition by investigating changes in speech perception across age groups. In three studies, we tested infants' discrimination abilities regarding lexical tones and vowels by using multiple experimental procedures (habituation, familiarization, and neural correlates). Studies 1 (Chapter 5) and 2 (Chapter 6) complement the previous results and broaden the perspectives of infants' speech discrimination abilities of lexical tones by combining behavioral and neural experimental procedures. Study 2 additionally examined and compared the neural correlates of non-native lexical tone and vowel processing, for which the German phonological system has a comparable vowel category. Study 3 (Chapter 7) investigated the effect of language specificity on the emergence of asymmetrical vowel perception during the first year of life and compared different theoretical approaches.

Previous lexical tone perception studies with infants used tone contrasts from many tone languages and tested infants learning tone or non-tone. The diverging findings do not yet allow the disentanglement of the acoustic perceptibility of the tested tone contrast from the influence of the native language. One aim of **Study 1** (Chapter 5) was to further investigate the U-shaped developmental pattern as found for Dutch infants (Liu and Kager, 2014) using a Cantonese tone contrast to test German-learning infants (6-, 9-, and 18-month-olds) and German- and Cantonese-speaking adults. As described in the previous chapters, studies on lexical tone perception in infants suggest that one source of the diverse findings on perceptual sensitivity of lexical tone contrasts may be the different experimental procedures used across studies. Hence, the second goal of Study 1 was to investigate how different experimental procedures (i.e., habituation and familiarization procedures) may affect infants' performance in tasks that test sound discrimination. The results of Study 1 support the U-shaped developmental pattern previously found by Liu and Kager (2014). German-learning infants discriminated between the Cantonese high-rising and mid-level contrast at 6 and 18 months but not at 9 months. Additionally, German-speaking adults discriminated between these Cantonese tones. However, native Cantonese listeners outperformed

German listeners in their discrimination ability. The two different experimental procedures for assessing infants’ and toddlers’ perceptual sensitivity revealed a decline in perceptual sensitivity from 6 to 9 months when tested with the habituation-dishabituation procedure as well as a sensitivity to the contrast among 18-month-olds when tested with the familiarization procedure. The habituation procedure led to larger effect sizes on the group level compared to the familiarization procedure.

Study 2 (Chapter 6) further investigated the developmental trajectory of infants’ perceptual abilities during the first year of life and focused on two objectives. The first objective was to investigate the perceptual reorganization process of lexical tone processing by augmenting it with the neural perspective. Given the diverging results from behavioral studies, we aimed to broaden the perspective on infants’ perceptual abilities by adding the neural component. The second objective was to compare non-native lexical tone perception to a segmental (native-like vocalic) speech contrast. The investigation of the simultaneous processing of the native-like and non-native sound categories enabled us to further explore the influence of language-specific processing in the perceptual reorganization process. In addition to infants, we systematically tested German-speaking adults on their perception of the same contrasts. The verification of adults’ neural responses allowed us to further assess whether any potential differences in infants’ neural responses to the speech contrasts result from the perceptual reorganization process or whether these differences can be attributed to the inherent properties of the stimuli themselves. The results of Study 2 showed that at 6 months, infants have a positive mismatch response (P-MMR) to both the tone and vowel contrasts. However, at 9 months, we observed an MMN for the vowel contrast but not for the tone, which still elicited a P-MMR. In adults, both contrasts elicited an MMN, with greater MMN amplitude for the vowel compared to the tone contrast. Taken together with the behavioral results from Study 1, the infant brain processes the lexical tone contrast at 9 months, whereas infants may not discriminate the tone contrast on the behavioral level at 9 months. The difference between the processing of lexical tones and that of vowels can also partly be found in adults. In adults, both contrasts elicited an MMN. However, the MMN differed in the magnitude of the amplitude between lexical tones and vowels. Vowels elicited a greater MMN compared to lexical tones and vowels were processed asymmetrically by the adult’s brain.

Study 3 (Chapter 7) investigated the question of how language-specific processing of speech categories interacts with the perceptual reorganization. This study further investigated the shift from universal to language-specific vocalic processing and contributes to the discussion of asymmetrical vowel perception during the first year of life. A special focus was placed on contrasting the two theoretical frameworks of asymmetrical vowel perception and relating them to the emergence of asymmetries in the perceptual reorganization process. The two central models for explaining the emergence of asymmetrical vowel perception are the NLM and NRV. While the NLM makes language-specific predictions, the NRV assumes more general language universals about the emergence of asymmetrical vowel perception. For experimental support of the comparison of the two theoretical frameworks, Study 3 compared native to non-native vowel perception in German-learning infants at 6 and 9 months. The results showed that asymmetrical vowel perception emerged with increasing contact with the native language. No asymmetries were observed in German-learning infants at 6 months, but they were at 9 months.

CHAPTER 5

Perceptual reorganization of lexical tones: Effects of age and experimental procedure[2]

Abstract

Findings on the perceptual reorganization of lexical tones are mixed. Some studies report good tone discrimination abilities for all tested age groups, others report decreased or enhanced discrimination with increasing age, and still others report U-shaped developmental curves. Since prior studies have used a wide range of contrasts and experimental procedures, it is unclear how specific task requirements interact with discrimination abilities at different ages. In the present work, we tested German and Cantonese adults on their discrimination of Cantonese lexical tones, as well as German-learning infants between 6 and 18 months of age on their discrimination of two specific Cantonese tones using two different types of experimental procedures. The adult experiment showed that German native speakers can discriminate between lexical tones, but native Cantonese speakers show significantly better performance. The results from German-learning infants suggest that 6- and 18-month-olds discriminate tones, while 9-month-olds do not, supporting a U-shaped developmental curve. Furthermore, our results revealed an effect of methodology, with good discrimination performance at 6 months after habituation but not after familiarization. These results support three main conclusions. First, habituation can be a more sensitive procedure for measuring infants' discrimination than familiarization. Second, the previous finding of a U-shaped curve in the discrimination of lexical tones is further supported. Third, discrimination abilities at 18 months appear to reflect mature perceptual sensitivity to lexical tones, since German adults also discriminated the lexical tones with high accuracy.

---

[2] This chapter has been published as: Götz, A., Yeung, H. H., Krasotkina, A., Schwarzer, G., & Höhle, B. (2018). Perceptual reorganization of lexical tones: effects of age and experimental procedure. Frontiers in psychology, 9, Article 477. https://doi.org/10.3389/fpsyg.2018.00477

5.1 Introduction

During the first year of life, infants' perception abilities may change for stimuli that are not present or not relevant in their environment. For example, in the linguistic domain, perceptual changes have been detected in infants' sensitivity to native and non-native speech sounds. With increased experience with their native language, infants show an enhanced ability to distinguish between native speech sounds, whereas the initial sensitivity to non-native speech sounds decreases. This pattern of perceptual reorganization has been shown for consonants (Rivera-Gaxiola et al., 2005; Werker & Tees, 1984), vowels (Polka & Bohn, 1996, 2011; Tsuji & Cristia, 2014), lexical tones (Liu & Kager, 2014; Mattock et al., 2008; Mattock & Burnham, 2006; Singh & Fu, 2016; Yeung et al., 2013), and word stress (Bijeljac-Babic et al., 2012; Höhle et al., 2009; Skoruppa et al., 2009).

However, research in recent years has converged on the idea that this picture is too simplistic. On the one hand, not all linguistically relevant sound contrasts are easily discriminable by young infants (Narayan et al., 2010); for a review, see (Maurer & Werker, 2014). On the other hand, there are non-native sound contrasts that are discriminable by children beyond the typical ages of perceptual reorganization, and even by adults (for consonantal contrasts, see Best et al., 2001); for vocalic contrasts, see (Mazuka et al., 2014). The present paper investigates the potential perceptual reorganization of lexical tones by infants learning non-tone languages. Previous research on lexical tone discrimination in infants is characterized by a rather complex pattern of findings: prior studies have found evidence for an increase, a decrease, and no-change in infants' and toddlers' ability to discriminate non-native tone contrasts across ages (for an overview, see Table 2). These divergent findings may be related to a number of dimensions on which these studies varied, including the tone contrasts used, the native language of the participants, and the experimental procedures. Our study focuses on the latter factor and compares the effects of familiarization vs. habituation in the initial exposure phase on German-learning infants' discrimination of a Cantonese tone contrast. In familiarization experiments infants are exposed to certain stimuli for a fixed time, thus the exposure is experimenter-controlled. In contrast, exposure in habituation is infant-controlled as the infant needs to reach a specific criterion (decrease in looking time) to proceed to the test phase. Thus, the latter type of pre-exposure may be more sensitive to the performance of individual infants.

We will first review prior studies on infants' and adults' perception of lexical tones and then present three experimental studies. In the first study, Cantonese tone discrimination in adult native speakers of Cantonese was compared to that in adult native speakers of German. In the second study, the discrimination of the high-rising and the mid-level Cantonese tones was tested in German-learning infants between 6 and 18 months of age using a familiarization procedure. The third experiment investigated discrimination of the same tone contrast in 6- and 9-month-old German infants using a habituation procedure.

Previous studies on infants' non-native lexical tone perception

A detailed review of infant tone perception can be found elsewhere (Singh & Fu, 2016). Here, we focus on studies that have investigated how infants learning a non-tonal language as their native language perceive different tones from various tone systems and we incorporate some more recent studies on infant tone perception. Furthermore, our review will also highlight the details of prior experimental methods.

The first studies that tested perceptual reorganization of lexical tones provided evidence for a decline in tone discrimination by infants learning a non-tone language. Mattock and Burnham (2006) compared English and Chinese (Mandarin- or Cantonese-learning) infants at 6 and 9 months on their discrimination of Thai rising versus falling as well as rising versus low tones using the Conditioned Head-Turn (CHT) paradigm. Infants were first trained to perform a head-turn whenever an auditory background stimulus (a syllable carrying one tone) was replaced by the target stimulus (the segmentally the same syllable with another tone). In the test phase—which was started after three consecutively correct head-turns in the training—the number of correct head-turns to a stimulus change was the dependent variable. Both 6- and 9-month-old Chinese-learning infants discriminated both tone contrasts, but English-learning infants showed a decrease in their discrimination from 6 to 9 months of age, with overall higher performance for the rising-falling than for the rising-low contrast.

No.	Authors	Year	Age (months)	Native Language	Contrast	Exposure phase	Results for non-tone group
1	Chen & Kager	2016	4, 6, & 12	Dutch	Mandarin rising – low-dipping	Habituation	Perceptual Enhancement
2	Chen et al.	2017	4 & 12	Dutch	Mandarin rising – low-dipping	Habituation	Perceptual Enhancement
3	Liu & Kager	2014	5–6, 8–9, 11–12, 14–15, & 17–18	Dutch	Mandarin high-level – high-falling	Habituation	Discrimination across all ages; U-shaped curve (Discrimination 5–6 and 17–18 months)
4	Liu & Kager	2017	5–6, 8–9, 11–12, 14–15, & 17–18	Dutch bilinguals	Mandarin high-level – high-falling	Habituation	Discrimination across all ages; U-shaped curve (Discrimination 5–6 and > 11 months)
5	Mattock & Burnham	2006	6 & 9	English & Chinese	Thai rising – falling & rising – low	Conditioning	Perceptual Decline
6	Mattock et al.	2008	4, 6, & 9	English & French	Thai rising – low	Familiarization	Perceptual Decline
7	Ramachers et al.	2018	6, 9, & 12	Dutch & Limburgian	Limburgian falling – falling-rising	Habituation	Discrimination across all ages
8	Shi et al.	2017	4, 8, &11	French	Mandarin rising – low-dipping; high-level – falling	Habituation	Discrimination across all ages
9	Tsao	2017	6–8 & 10–12	English & Mandarin	Mandarin high-level – low-dipping	Conditioning	Perceptual Enhancement, but discrimination at both ages
10	Yeung et al.	2013	4 & 9	English, Cantonese, Mandarin	Cantonese high-rising – mid-level	Familiarization	Perceptual Decline

Table 2. Summary of the previous results on infant lexical tone perception.

Mattock et al. (2008) extended this study to 4-month-old infants learning English or French while continuing to test 6- and 9-month-olds acquiring these languages. They used a visual fixation paradigm (i.e., they measured infants’ looking time at a central visual display during auditory stimulus presentation), where infants were initially exposed to a syllable representing either a low or a rising Thai tone for 30 s in a familiarization phase. In the test phase, two trial types were presented: four alternating trials that contained both the familiarized and the non-familiarized tone, and four non-alternating trials that only contained tokens of the familiarized tone. In this Stimulus Alternation Preference Procedure (SAPP), the 4- and 6-month-olds but not the 9-month-olds showed significantly longer looking times for the alternating trials compared to the non-alternating trials with no difference across the language groups.

Yeung et al. (2013) tested 4- and 9-month-olds learning Cantonese, Mandarin, and English on Cantonese tones that were similar to the Thai contrast (high-rising vs. mid-level tones) investigated by Mattock and colleagues. Using a modification of the SAPP, infants heard three trial types in the test phase: four alternating trials (familiarized and non-familiarized tone intermixed), two non-alternating trials only containing the familiarized tone, and two non-alternating trials only containing the non-familiarized tone. With this modification, discrimination and preference could be measured in the looking times obtained within the same experiment: that is, differences between the alternating and non-alternating trials would indicate discrimination while the direction of differences between the non-alternating trials would indicate a preference. The English-learning infants showed a decline in the ability to discriminate these contrasts while this was not the case for the Mandarin or Cantonese infants. Moreover, infants learning one of the tonal languages showed an asymmetrical performance pattern with better discrimination when they were familiarized with the high-rising tone than with the mid-level tone. While these studies showed a decline in discrimination ability for non-tone language learners, others have found enhanced perceptual abilities with increasing age (Chen et al., 2017; Chen & Kager, 2016; Tsao, 2017). Chen and Kager (2016) as well as Chen et al. (2017) tested Dutch-learning infants’ discrimination of the Mandarin low-rising and low-dipping tones. Different from Mattock et al. (2008) and Yeung et al. (2013), who used familiarization in the initial exposure phase, infants were first habituated by repeatedly being exposed to one of the tones until their looking time had decreased for a predefined percentage. Then in the test phase, one trial of the habituated tone and one trial of the non-habituated tone was presented. The results from both studies suggest successful discrimination in 6- and 12-month-olds but not in 4-month-olds. The authors concluded from their results that, with increasing age, infants develop more fine-grained acoustic discrimination abilities for pitch information. Increasing perceptual sensitivity was also observed by Tsao (2017), who tested 6–8 and 10–12-month-old Mandarin- and English-learning infants using the CHT paradigm on the Mandarin high-level versus low-dipping tones. Both language groups showed discrimination at both ages and their discrimination ability was enhanced with increasing age.

A third pattern found in the literature is that infants show no changes in their discrimination ability with increasing age (Liu & Kager, 2014, 2016; Ramachers et al., 2018; Shi, Santos, et al., 2017; Tsao, 2017). Ramachers et al. (2017) tested Dutch and Limburgian[3] 6-, 9-, and 12-month-old infants with Limburgian falling vs. falling-rising tones. After the infants were habituated with one tone, they were presented with trials that only contained the habituated tone (non-alternating) or with a mixture of the habituated and the non-habituated tones (alternating). Looking time to a central visual display was the dependent measure, and results showed that Dutch infants at all ages (with no previous exposure to this specific dialect) discriminated the Limburgian tone contrast. Ramachers and colleagues (2017) argue that Dutch intonation has pitch contours (H*L and H*LH%) that are acoustically comparable to the Limburgian tones (Gussenhoven, 2004), which may have led to maintenance of discrimination. Shi et al. (2017) came to a similar result when testing French-learning 4-, 8-, and 11-month-old infants. They habituated the infants to one instance of two Mandarin tone contrasts: either one token from the perceptually close rising versus low-dipping contrast or one from the perceptually more distinct high-level versus falling contrast. Infants were then tested on their discrimination of the habituated and the non-habituated tones. The infants showed successful discrimination across all three age groups with slight indications of a decline only for the perceptually close contrast. They discuss their findings as an indication of the emerging impact of native phonology and of the acoustic salience of the tested contrast in the perception of the non-native tone patterns.

---

[3] Limburgian is a dialect of Standard Dutch that uses word-level pitch for marking lexical and grammatical differences.

Finally, a fourth developmental pattern was observed by Liu and Kager (2014), who tested the discrimination of the Mandarin high-level vs. high-falling tonal contrast in Dutch infants between 5 and 18 months of age using the visual fixation paradigm implemented with a habituation procedure. Their study revealed perceptual sensitivity at all ages when using naturally recorded speech stimuli. However, they found a U-shaped developmental curve in a second experiment, in which synthesized stimuli with smaller acoustic differences of the same contrast were used. Specifically, Dutch-learning infants at 5–6 and 17–18 months of age discriminated the contrast in these materials, but not the intermediate age groups. This U-shaped development was also found in a group of bilingual infants learning Dutch and another non-tone language (Liu & Kager, 2016a). In line with Shi et al. (2017), the authors interpreted the finding that Dutch-learning infants regain their ability to discriminate the tones as a result of their experience with the native (Dutch) intonation system and its modulation by the acoustic salience of the contrast. To our knowledge, the two studies by Liu and Kager (2014, 2016a) are the only ones that have tested tone perception across a larger age range extending into the second year of life and that have found evidence for a U-shaped learning curve.

In sum, previous studies have shown that infants’ non-native tone perception is probably influenced by a large number of factors, including age, task demands, the acoustic salience of the target tone contrast, and the prosodic systems of the native languages of the infant participants. Thus, developmental change in language acquisition and the experimental observation of this change seem to be dependent on a complex interaction of different factors. This links up with findings that show that older children and adult speakers of non-tone languages can also identify and discriminate lexical tones, even though their performance is typically below that of native speakers of the particular language (Burnham & Francis, 1997; Francis et al., 2008; Hallé et al., 2004; Hay et al., 2015; So & Best, 2010). The adult perception of L2 tones is influenced by various factors, among others by the L1 lexical tone system (if the L1 is a tone language) or the use of pitch variation for post-lexical functions, (e.g., different intonation or phrasing patterns) in the native language (Caldwell-Harris et al., 2015; Wayland & Li, 2008), but also by specific task conditions (e.g., duration of the interstimulus interval, requirement to count backwards during the interstimulus interval) that can show differential effects on non-native and native speakers’ performance (Lee et al., 1996). One explanation for good tone discrimination abilities in adult speakers of non-tonal languages is that hearers might adopt their knowledge about the native intonation system for identifying and discriminating lexical tones (Francis et al., 2008). For instance, Francis et al. (2008) found that English listeners were highly accurate in identifying the Cantonese high-rising tone, which the authors linked to the acoustic similarity of this Cantonese tone to the rising intonation pattern of questions in English. Another possibility derives from the acoustic salience of the tested contrast. Highly acoustically salient tone contrasts are easier to discriminate independent of the native language background (Hallé et al., 2004). Given these findings that tone discrimination in adult speakers of non-tonal languages is possible, but is modulated by several factors, adult speakers' performance also needs to be considered when studying perceptual reorganization of tone discrimination in early infancy.

The current study

The above-reviewed research on infants' non-native tone perception reflects the influence of several factors on experimental outcomes: acoustic properties of the tones used in the experiments, characteristics of the prosodic systems of the native languages of the participants, and also aspects of the experimental procedures. The studies that have found a perceptual decline with increasing age have mainly used familiarization procedures (Mattock et al., 2008; Yeung et al., 2013), whereas all studies that have found patterns of (re-)increased or maintained sensitivity across age have used infant-controlled habituation or conditioning procedures (Chen et al., 2017; Chen & Kager, 2016; Hay et al., 2015; Liu & Kager, 2014, 2016; Ramachers et al., 2018; Shi, Santos, et al., 2017; Tsao, 2017). This suggests that habituation may be the more robust procedure to reveal discrimination abilities in infants. In line with this consideration, a recent test–retest reliability study suggests that habituation results are more consistent and reveal larger effects at the group level than familiarization (Cristia et al., 2016). One reason for this could be that infants in a habituation procedure enter the test phase of the experiment on an individually controlled encoding status of the stimulus. The duration of the exposure during the habituation procedure is dependent on infants' response to the stimulus. In contrast, familiarization has a fixed duration that does not take into account individual differences in the speed of encoding the stimuli. According to the model by (Hunter & Ames, 1988), the degree of familiarity with the exposed stimulus (which depends on an interaction of stimulus complexity and the infants' age as an indicator of developmental level) determines whether an infant prefers the familiar or the novel stimulus in the test phase. Therefore, group results may reflect heterogeneous individual patterns of novelty or familiarity preferences, which may lead to null effects. This inconsistency in the direction of preferences is actually predicted after familiarization in some cases but is never predicted after habituation. Thus, the conflicting results on infants' tone perception obtained across different studies may at least partly be related to the use of different pre-exposure techniques.

The present study had two main objectives. First, we further investigated the U-shaped development found by Liu and Kager (2014) using another tone contrast and testing a population with a different native language than Dutch. To this end, discrimination of a Cantonese tone contrast was tested with German-learning infants between 6 and 18 months of age, as well as with a group of German and Cantonese adults. Second, we wanted to pursue the question of methodological impacts on the results in infant discrimination studies. For that reason, the effect of using a familiarization or a habituation technique on the discrimination performance of 6- and 9-month-olds was investigated by testing these two age groups with two different experimental procedures.

Before testing infants, we first asked whether the target tone contrast would be discriminated by adult speakers of German. We tested a group of German adults on their ability to discriminate Cantonese tone contrasts and compared the results to the performance of a group of adult native speakers of Cantonese. Our prediction was that German adults may be able to discriminate these tones in an AXB task but that Cantonese speakers should outperform the German speakers. An AXB task was chosen to reduce the effects of memory load. Different tokens of syllables from the same tonal category were used to force listeners to discriminate categorically rather than acoustically.

5.2 Experiment 1: Adults' discrimination of Cantonese lexical tones

5.2.1 Method

5.2.1.1 Participants

Ten native Cantonese speakers (19 to 31 years, 5 female) and 14 native German speakers (22 to 31 years, 8 female) participated in this study. None of the native German speakers had any language competence in Cantonese or another tone language. Although all participants reported L2 proficiency in English, they considered themselves to be monolingual. All participants reported normal hearing abilities. The study was approved by the Ethics Committee of the University of Potsdam. Written informed consent in accordance with the Declaration of Helsinki was obtained from all participants.

5.2.1.2 Stimuli

The stimuli for the adult experiment comprised five different Cantonese lexical tones: high-rising (Tone 25), mid-level (Tone 33), low-falling (Tone 21), low-rising (Tone 23), and low-level (Tone 22). Although our experiments with the German infants (see below) were restricted to testing the discrimination of only Tone 33 and Tone 25,[4] we examined more tone contrasts in the adult experiment. This was done in order to minimize any effects of only presenting two tones repeatedly, which may draw the participants' attention to their specific acoustic differences and thus foster enhancement of discrimination during the experiment. A second reason for including multiple tones was to generate a broader picture of German adults' processing of lexical tones.

A female native speaker of Cantonese produced 40 segmentally different CV and CVC syllables in each of these five tones leading to 200 different syllables overall (e.g., the syllables /jin/ and /se/, each produced with five different tones). Half of the stimuli were CV and the other half CVC syllables. All syllables had a legal German phonotactic structure and were meaningful Cantonese words. To create acoustic variability the speaker produced each stimulus four times. An acoustic analysis of the pitch patterns of the stimuli was conducted using PRAAT; see Table 3 (Boersma & Weenink, 2016). Pitch contours were measured by sampling at three different time

---

[4] This tone contrast was also used in the study by Yeung et al. (2013) that tested English-learning infants. Given the prosodic similarity between English and German, we expected this tone contrast to generate similar effects in German-learning infants.

points within the vowel: at initial, middle (at 50 %), and final position. Figure 2 illustrates an example of the five different pitch contours of the syllable /jin/. The pitch contour of level tones showed no change across the syllable (Tone 22, Tone 33), whereas for contour tones a pitch rise (Tone 23, Tone 25) or fall (Tone 21) occurred at the end of the syllable. For the experiment, all stimuli were normalized in intensity.

Tone	F0 initial in Hz	F0 middle in Hz	F0 final in Hz
21	183 (20)	168 (17)	162 (20)
23	176 (16)	187 (17)	214 (16)
25	183 (12)	193 (14)	229 (12)
22	198 (16)	191 (16)	193 (16)
33	211 (17)	206 (18)	207 (17)

Table 3. Results from the acoustic analysis of the different Cantonese lexical tones. All values are f0 means, Standard Deviations are given in parentheses. The analysis was done at three different positions: at the initial, middle and final position of the pitch contour.

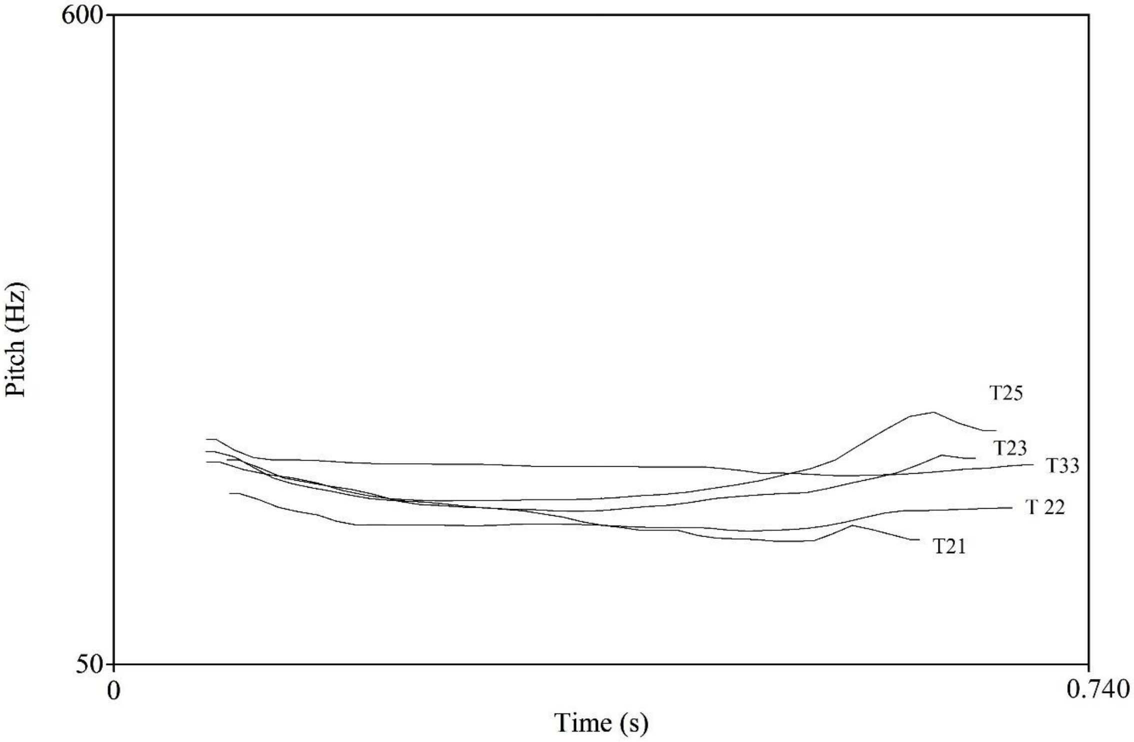

Figure 2. An example of the F0 contours of the syllable /jin/ of the five different tested Cantonese tones.

5.2.1.3 Procedure

Both Cantonese and German adults performed an AXB discrimination task. In this task, participants needed to discriminate between ten different tone pairs. The five tone types were combined with each other, such that Stimulus A and B of a trial were always segmentally identical syllables but belonged to different tone categories; X also had the same segmental structure and belonged either to the same tone category as A or as B. An AXB task was chosen to reduce the effects of memory load compared to an ABX task. The X in an AXB task is equally distant from A and B, which prevents a mapping bias to the B stimulus (Best et al., 2001; Hallé et al., 2004; Strange & Shafer, 2008). Within a trial, different tokens of the syllables from the same tonal category were used to force listeners to discriminate categorically rather than acoustically (Best et al., 1988; Polka, 1991, 1992), thereby increasing the likelihood of finding language-specific effects.

Four different trial types with the four possible orders of the stimuli were presented: AAB, ABB, BAA, and BBA. Each participant heard each of the 40 types of syllables combined with only one tone contrast. The pairing was randomized and counterbalanced across the participants (e.g., one participant heard the contrast Tone 25 – Tone 33 on the syllable /se/, while another participant heard the contras Ton 22 – Tone 33 on the same syllable). Therefore, every participant heard each of the 40 syllables during the experiment but the tone contrast that was instantiated on these syllables varied across the participants. Each tone contrast occurred with four different syllables for each participant. During the experiment, each syllable-tone pairing was presented four times, once in each trial type. This resulted in an overall number of 160 trials for each participant (4 syllables x 10 tone contrasts x 4 trial orders). These trials were divided into four blocks of 40 trials, in order to allow pauses in between. Each block only contained one of the trial types for a syllable-tone pair. The trials within a block were presented in a pseudo-randomized order with the same tone contrast never repeating twice in row. The stimuli within trials were separated by an interstimulus interval of 1000 ms; the intertrial interval was 3000 ms. An interstimulus interval of 1000 ms was chosen because previous studies have shown that language-specific effects are more clearly revealed with long interstimulus intervals (Werker & Logan, 1985). The maximum response time for the participants was 2500 ms, measured from the offset of the last syllable. The pause between blocks was controlled by the participant, and the experiment continued when the participant pressed a button. In total, the experiment lasted around 20 minutes.

Participants were instructed to decide whether the second syllable was more similar to the first or to the third syllable, otherwise they were not instructed to attend to any specific part of the syllables. The experiment and the participants’ responses on a keyboard were controlled with OpenSesame (Mathôt et al., 2012) and run on a laptop. All trials were presented over headphones in a silent room.

5.2.2 Results

Figure 3 summarizes the percentages of correct responses given for all contrasts by both language groups. Statistical analyses were run on the number of correct responses as the dependent variable. The performance of both language groups was significantly higher than predicted by chance for all tone contrasts (one sample t-test against chance level, all *p’s* < .001). This was also true for the relevant tone contrast for the infant study (Tone 33 – Tone 25). Most importantly, a one sample t-test against chance revealed above chance performance in German adults ($t = 18.55$, $p < .001$) for this contrast.

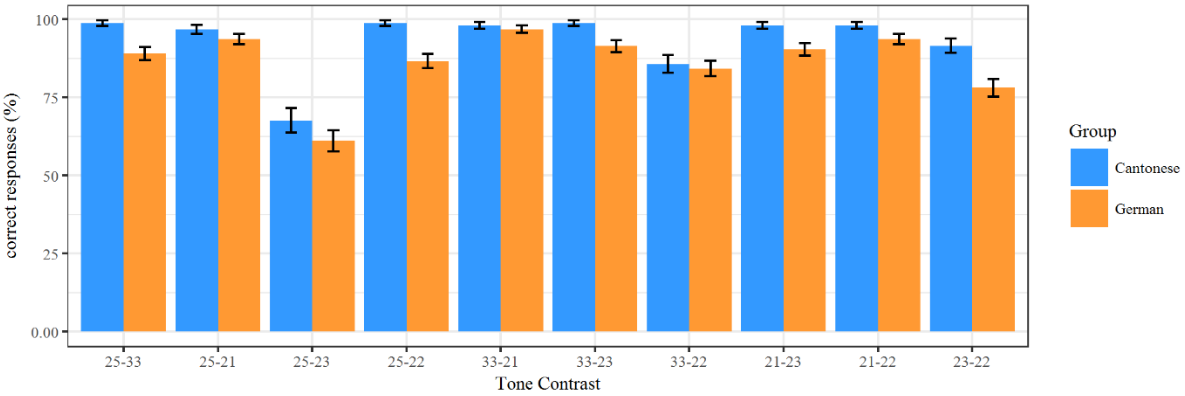

Figure 3. Results from the AXB discrimination task, separated by group and tone contrast.

As a next step, we compared different models that were computed with the function glmer from the lme4 package (Bates et al., 2015) in R (R Core Team, 2017). Models and their results were obtained by the anova function. The best fitting model (lowest Akaike Information Criterion (AIC, Akaike, 1998) and significant difference in the Chi-square test) included item and subject as random factors and interaction of language group (Cantonese and German) and tone contrast (the 10 different tone contrasts) as fixed factors; see Table 4. Additionally, we asked for musical experience. Participants were asked whether they had learned to play an instrument and if yes, how long they do or did play it. Model comparison revealed that musical experience (years playing an instrument) did not modulate the outcome of our data. Compared to the model including the interaction of Tone Contrast and Language group, the model including musical experience has higher AIC (2183.4 compared to 2175.6) and no significantly better fit with Chi-square test results (*p* = 0.19).

Model	Df	AIC	BIC	logLik	Deviance	Chisq Chi	Df	Pr (> Chisq)
~ Tone contrast + (1|subject)	11	2187.6	2255.6	-1082.6	2165.2			
~ Tone contrast + (1|subject) + (1|item)	12	2160.8	2235.5	-1068.4	2136.8	28.397	1	<.001 ***
~ Tone contrast * Group + (1| subject) + (1|item)	22	2275.6	2275.6	-1047.3	2094.7	42.114	10	<.001 ***

Table 4. Results from the model comparison of the adult perception experiment. The comparison is organized hierarchically. The first model was compared to the second model – which fit better to the data. The second model was then compared to the third and so forth. The comparison revealed best fit for the model which includes the

interaction of tone contrast and group as fixed effect and subject and item as random effects (*** indicates p < .001).

In general, our results reveal good performance in both groups, but show that German native listeners performed less accurately than the native Cantonese listeners (86.5% vs. 93.4%, respectively). The statistical analysis showed that the overall performance differed significantly between the two language groups (β = -2.253, *SE* = 0.758, *z* = -2.973, *p* < .01). However, this group difference was not significant across all contrasts as indicated by the interaction of tone contrast with group. Cantonese listeners best discriminated high-rising (25) vs. mid-level (33), high-rising (25) vs. low-level (22), and mid-level (33) vs. low-rising (23), each at a level of 98.7%. German adults performed best on the discrimination of mid-level (33) vs. low-falling (21). For both groups, the contrast high-rising (25) vs. low-rising (23) was the most difficult contrast.

With respect to the infant experiments, we were especially interested in how native and non-native adults perceive the difference between high-rising and mid-level tones. Our results revealed that the Cantonese adults discriminated Tone 25 vs. Tone 33 significantly better than the German listeners (β = -2.503, *SE* = 0.871, *z* = -2.874, *p* < .01). Furthermore, native listeners discriminated Tone 25 vs. Tone 22 (β = -2.567, *SE* = 0.786, *z* = -3.265, *p* < .01), Tone 33 vs. Tone 23 (β = -2.047, *SE* = 0.850, *z* = -2.409, *p* < .01), Tone 21 vs. Tone 23 (β = -1.818, *SE* = 0.713, *z* = -2.549, *p* < .05), and Tone 23 vs. Tone 22 (β = -1.127, *SE* = 0.336, *z* = -3.358, *p* < .001) significantly better than the non-native German listeners. The discrimination for the other tone contrasts was not significantly different between the Cantonese and the German listeners.

5.2.3 Discussion

The first experiment tested the discrimination of Cantonese lexical tones by adult German listeners without knowledge of Cantonese and by native speakers of Cantonese. Three main findings were obtained: First, German native speakers were able to distinguish between different lexical tones. Second, native Cantonese speakers outperformed German listeners in their overall discrimination abilities. Third, there was variation in German listeners' discrimination performance depending on the specific contrast: while the discrimination reached native-like levels for some contrasts, performance was below that of native speakers for other contrasts. This is in line with other discrimination studies that have shown good discrimination by non-native listeners, but an overall better performance by native listeners (Burnham & Francis, 1997; Cutler & Chen, 1997; Francis et al., 2008; Lee et al., 1996).

However, the picture becomes less clear when comparing performances of each tone contrast separately. Some lexical tones (high-rising vs. mid-level, high-rising vs. low-level, low-rising vs. mid-level, low-rising vs. low-level, and low-rising vs. falling) are harder to discriminate for German than for Cantonese native speakers. However, there are also contrasts for which both language groups show comparable levels of high performance (high-rising vs. low-falling, mid-level vs. low-falling, and low-level vs. falling). Further, there are two contrasts for which both language groups show comparably lower performance (high-rising vs. low-rising, mid-level vs. low-level). It is striking that the pairs that are highly discriminable by both groups contain one level and one contour tone or two contour tones with frequency changes in opposite directions, while the tone pairs that are harder to discriminate are both level tones or show the same direction of frequency change. This pattern suggests that for non-native as well as for native tone discrimination, acoustic properties and the acoustic distance of the specific tone contrast are relevant for their discriminability. In addition, it is possible that German listeners assimilate some of the tones to their native intonation system. This would then support a language-specific account of adult tone perception. It is noteworthy that all contrasts that are highly discriminable for the German listeners contain the falling Tone 21. The good discrimination seen here might stem from familiarity with the German intonation system, which uses falling contours for neutral statements (Grice & Baumann, 2002). That is, similar to what Francis et al. (2008) have proposed for English listeners, German native speakers might use their knowledge of the native intonation system to discriminate non-native lexical tones.

To summarize, our findings from the first experiment revealed that German native speakers discriminate Cantonese lexical tones highly accurately, but native listeners perform significantly better. The overall good discrimination performance for German listeners could be explained by acoustic salience and/or assimilation to the native prosody. Our results thus showed that native and nonnative adults’ performance may differ depending on the specific contrast. Discrimination abilities in adults should therefore be considered before testing potential changes in infants’ non-native sound discrimination. Overall, the most important finding from our first experiment is that

German adults can discriminate the tone contrast that was used in our infant studies (Tone 33 vs. Tone 25), but that their performance was below that of native speakers of Cantonese. The finding that German adults can hear the difference between these tones increases the likelihood of observing a U-shaped developmental pattern, or perceptual enhancement with increasing age. But the finding that native Cantonese listeners show higher achievements in discriminating these two tones suggest that their discrimination is not only due to a large acoustic distance but is also affected by the native language of the listener.

5.3 Experiment 2: Testing 6-, 9-, and 18-month-olds using a familiarization procedure

Here we contribute new data to the infant tone perception literature by testing German infants' perception of the Cantonese Tone 33 vs. Tone 25 contrast that had previously been used in a study with English-learning infants by Yeung et al. (2013). Similar to Liu and Kager (2014), we included a wider age range than Yeung et al. had done in order to test for evidence of a U-shaped developmental curve in German 6-, 9-, and 18-month-olds. Following the Yeung et al. study, we used a procedure involving familiarization, but the discrimination abilities during the test phase were assessed with the head-turn preference procedure.

5.3.1 Method

5.3.1.1 Participants

In total, 88 monolingual German-learning infants participated in this experiment: 30 6-month-olds (M_{age} = 182 days; *range* = 168–194 days; 14 girls), 30 9-month-olds (M_{age} = 275 days; *range* = 258–289; 18 girls), and 28 18-month-olds (M_{age} = 540 days; *range* = 526–556 days; 13 girls). An additional 16 infants were tested but excluded from the analysis for the following reasons: crying (n = 8), fussiness (n = 5), technical error (n = 1), and pre-term (n = 2). Another two infants were excluded because at least one of the main caretakers grew up in an area in which the local German dialect uses word-level pitch contrasts (Werth, 2011). The remaining infants were all born full-term. According to parental report, infants did not suffer from repeated or acute ear infections, and there were no indications of atypical development or any experience with a tone language. This study was carried out in accordance with the recommendations of the

Ethics Committees of the University of Potsdam with written informed consent given by the parents in accordance with the Declaration of Helsinki.

5.3.1.2 Stimuli

For this study, we used the stimuli from Yeung et al. (2013): Cantonese CV syllables (/tɕʰi/) with either a high-rising (Tone 25) or mid-level (Tone 33) tone. In total, there were four different tokens of each tone. For detailed acoustic properties of the syllables, see Yeung et al. (2013).

The familiarization phase included only tokens of either Tone 25 or Tone 33. During the test phase, single syllables were concatenated into two different types of sequences: non-alternating (tokens from one tone category) and alternating sequences (tokens from both tone categories). In total, the test phase contained eight trials: four non-alternating and four alternating trials. Two of the non-alternating trials included only tokens of Tone 25 and the other two only of Tone 33. In the alternating trials, tone types were intermixed: the first four tokens at the beginning of the trial alternated between the two tones, the following ones were in a random order. The tokens were separated by an interstimulus interval of 1 s. Half of the alternating trials started with Tone 25, the other half with Tone 33, and they contained the same number of both tone types. During the familiarization phase, the maximal trial length was 15 s and during the test phase it was 30 s.

5.3.1.3 Procedure

The experiment was run with the head-turn preference procedure (Hirsh-Pasek et al., 1987; Jusczyk & Aslin, 1995), which differed from Yeung et al.'s use of visual fixation, but still measured auditory preference by recording the duration of attention to a visual stimulus while being presented to an acoustic stimulus. Infants sat on their caretakers' lap in a booth and first fixated on a flashing green lamp in front of them. Next, the experimenter – who sat in a second room and monitored the infants' gaze via a camera mounted above the green light – started the experimental trial by pressing a button. Then, one of the red lights mounted on the left or the right side inside the booth began to flash. As soon as the infant fixated the now blinking red light, the experimenter started the acoustic stimulus. The trial ended when the infant either looked away for more than 2 s, or when the end of the acoustic stimulus was reached. To start the next trial, the experimenter pressed a button and the green light in front of the infant again began to flash. Infants' looking duration (listening time) was coded online by the experimenter via a button box connected to a computer.

The experiment consisted of a familiarization and a test phase. During the familiarization phase, infants were presented with only one of the tones (either Tone 25 or Tone 33, counterbalanced across participants) until they had accumulated 30 s of listening time. A maximal trial length of 15 s assured that the infant looked at least once to both sides of the sound source during the familiarization. The test phase followed immediately after the familiarization phase and consisted of eight trials: two non-alternating trials of Tone 25, two non-alternating trials of Tone 33, and four alternating trials. These eight test trials were the same for all infants. During the test phase, the presentation order of alternating and non-alternating trials was pseudo-randomized; two alternating or non-alternating trials never followed each other directly (i.e., N-A-N-A-N-A-N-A or A-N-A-N-A-N-A-N). The test phase was additionally divided into two blocks: in each block, each trial type (alternating, non-alternating Tone 25, non-alternating Tone 33) was presented at least once. The presentation order of alternating, non-alternating Tone 25, and non-alternating Tone 33 was counterbalanced across infants, so that each of the trial types was presented in every position during the test phase. To check the reliability of the online measures of listening time (which was automatically calculated based on the experimenter's button pressing), 50% of the videos (randomly selected) obtained during the experimental session were re-coded by a second experienced coder using specialized software ELAN (Wittenburg et al., 2006). The inter-coder reliability was Pearson's $r = 0.99$, $p < .001$.

5.3.2 Results

The averaged listening times for each trial type were entered as dependent variable into the statistical analysis. The mean listening times separated by age group and condition are displayed in Figure 4. For the statistical analysis, listening times were logarithmically transformed in order to create a normal distribution of the residuals. Data were analyzed with R (R Core Team, 2017) and linear mixed models with the lmer function from the package lme4 (Bates et al., 2015). Model comparison revealed that the model including the interaction of Condition (alternating, non-alternating Tone 25, and non-alternating Tone 33) × Age Group (6-, 9-, and 18-months) as fixed effect and trial number and subject as random factors fit best to our data (Table 5). This indicates that the listening times are differently affected by the conditions and the age.

Furthermore, the comparison revealed that the tone used for the familiarization did not modulate the results, as including this factor did not improve the model fit (indicated by higher AIC and no significant difference in the Chi-square test).

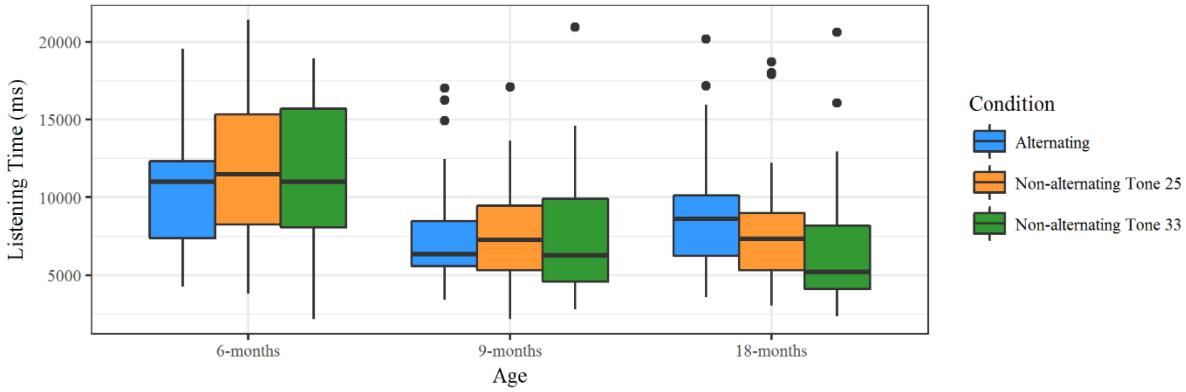

Figure 4. Results from the familiarization experiment divided by age group. Mean listening times for the alternating trials were only significantly longer at 18 months, indicating that only the 18-month-olds discriminated the lexical tones.

Model	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr (> Chisq)
~ Condition + (1|subject)	5	1515.6	1538.3	-752.78	1505.6			
~ Condition*Age + (1|subject)	11	1502.5	1552.7	-740.26	1480.5	25.04	6	<.001 ***
~ Condition + (1|subject) + (1|trial_number)	12	1457.6	1512.2	-716.78	1433.6	47.00	1	<.001 ***
~ Condition*Age* Familiarization + (1|subject) + (1|trial_number)	21	1469.5	1565.1	-713.72	1427.5	6.11	9	0.729

Table 5. Results from the model comparison of the familiarization procedure. The comparison is organized hierarchically. The first model was compared to the second model – which fit better to the data. The second model was then compared to the third and so forth. Trial number refers to each individual trial, familiarization refers to the type of familiarization tone. Results from the Chi-square test and AIC score revealed best model fit for the model which includes the interaction of Age and Condition as fixed effect and subject and trial number as random effects. (*** indicates p < .001).

As the model showed a significant interaction of Age Group × Condition, we calculated separate models for each age group. Detailed statistical information for all age groups is provided in Table 6. These models also included subject and trial number as random factors and Condition as fixed effect. Familiarization was not included as a fixed effect, as the previous general model did not show an effect for the familiarization tone.

6-month-olds				
Fixed effects	Estimate β (SE)	df	t value	Pr(>|t|)
(Intercept) Alternating	3.923 (0.062)	17.1	63.809	<.001
Non-Alternating Tone 25	0.066 (0.040)	203.1	1.677	0.095
Non-Alternating Tone 33	0.041 (0.040)	203.1	1.036	0.302
9-month-olds				
(Intercept) Alternating	3.788 (0.037)	21.14	102.033	<.001
Non-Alternating Tone 25	-0.004 (0.041)	202.52	-0.099	.921
Non-Alternating Tone 33	-0.038 (0.041)	202.52	-0.926	.356
18-month-olds				
(Intercept) Alternating	3.834 (0.043)	21.43	89.972	<.001
Non-Alternating Tone 25	-0.031 (0.049)	189.65	-0.616	.539
Non-Alternating Tone 33	-0.104 (0.049)	186.65	-2.097	.037 *

Table 6. Detailed results of the statistical analysis of the familiarization experiment for each age group. The estimates represent the log-transformed listening times. The results indicate that only the 18-month-olds discriminate the contrast by longer listening times to the alternating trials, but not the 6- and 9-month-old infants. All models included Condition as fixed effect and subject and trial number as random effects as revealed as the best fit by the overall model comparison (* indicates p < .05).

For the 6-month-olds, the listening times for the alternating trials (*M* = 10.6 s, *SD* = 7.9 s) did not differ significantly from the listening times for the non-alternating Tone 25 trials (*M* = 11.9 s, *SD* = 7.6 s) nor from those for the non-alternating Tone 33 trials (*M* = 11.3 s, *SD* = 7.9 s). The effect sizes (Cohen's *d*) for alternating vs. non-alternating Tone 25 were *d* = -0.249, and for alternating vs. non-alternating Tone 33 *d* = -0.108.

The 9-month-olds also did not show significant differences in their listening times for the alternating trials (*M* = 7.63s, *SD* = 5.17 s) compared to the non-alternating Tone 25 (*M* = 7.55 s, *SD* = 4.89 s) or the non-alternating Tone 33 (*M* = 7.53 s, *SD* = 5.93 s) trials. The effect sizes (Cohen's *d*) for alternating vs. non-alternating Tone 25 were *d* = -0.009, and for alternating vs. non-alternating Tone 33 *d* = 0.132.

However, the 18-month-olds showed significantly longer listening times for the alternating trials (*M* = 9.07 s, *SD* = 6.87 s) than for the non-alternating Tone 33 trials (*M* = 6.89 s, *SD* = 5.47 s). The difference between alternating and non-alternating Tone 25 trials (*M* = 8.15 s, *SD* = 5.90 s) was not significant. The effect sizes (Cohen's *d*) for alternating vs. non-alternating Tone 25 trials were *d* = 0.087, and for alternating vs. non-alternating Tone 33 trials *d* = 0.323.

5.3.3 Discussion

The results from this experiment did not provide evidence that 6- and 9-month-old German-learning infants discriminate the Cantonese Tone 25 – Tone 33 contrast. Only the 18-month-olds showed discrimination abilities for this contrast. However, discrimination showed up only in the comparison of the listening times to alternating sequences and non-alternating sequences containing Tone 33. No evidence of discrimination occurred between alternating sequences and non-alternating sequences that only contained Tone 25. This indicates some kind of asymmetry in the perception of these tones by German 18-month-olds.

Taken together, these results are only partly congruent with our prediction of perceptual reorganization and a U-shaped learning curve in tone perception. On the one hand, the differences in the results between the 9- and the 18-month-olds are in line with the observations by Liu and Kager (2014), who report an increase in the discrimination of Mandarin tone contrasts by Dutch-learning infants across these ages. Furthermore, our finding that 18-month-olds discriminate the tones is in line with our findings from Experiment 1 since German adults can also discriminate this contrast. However, what is missing is evidence of a decline in perceptual sensitivity between 6 and 9 months of age, as neither the 6- nor the 9-month-olds gave any indication of discriminating the contrast. So far, our result pattern for German-learning children is mostly compatible with the hypothesis of an age-related enhancement in tone perception, which is consistent with the findings of previous studies with Dutch-learning (Chen & Kager, 2016; Chen et al., 2017) or English-learning infants (Tsao, 2017). Given the fact that German 7- to 8-month-old infants have been shown to be sensitive to pitch variations (Abboub et al., 2016; Wellmann et al., 2012), the assumption that even 9-month-olds may not yet be able to discriminate the tone contrasts based on pitch information is not likely. However, it might be that infants at this age focus on sound contrasts that mark lexical distinctions in their native language. Since this is not the case for pitch differences on the syllabic level, 9-month-olds might ignore these pitch differences.

There may be at least two other potential explanations for our failure to find indications of a decline in discrimination in the two younger age groups that we tested. The first one is that perceptual reorganization for these tone contrasts has set in before 6-months of age. Remember that Yeung et al. (2013) tested 4- and 9-month-old but not 6-month-old English-learning infants with the same tone contrasts as were used with the German infants. They found discrimination in 4-month-olds but not in the 9-month-olds. Comparing the English-, Mandarin- or Cantonese-learning 4-month-olds in that study revealed that all language groups discriminated between the tones, but that the preference patterns for the different stimulus types were not the same across the groups. This suggests language-specific influences on tone perception already at this early age, leaving open the possibility that we would have found evidence for perceptual reorganization in German infants younger than 6 months. Nevertheless, a number of other studies using different stimuli and testing infants exposed to different languages found non-native tone discrimination in 6-month-olds (Mattock et al., 2008; Mattock & Burnham, 2006). This suggests that the perceptual decline for lexical tone contrasts is not necessarily completed by the age of 6 months.

A second explanation for our failure to find evidence for changes in the younger infants' tone perception is methodological in nature: the method used in our experiment may not have been suitable to demonstrate infants' ability to discriminate the tones. As argued above, the effect of familiarization may be modulated by characteristics of the stimuli and the participants, making this type of pre-exposure not optimally suitable to uncovering discrimination abilities for all types of stimuli at all ages. Hence, our third experiment reinvestigated 6- and 9-month-olds' discrimination of the same contrasts as in the previous experiment but using a habituation procedure during the exposure phase.

Before we come to the third experiment, the results of the 18-month-olds deserve some consideration. As stated above, their listening times were longer for the alternating trials compared to the Tone 33 non-alternating trials, but not compared to the Tone 25 non-alternating trials. This pattern seems to be caused by enhanced listening times for the non-alternating Tone 25 sequences (compared to the non-alternating Tone 33 sequences). Listening times reflect specific preferences that infants have for stimuli that are presented during the experiment, and such preferences can emerge in the course of the experiment (when a familiarization phase is included) or can also be caused by some inherent properties of the stimuli (e.g., acoustic saliency, familiarity, etc.). Our results suggest that for German-learning infants, high-rising tones attract more attention compared to mid-level tones. Interestingly, Yeung et al., (2013) also found that the Mandarin-learning (but not the Cantonese-learning) infants showed longer listening times to Tone 25 compared to Tone 33. In contrast, the English-learning 4-month-olds showed a preference for listening to Tone 33 compared to Tone 25. The authors suggested that these differences in preference speak against an acoustic explanation that applies across languages, but rather suggests a language-specific preference for a certain tone type. A similar explanation may hold for the results of the German 18-month-olds. Their greater attention to Tone 25 than to Tone 33 indicates that they prefer pitch contours over level tones, which may be driven by the function that pitch contours have in German. In intonation languages like German, rising pitch contours often occur at the end of clauses, where a pragmatic function is to mark the utterance as a question or to indicate that the sentence is not yet finished (Grice & Baumann, 2002; Spinelli et al., 2017). The preference for the Cantonese contour Tone 25 may thus be interpreted as an indication that the 18-month-old German infants have started to learn about these pragmatic functions of rising contours. We will discuss this point in more detail in the general discussion.

5.4 Experiment 3: Testing 6- and 9-month-olds using a habituation procedure

5.4.1 Method

5.4.1.1 Participants

Thirty monolingual German-learning infants participated in this experiment: 15 6-month-old (M_{age} = 182 days, *range* = 168–195 days; 8 girls) and 15 9-month-old (M_{age} = 207 days, *range* = 255–289 days; 7 girls) infants. An additional 12 infants were tested but excluded from the analysis for the following reasons: crying (*n* = 3), failure to reach the habituation criterion (*n* = 7), listening times < 500 ms for at least one of the four test trials (*n* = 1), and fussiness (*n* = 1). Infants from Experiment 3 did not participate in the previous Experiment 2. All infants were born full-term and according to parental report none of the infants suffered from any repeated or acute ear infections. None of the infants showed indications of atypical development or had experience with a tone language. This study was carried out in accordance with the recommendations of the Ethics Committees of the University of Potsdam with written informed consent from all parents in accordance with the Declaration of Helsinki.

5.4.1.2 Stimuli

The tone contrast for this experiment was identical to the contrast in Experiment 2. For habituation and test phases, the same four tokens as used in Experiment 2 were re-arranged into new sound files. Since we had four tokens of each tone, we decided to use all tokens in the habituation and test phases in order to allow more acoustic variation within each phase. Stimuli were separated by an interstimulus interval of 1 s, resulting in a speech string of 40 s. During the experimental trials, a black and white checkerboard was displayed on a screen (e.g., (Horowitz, 1974; Stager & Werker, 1997). Between trials, infants saw a silent bouncing ball to redirect their attention to the screen.

5.4.1.3 Procedure

Infants sat on the caretaker’s lap, facing a monitor at a distance of approximately 1.2 meters in a silent room. A camera positioned above the presentation screen monitored infants’ looking behavior. The stimulus presentation and infants’ looking behavior was coded online using Habit 2 (Version 2.1.25, (Oakes et al., 2015). All acoustic stimuli were presented with an intensity of 65 dB over loudspeakers, which were placed behind the screen. One trial consisted of a 40 s speech string. Trials started as soon as the infant fixated the screen and the experimenter pressed a key. The length of each trial was controlled by the infant’s behavior: the trials ended when infants either looked away for more than 2 s, or the maximum trial duration was reached.

The experiment consisted of three phases: habituation, test, and post-test phase. The maximum number of trials within the habituation phase was 18 trials. The habituation criterion was reached when infants’ mean listening time across three consecutive trials decreased to 50% of the mean listening time of the first three trials. Infants who did not reach the criterion were excluded from the analysis. All infants were habituated with Tone 25. The test phase started immediately after infants reached the habituation criterion or after the maximum number of trials was presented. In the test phase two trials with the novel (Tone 33) and two trials with the habituated (Tone 25) tone, each with a maximum duration of 40 s, were presented. The presentation order of the two novel and habituated tone trials was counterbalanced across infants. Half of the infants started the test phase with a trial containing the novel tone and the other half with a trial containing the habituated tone. A post-test phase followed directly after the test phase. During the post-test phase, a completely novel auditory stimulus was presented to verify the infants' attention to the task. The post-test trial differed segmentally from the tone stimuli. In total, 50% of the participants (randomly selected) were re-coded (frame by frame, 25 fps) by a second coder using the specialized software ELAN (Wittenburg et al., 2006). The inter-coder reliability was $r = .98$, $p < .001$.

5.4.2 Results

The averaged listening times for the novel and the habituated stimuli served as dependent variable. Mean listening times to the different trial types for the two age groups are displayed in Figure 5. Discrimination is indicated by a longer listening time for either the novel or the habituated tone. On average, infants needed about 6.08 trials (*SD* = 4.1) to reach the habituation criterion. Both age groups accumulated a comparable amount of listening time to the stimuli during habituation (91.95 s at 6 months, and 91.55 s at 9 months).

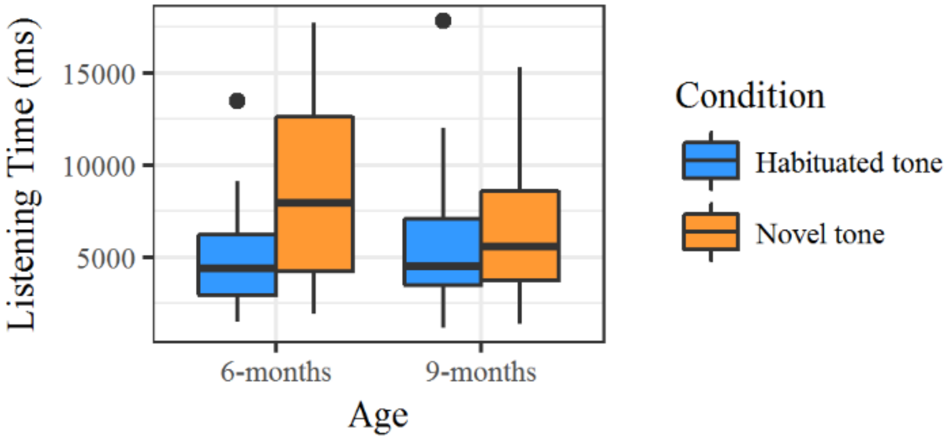

Figure 5. Results from the habituation experiment divided by age group. Mean listening times to the novel tone were significantly longer compared to those to the habituated tone.

Again, all listening times were logarithmically transformed to fulfill the assumption of normal distribution of the residuals. The statistical analysis was performed with R (R Core Team, 2017) by using linear mixed models with the lmer function from the package lme4 (Bates et al., 2015). Again, we compared different models in order to test the best model fit using the anova function. The results from a

Chi-square test as well as the lowest AIC revealed best fit for a model including the interaction of Age Group (6- and 9-months) and Condition (novel and habituated tone) as fixed factor and subject as random factor. In contrast to Experiment 2, trial number did not lead to a better model fit and was therefore excluded from further analysis. The missing effect of trial number was probably due to the smaller number of test trials. For details on the statistical analysis, see Table 7.

Model	Df	AIC	BIC	logLik	deviance	ChisqChi	Df	Pr(>Chisq)
~ Condition + (1|subject)	5	345.11	360.13	-167.55	335.11			
~ Condition* Age + (1| subject)	8	341.99	366.02	-163.00	325.99	9.12	3	0.028 *
~ Condition* Age + (1| subject) + (1| trial_number)	9	343.99	371.03	-163.00	325.99	0	1	1

Table 7. Results from the model comparison of the habituation paradigm. The comparison is organized hierarchically. The first model was compared to the second model – which fit better to the data. The second model was then compared to the third. The second model fit best to the data and included Age and Condition as fixed effects and subject as random effects. In contrast to Experiment 2, trial number did not lead to a better model fit and was therefore excluded from further analysis. Note that habituation type was not included in the models because all infants were habituated with the same tone (Tone 25) (* indicates p < .05).

Since the interaction of Condition × Age Group was found to be significant, we performed separate analyses for each age group. Detailed statistical information can be found in Table 8. All comparisons were also calculated with the lmer function with Condition as fixed factor and subject as random factor. The 6-month-olds showed significantly longer listening times to the novel tone (*M* = 8.52 s, *SD* = 5.24 s) compared to the habituated tone (*M* = 5.11 s, *SD* = 4.30 s). In contrast, the 9-month-olds' listening times to the novel tone (*M* = 6.31 s, *SD* = 5.15 s) were not significantly different from those to the habituated tone (*M* = 5.98 s, *SD* = 4.12 s). Effect sizes (Cohen's *d*) were calculated for the 6-month-olds, *d* = -0.435, and for the 9-month-olds, *d* = 0.048.

6-month-olds				
Fixed effects	Estimate β (SE)	df	t value	Pr(>|t|)
Intercept (Habituated_tone)	8.328 (0.162)	24.95	51.437	<.001 ***
Novel_tone	0.337 (0.158)	45.0	2.136	.038 *
9-month-olds				
Intercept (Habituated_tone)	8.386 (0.17576)	23.92	47.715	<.001 ***
Novel_tone	0.040 (0.164)	45	0.243	0.81

Table 8. Detailed results from the statistical analysis of the habituation experiment for each age group. The estimates represent the log-transformed listening times. Results indicated that the 6-month-olds discriminate between Tone 25 and Tone 33, but the 9-month-olds do not. All separate models included Condition as fixed effect and subject as random effect (* indicates $p < 0.05$, *** indicates $p < .001$).

5.4.3 Discussion

Our results from the habituation experiment clearly show an age-related decline in perceptual sensitivity for the contrast of Cantonese high-rising and mid-level lexical tones. While the 6-month-olds succeed in discriminating the tones, the 9-month-olds did not show any evidence of discrimination. The decline in perceptual sensitivity between 6 and 9 months is in line with previous studies on lexical tone perception in infants (Liu & Kager, 2014; Mattock & Burnham, 2006; Mattock et al., 2008). These findings support the idea of perceptual reorganization for lexical tones between the ages of 6 and 9 months (Mattock et al., 2008; Mattock & Burnham, 2006; Yeung et al., 2013) and extend this observation to German-learning infants.

5.5 General discussion

The studies presented here pursued two main goals. The first one was to investigate whether further evidence can be obtained for a U-shaped development in the discrimination of non-native tone contrasts that is characterized by an initial decline and a later re-increase of perceptual sensitivity. The second goal was to investigate whether a procedure that involves habituation in the exposure phase of the experiment provides clearer evidence of infants' discrimination of lexical tones than a procedure that uses familiarization during the exposure phase of the experiment.

Summarizing the results across the three experiments, our overall findings suggest a U-shaped developmental pattern for tone discrimination in speakers and learners of German. First, German adults are able to discriminate the Cantonese high-rising versus mid-level tones although their performance was below that of native Cantonese speakers. Second, we found a decline in the ability to discriminate these tones between the ages of 6 and 9 months: while 6-month-olds showed a clear dishabituation and thus discrimination effect in our last experiment, the results from the 9-month-olds did not indicate any discrimination of the tones across the two experiments. Third, evidence for a decline between the ages of 6 and 9 months was only obtained after habituation, but not after familiarization. We will first discuss the implications of our findings for the understanding of perceptual reorganization in infants and then consider methodological implications.

Understanding developmental trajectories for tone discrimination

Overall, the results from our study suggest a developmental trajectory in the tone discrimination of German-learning infants that is identical to what Liu and Kager (2014) found for Dutch-learning infants: good discrimination at 6 and 18 months of age, but not at 9 months. Our study extends the findings from Liu and Kager (2014), who used the Mandarin high-level and high-falling tones, to a different tone contrast from another language and to learners of a different L1. This is an important finding as it shows that the U-shaped developmental pattern that was reported for the first time by Liu and Kager (2014) can be replicated and does indeed generalize to a new tone type. In addition, our study revealed that the tone contrast that was used in our infant study can also be discriminated by adult speakers of German, but on a significantly lower level than by native speakers of Cantonese. Contrastingly, for other tone contrasts tested in Experiment 1, discrimination reached native-like performance in adult speakers of German. This suggests that the adult discrimination of Tone 25 and Tone 33 is not only based on the acoustic saliency of the phonetic contrast. This in turn suggests that the U-shaped developmental pattern for this tone contrast is based on perceptual reorganization influenced by the acquisition of phonological properties of the native language and is not only due to a change in the acoustic sensitivity to pitch information.

As already discussed in previous studies (Liu & Kager 2014, 2017; Ramachers et al., 2018; Shi et al., 2017), we assume that the intonation system of the native language and the relation of the tested non-native tone contrast to this system is crucial. Changes in pitch contours are not a unique characteristic of tone languages, as they are also relevant for the intonation of languages like German. In intonation languages, pitch movements have post-lexical functions indicating prosodic (and syntactic) phrasing and pragmatic functions, such that infants growing up with a non-tone language are not fully naïve to pitch variations. In German, rising pitch contours with a nuclear pitch accent (L*H) are related to sentence internal boundaries of prosodic phrases and to Yes-No Questions (Grice & Baumann, 2002; Gussenhoven, 2004; Petrone et al., 2017). Since questions are frequently used in communication with infants and toddlers to catch their attention (Spinelli et al., 2017), and even infants and toddlers show discrimination of question over declarative intonation contours (Geffen & Mintz, 2011; Soderstrom et al., 2011), our finding that German toddlers discriminate high-rising from mid-level tones at 18 months of age lines up with findings from other studies that assume that the native language intonation system has an impact on lexical tone perception in speakers of non-tone languages. Their growing knowledge of German intonation and its relation to the syntactic and pragmatic system may have sharpened, or re-sharpened, their processing of the tonal information in the Cantonese stimuli. However, 5-month-old English-learning infants can discriminate between statements and questions marked by their different prosodic contours (flat vs. rising contour: (Geffen & Mintz, 2011; Soderstrom et al., 2011) and German 8-month-olds can detect phrase boundaries that are marked by pitch changes in combination with final lengthening (Wellmann et al., 2012). Given these results, the question arises why a decline in perceptual sensitivity to pitch as marking lexical tone is observed in learners of non-tone languages.

If the assumption that growing knowledge about the language-specific intonation system affects tone discrimination is correct, then the discrimination abilities of 6-month-olds and that of 18-month-olds probably do not rely on the same mechanisms. Discrimination of non-native contrasts in young infants has typically been attributed to extremely sensitive acoustic perception in early development (Aslin et al., 1998), which allows the discrimination of all kinds of minimal sound contrasts. Perceptual reorganization then maintains or sharpens the discrimination of contrasts that are relevant in the linguistic system of the native language, but leads to a decline in the discrimination of sound contrasts that are not relevant in the linguistic system. Thus, we assume that the younger infants still process tone stimuli in a more acoustic manner, and while an infant’s native language is expected to influence these results (cf. Yeung et al.’s 2013 findings of language-specific differences in preferences for pitch contours across languages at 4 months of age), there should not be any decline in the ability to perceive differences in contours until a point in the development when infants must learn the linguistic functions of either tonal or intonational contrasts.

The results from the experiment using the habituation procedure with 6- and 9-month-old German infants, along with prior work illustrating the classic pattern of perceptual reorganization, suggest that 9 months of age is perhaps a critical age of interest (Liu & Kager, 2014; Mattock & Burnham, 2006; Mattock et al., 2008; Yeung et al., 2013). Because our (null) results for 9-month-olds were obtained across both experimental paradigms, we do not consider them to be a reflection of methodological issues. We propose that this decrease in tone discrimination around 9 months is an indication of a milestone in infants’ linguistic development, when infants begin to reorganize their perceptual systems to understand how pitch is functionally used in their target language, with an emphasis on word-level meanings. For infants learning German, within-syllable pitch information is not lexically informative, and so like other 9-month-old learners of non-tonal languages, they may start to ignore pitch cues from this age.

A study by Hay et al. (2015) provides data that is related to this general idea. They found that 14-month-old English-learning infants can still use a Mandarin rising and falling tone contrast in word learning by mapping novel objects to labels that differ only in pitch contours. However, 17- and 19-month-olds tested with the same procedure did not respond to this labeling violation (for similar results with English-learning 2-year-olds, see (Quam & Swingley, 2010). Testing the 19-month-olds on pure discrimination of the tones using a habituation task further revealed that these older infants could nevertheless discriminate the target tones. Hay et al. (2015) discuss this change across ages as an indication that infants get increasingly more specific about the sound contrasts that they consider to be lexically contrastive. Therefore, the older toddlers do not attend to tone contrasts in a word learning scenario, although they can discriminate them in other contexts. Infants and toddlers in our study were younger, but it may still be the case that their performance reflects shifts in attention related to lexical development. As Bergelson & Swingley (2012) have shown, infants from 6 to 9 months of age may already be strongly focused on word learning, and may be particularly attuned to sound contrasts that are lexically contrastive in their language (i.e., German), while largely ignoring sound contrasts that are not. In intonation languages, attention to tonal information may then potentially increase again when children start to detect semantic or pragmatic functions of the intonational patterns in their language which could explain why at 18-months German and Dutch infants again showed discrimination of the lexical tones. Further research would be necessary to test this hypothesis.

Future research must explore these ideas further, as lexical development might not be the only factor explaining the dip in discrimination abilities. Other factors, like salience of the contrast, might interact with the lexical development: for example, previous tone discrimination studies have not shown a perceptual decline at 9 months for certain tone contrasts (Liu & Kager, 2014, 2016; Ramachers et al., 2018; Shi, Santos, et al., 2017; Tsao, 2017). Relatedly, a perceptual shift has also been reported in the visual domain around the same age, suggesting parallel development across perceptual domains. Data from (Lewkowicz & Hansen-Tift, 2012) have also shown a U-shaped function in visual scanning, such that infants around 8 to 9 months of age look at the mouth, whereas 4- and 12-month-olds look at the eyes. This shift may be symptomatic of a general increase in attention to certain units (segmental relative to suprasegmental information). Much remains unclear about why infants from 8 to 10 months of age show a specific developmental pattern with respect to tone perception.

Methodological comparisons

The difference in the 6-month-olds' results between the familiarization and the habituation experiment line up with previous research, since most other studies have shown lexical tone discrimination with habituation procedures (Chen & Kager, 2016; Liu & Kager, 2014, 2016; Ramachers et al., 2017; Shi et al., 2017; Tsao, 2017), whereas a decline in perceptual sensitivity has mostly been found with studies using a familiarization procedure (Mattock et al., 2008; Mattock & Burnham, 2006; Yeung et al., 2014). Similar to the findings from Cristia et al. (2016), our results show that the habituation procedure generates larger effect sizes at the group level. Both habituation and familiarization procedures are based on the customization of the participants to one type of stimulus and then measuring differences in the response to the old versus a new stimulus. As stated in the introduction, we assume that habituation procedures are more adapted to individual variation by only stopping the initial exposure phase when the behavior of the infant indicates a specific level of customization. In contrast, familiarization-based procedures use a fixed amount of time or number of presentations and do not take individual differences in processing the stimuli into account. A comparison of the exposure time in our two experiments shows large differences: recall that the familiarization in Experiment 2 was fixed to 30 s of exposure to one of the tones. However, in the habituation experiment, infants needed about 6 presentation trials and accumulated an overall listening time to the tones of about 90 s before they reached the criterion, suggesting that they had more exposure to the crucial stimulus then the infants in Experiment 2. This difference may explain why the 6-month-olds discriminate the two tones after habituation, but not after familiarization: the amount of exposure may not have been sufficient for this age group to encode the stimulus in a way that allowed for its discrimination from another stimulus during the test phase. This also suggests that S6-month-olds may show discrimination after a longer familiarization (for effects of familiarization duration on infants' discrimination performance, see Bijeljac-Babic et al., 2012). The effect of trial number observed for Experiment 2 corroborates these considerations. Across the test phase, the listening times in 6-month-olds changed: while there was no evidence of discrimination in the first four trials, infants showed significantly different listening times to the two tones in the last trials.[5] This change over the experiment did not hold for the 9-month-olds,[6] which underlines that the discrimination performance by the 9-month-olds was not affected by the methodological modulation but that the effects of perceptual reorganization are rather robust in this age group.

However, it can also be the case that other reasons might explain the different findings in our two experiments: for example, the higher number of different trial types in the SAPP may have made infants' responses less sensitive across the conditions. The SAPP as used in our experiment and in the study by Yeung et al. (2013) included three trial types (one non-alternating containing the familiarized tone, one non-alternating containing the novel tone, and one alternating), whereas the studies that used habituation during the initial exposure phase only presented two different trial types, as we did in our Experiment 3 (habituated tone and novel tone: Chen et al., 2017; Chen &

---

[5] First four trials: Alternating vs. Non-Alternating Tone 25 ($\beta = 0.112$, $SE = 0.148$, $t = 0.760$, $p = 0.45$), Alternating vs. Non-Alternating Tone 33 ($\beta = 0.034$, $SE = 0.148$, $t = 0.233$, $p = 0.82$). Last four trials: Alternating vs. Non-Alternating Tone 25 ($\beta = 0.226$, $SE = 0.111$, $t = 2.017$, $p = 0.04$), Alternating vs. Non-Alternating Tone 33 ($\beta = 0.121$, $SE = 0.111$, $t = 1.084$, $p = 0.28$).

[6] First four trials: Alternating vs. Non-Alternating Tone 25 ($\beta = -0.020$, $SE = 0.148$, $t = -0.137$, $p = 0.89$), Alternating vs. Non-Alternating Tone 33 ($\beta = -0.183$, $SE = 0.148$, $t = -1.242$, $p = 0.22$). Last four trials: Alternating vs. Non-Alternating Tone 25 ($\beta = 8.5\text{e-}03$, $SE = 1.1\text{e-}01$, $t = 0.073$, $p = 0.94$), Alternating vs. Non-Alternating Tone 33 ($\beta = 9.8\text{e-}05$, $SE = 1.1\text{e-}01$, $t = 0.001$, $p = 0.99$).

Kager, 2016; Shi et al., 2017; or habituated tone and alternating: Ramachers et al., 2018), or only one trial type (the novel tone: Liu & Kager, 2014, 2017) during the test phase. Our two experiments with the infants also differed in another aspect of the experimental procedure. In Experiment 2, the duration of a head-turn to the presentation side of the acoustic stimulus was measured, while in Experiment 3 we measured visual fixation on a central monitor. We consider it unlikely that this methodological difference was responsible for the differential results across the two experiments, since listening times were the dependent variable in both cases. Moreover, head-turning versus visual fixation was not considered as a highly relevant factor in modulating test–retest reliability data in the analysis by Cristia et al. (2016).

However, the difference in the results of our experiments across the two testing conditions underlines the importance of the methodological decisions made for experiments with infants. To make research undertaken by different labs more comparable, a higher standardization of the methods used for specific research questions is desirable. We agree with Cristia et al. (2016) that this is specifically important for infant research as it is slow and costly, and therefore needs the close collaboration of researchers across institutions and languages.

Conclusions

Taken together, our findings suggest an age-related decline in the discrimination of lexical tones between 6 and 9 months with an additional perceptual recovery at the age of 18 months in German-learning infants. The perceptual recovery in toddlers might be driven by their acquisition of the native intonation and pragmatic system, whereas the discrimination at 6 months of age may be attributed to universal listening abilities. The decline in the ability to discriminate a non-native contrast was only evident when using habituation, but not when using familiarization, suggesting that methodological aspects are important to consider in the interpretation of findings from infant studies.

CHAPTER 6

Neural correlates of lexical tone and vowel processing in 6- and 9-month-old German-learning infants and German-speaking adults

Abstract

Previous behavioral experiments have shown that perceptual sensitivity to lexical tones declines between 6 and 9 months in infants learning non-tone languages (Mattock et al., 2008; Yeung et al., 2013; however, see Chen & Kager, 2016; Shi et al., 2017). Nevertheless, some studies have found a U-shaped developmental pattern characterized by a regaining of discrimination abilities at 18 months (Liu & Kager, 2014; Götz et al., 2018). The goal of the present study was to examine the neurophysiological underpinnings of this perceptual reorganization process by comparing the neural correlates of two different contrasts (a non-native lexical tone contrast and a native-like vowel contrast) in German-learning infants and adults. To this end, we conducted three ERP experiments with 6- and 9-month-old German-learning infants as well as German-speaking adults using a double oddball paradigm. Our results show that in the 6-month-olds, the tone and vowel contrasts elicited P-MMRs. In contrast, among the 9-month-olds the vowel contrast elicited an adult-like MMN while a P-MMR was observed for the tone contrast. In adults, both speech contrasts elicited MMNs, and we observed greater MMN amplitude for the vowel compared to the tone contrast. The P-MMRs for tones in infants and the MMN in adults indicate that both groups show neural sensitivity to this non-native contrast. Most interestingly, the decrease in behavioral discrimination for the same contrast found by Götz et al. (2018) between 6 and 9 months is not reflected in the neurophysiological responses. The emergence of the adult-like MMN for the native-like vowel contrast at 9 months corroborates findings of previous studies (e.g., Morr et al., 2002) and may be interpreted as response to language-specific phonological information to this native-like contrast, thus reflecting the perceptual reorganization process. Accordingly, the P-MMRs in the 6- and 9-month-olds provide no evidence for the perceptual reorganization for the non-native tone contrast.

6.1 Introduction

Adults' native language phonological system modulates their speech perception. This modulation affects the perception of speech segments like consonants (e.g., Best et al., 2001) and vowels (e.g., Ingram & Park, 1997) and of suprasegmental information like lexical stress (e.g., Cutler et al., 1997), lexical tone (So & Best, 2010), and the rhythmic organization of speech (e.g., Boll-Avetisyan et al., 2016). Native language phonological modulation emerges early in development: infants' first year of life is characterized by developmental changes that attune the perceptual system to the sound properties of the native language (for a review, see Werker & Gervain, 2013). A typical result of this attunement process is a decrease in the ability to discriminate sound contrasts that are not phonologically relevant in the native language while discrimination ability increases for sound contrasts that have phonological status in the native language (e.g., Kuhl et al., 2006). This change from universal to language-specific speech perception is referred to as the perceptual reorganization process.

Previous studies on the perceptual reorganization process for vowels and lexical tones

The facilitation of native vowel discrimination with increasing age has been verified in a meta-analysis. Tsuji and Cristia (2014) compared 22 different studies of infants' vowel discrimination. Their results confirm that between 6 and 10 months of age, the effect sizes for discrimination of native and non-native vowels begin to diverge with a positive slope for native vowel discrimination, which is typical of the perceptual reorganization process as described in the literature. However, a general decline of perceptual sensitivity for non-native vowels was not found in this meta-analysis or in other studies (de Klerk et al., 2019; Mazuka et al., 2014; Polka & Bohn, 1996, 2011).

Beyond vowel contrasts, previous studies addressing infants' ability to discriminate different (non-native) lexical tones also yielded mixed results. Specifically, a number of different developmental patterns were found. Some studies reported a decline in discrimination abilities with a loss of perceptual sensitivity at 9 months (Mattock & Burnham, 2006; Mattock et al., 2008; Yeung et al., 2013), which is in line with the perceptual reorganization process, while others found the opposite pattern, namely an increase in the perceptual sensitivity to lexical tone contrasts between 4 and

12 months (e.g., Chen et al., 2017; Chen & Kager, 2016). Furthermore, another set of studies found no change in perceptual sensitivity across different age groups (e.g., Ramachers et al., 2018; Shi et al., 2017). A small number of studies tested lexical tone discrimination beyond the first year of life. These studies demonstrated that 18-month-olds again show discrimination of lexical tones for which infants demonstrated a decrease in performance, thus confirming a U-shaped developmental pattern (e.g., for Cantonese mid-level vs. high-rising tones (Götz et al., 2018, Chapter 5) or an acoustically reduced high-level vs. falling Mandarin tone contrast (Liu & Kager, 2014). This U-shaped developmental pattern may be explained by the fact that both tone and non-tone languages use pitch on a phonological level; for example, pitch changes are used on the metric level and/or to transport paralinguistic information (Best, 2019; Chien et al., 2020). Hence, speakers of non-tone languages are not fully naïve to pitch but use pitch for different functions than speakers of tone languages do. Several behavioral studies have demonstrated that adult speakers without knowledge of a tone language can still discriminate lexical tones even though their performance is typically worse than that of native speakers (Burnham & Francis, 1997; Francis et al., 2008; Götz et al., 2018; Hallé et al., 2004; Hay et al., 2015; So & Best, 2010).

The current paper further investigates the perceptual reorganization process by comparing the neurophysiological underpinnings of German-learning infants' processing of a non-native contrast (Cantonese lexical tones) where behaviorally no discrimination has been found to a native-like contrast (Cantonese vowels) where we assume, based on general meta-analysis results, behaviorally maintained or enhanced discrimination with increasing age. Additionally, we compare infants' results to adults' neural processing of the same contrasts.

Mismatch negativity in adults

The mismatch negativity (MMN; Näätänen et al., 2007) is a widely studied neurophysiological event-related potential (ERP) of sound discrimination. The MMN is a fronto-central component that peaks approximately 100 to 250 ms after the point of divergence between a frequent standard and a rare deviant (either linguistic or non-linguistic) stimulus. The MMN is a pre-attentive component that occurs independent of participants' attentional state and that can also, for example, be evoked in sleeping participants (e.g., Nashida et al., 2000; Sallinen et al., 1996). Linguistic experience can modulate the MMN, resulting in higher amplitudes for native compared to non-native speech contrasts (e.g., Winkler et al., 1999). In addition, stimuli with greater acoustic distance elicit an MMN of greater amplitude than stimuli with less acoustic distance do (e.g., Näätänen, 2001). The amplitude of the MMN correlates with behavioral discrimination abilities, where contrasts that are easier to discriminate elicit a higher-amplitude MMN (e.g., Näätänen, 2001). Furthermore, an MMN can also be elicited in the absence of behavioral discrimination (e.g., Rivera-Gaxiola et al., 2005).

Neural correlates of lexical tone discrimination in adults

Several recent studies have used the MMN to compare tone processing in non-tone and tone language speakers (Chandrasekaran et al., 2007; Chen et al., 2018; Gao et al., 2019; Kaan et al., 2008; Politzer-Ahles et al., 2016; Yu et al., 2017). These studies commonly report that speakers of non-tone languages (in most of these studies, speakers of English) show an MMN in response to a non-native tone contrast. These results support the findings that speakers of non-tone languages can behaviorally discriminate tone contrasts (Burnham & Francis, 1997; Francis et al., 2008; Götz et al., 2018; Hallé et al., 2004; Hay et al., 2015; So & Best, 2010). However, with respect to language-specific effects on tone perception, the available results reveal a rather complex picture. Non-tone language speakers may not show MMNs to all tone contrasts (Gao et al., 2019; Kaan et al., 2008; Yu et al., 2017). For example, Kaan et al. (2008) found that native speakers of English showed an MMN to a Thai tone contrast consisting of a mid-level tone as standard and a low-falling tone as deviant, but not when a high-rising tone served as deviant. In contrast, native Thai speakers exhibited MMNs for both contrasts, although the MNN was weaker for the high-rising deviant. However, a study by Gao et al. (2019) testing native Mandarin and English speakers on their discrimination of sounds from a continuum between the Mandarin rising and falling tones demonstrated no effect of phonological knowledge. The authors tested contrasts within and across the Mandarin phonological categories and report an MMN for both language groups evoked by the within-category and across-category contrasts. The within-category contrast elicited greater amplitude in both groups. A possible explanation for the MMNs found in non-tone language speakers is that they might emerge over the course of an experiment as an effect of learning (Liu et al., 2018). Liu and colleagues (2018) tested English-speaking adults on an acoustically reduced Mandarin high-level (T1) versus high-falling (T4) contrast and found that this contrast elicited an MMN in the second half of the experiment but not in the first half, which suggests that non-tone language speakers need a certain exposure time to process a non-salient tone contrast.

Speakers of tone languages may also show modulations in their MMNs (amplitude or peak) to different native tone contrasts, and these modulations are not present in non-native speakers. Chandrasekaran et al. (2007) report that native speakers of Mandarin showed a stronger MMN (larger amplitude) to the high-level versus low falling-rising contrast than to the acoustically less similar high-rising versus low falling-rising contrast. Conversely, English speakers showed no differences in their MMNs to these two contrasts. In addition, the MMNs in native and non-tone language speakers may show differences in latencies: Chen et al. (2018) compared Mandarin and Dutch speakers' neural responses to the Mandarin high-rising versus low falling-rising contrast. These authors found that the native speakers' MMN showed a longer peak latency than that of the Dutch speakers. They assumed that Mandarin speakers perceive the stimuli categorically and thus may need to hear a larger portion of the stimuli to respond to their differences. Another finding is that in tone language speakers, the occurrence of an MMN is more strongly modulated by the manipulation of experimental conditions than in non-tone language speakers. Politzer-Ahles et al. (2016) tested MMN responses to the Mandarin high-rising, low falling-rising, and low tones in native Mandarin speakers and speakers of different non-tone languages. When the contrasts were represented by a single token per lexical tone, the authors observed no differences between the MMNs of native and non-tone language speakers. For both groups, an MMN was only elicited when the low falling-rising tone served as standard but not when it served as deviant, suggesting asymmetrical perception when the low falling-rising tone was involved in the contrast. However, when greater acoustic variability was introduced into the stimuli by using different tokens of the same tone contour and by using syllables with different vowels, the asymmetrical pattern disappeared for the non-tone language speakers but remained for the native speakers. This result pattern is in line with the assumption that the MMN response to lexical tones reflects the activation of phonological, long-term representations in native tone language speakers.

So far, the results from previous studies assessing lexical tone processing reveal a complex picture in which the emergence of an MMN in response to a tone contrast in speakers of tone languages is affected by the acoustic properties of the contrast tested (Chandrasekaran et al., 2007; Kaan et al., 2008) and the direction of the change from standard to deviant (Politzer-Ahles et al., 2016). In addition to lexical tones, tone languages also have segmental contrasts. In the next section, we describe the difference between the perception of lexical sounds and segmental language categories by native and non-native listeners in more detail.

Comparison between lexical tone and vowel perception

At present, little is known about how adults process lexical tones in comparison to segmental speech categories since only a few studies have compared processing between segmental and suprasegmental categories directly (e.g., Cutler & Chen, 1997; Ye & Connine, 1999). The authors found that on the behavioral level, segmental properties were detected faster than tone features were, demonstrating an overall advantage for segmental over tone features in language processing in adult speakers of tone languages. This finding has been confirmed in neural correlates: Gao et al. (2012) and Lidji et al. (2010) found that the MMN to consonants and vowels peaked earlier compared to the MMN to pitch differences. This finding is further supported by a study by Yu et al. (2017), who tested native Mandarin- and English-speaking adults' discrimination of Mandarin tone contrasts (low falling-rising tone as standard and high level and rising tones as deviants). The tones were articulated over different syllables (/gupa/, /gipa/, and /gypa/). The results revealed robust MMNs for native Mandarin speakers in response to both tone contrasts. In comparison, English native speakers showed a diminished MMN in the condition with longer interstimulus intervals. Furthermore, the comparison between lexical tones and vowels revealed a larger MMN in both language groups (i.e., native and non-native speakers) for vowel deviants compared to rising tone deviants. This direct comparison between vowels and tones shows that vowels may be processed more easily than tones.

Mismatch response in infants

In infants, the mismatch response (MMR) can differ from that observed in adults with respect to latencies and/or polarity. With respect to latency, infants' MMRs are slower and can occur in a different time window compared to the adult MMN. In infants, an MMR in an earlier time window could represent general discrimination abilities and could be seen as acoustic processing of sound changes that does not rely on other extrinsic factors (e.g., familiarity with specific sounds; e.g., Garcia-Sierra et al., 2016; Shafer et al., 2011). On the other hand, an MMR in a later time window could represent the precursor of the adult MMN (e.g., Marklund et al., 2019; Shafer et al., 2011). The polarity of the MMR in infants can be positive (P-MMR) or negative (adult-like MMN[7]), whereas it is known to be negative in adults. The characteristics of the shift from a positive polarity in infancy to a negative polarity in adulthood are unclear. The shift seems to be influenced by several factors, including the infant's maturational status (e.g., Friedrich et al., 2009; Leppänen et al., 2004) and the strength of the acoustic distance between the standards and deviants (e.g., Morr et al., 2002). Leppänen et al. (2004) have shown that older infants tend to show a more negative MMR. In other words, the P-MMR decreases with age, whereas the MMN becomes more prominent (e.g., Leppänen et al., 2004). Furthermore, a longitudinal study by Friedrich et al. (2009) demonstrated that the polarity of the MMR is linked to future language skills. Infants who showed a P-MMR to native stress patterns at the age of 4 to 5 months were found to have below-average language skills at 2.5 years of age, while infants who demonstrated an MMN at 4 to 5 months of age exhibited age-adequate language skills at 2.5 years. In addition to maturation, the acoustic properties of the stimuli seem to influence the MMR polarity. A greater acoustic distance between standards and deviants (and therefore better discriminability) elicits a more negative MMR in infants (e.g., Morr et al., 2002). Morr et al. (2002) presented 2- to 7-month-old infants with different degrees of acoustic distance between standard and deviant stimuli. They found that infants showed a P-MMR for smaller acoustic differences (1,000 Hz vs. 1,200 Hz), whereas larger acoustic differences (1,000 Hz vs. 2,000 Hz) elicited an adult-like MMN.

---

[7] The term "adult-like MMN" refers to the MMN in infants and the evidence that the MMN for infants may appear in a different time windows than the MMN for adults and are therefore not identical.

Independent of its polarity, the MMR amplitude is related to neural discrimination abilities. The discrimination status is especially relevant for studies that investigate neural correlates of the perceptual reorganization process during the first year of life. However, it can be challenging to examine neural correlates of the perceptual reorganization process due to the change of a P-MMR to an adult-like MMN across development. Since the positive and negative polarities cancel each other out, no MMR might be observed on the group level. For example, Rivera-Gaxiola et al. (2005) only found neural evidence of non-native sound discrimination in 11-month-olds when dividing their participant group into positive and negative mismatch responders. In contrast, a native sound contrast elicited an adult-like MMN in all infants. In this context, the polarity of the MMRs might reflect the perceptual reorganization process. Native contrasts elicit an adult-like MMN earlier in the developmental trajectory compared to non-native contrasts. The P-MMR remains present longer for those non-native contrasts.

Neural correlates of vowel and lexical tone perception in infants

The pattern of an initial P-MMR and a later adult-like MMN has also been observed in studies investigating infants’ neural response to native vowel contrasts (e.g., Marklund et al., 2019; Shafer et al., 2011; Yu et al., 2019), where the P-MMR was found to decrease with age, while the adult-like MMN increased with age. These findings further suggest that the polarity change of the MMR reflects the perceptual reorganization process. However, other studies have shown that the adult-like MMN is already present in young infants and decreased for non-native vowel contrasts, while it remained stable or increased for native vowel contrasts (e.g., Cheour et al., 1998; Jansson-Verkasalo et al., 2010).

Only a few studies have investigated neural processing of lexical tones in infants (Cheng et al., 2013, 2015; Cheng & Lee, 2018; Liu et al., 2014). Even fewer studies have investigated the neural underpinnings of lexical tone processing in infants learning a non-tone language (Liu et al., 2019). In a longitudinal study, Cheng et al. (2013) tested Mandarin-learning infants at birth and at 6 months of age on their neural response to two Mandarin tone contrasts: a contrast with a large acoustic distance between the tones (high-level, T1 vs. low-dipping, T3), which the researchers call the distinct contrast, and a contrast with a small acoustic distance (high-rising, T2 vs. low-dipping, T3), which they call the similar contrast. The MMR to the distinct tone contrast changed from a P-MMR at birth to an adult-like MMN at 6 months. In contrast, the similar tone contrast elicited neither a P-MMR nor an adult-like MMN at birth but a P-MMR at 6 months. In a subsequent study, Cheng and Lee (2018) used the same tone contrasts in another paradigm to investigate the developmental trajectory of neural responses to tone contrasts beyond the first year of life (at 12, 18, and 24 months). The distinct (T1 vs. T3) tone contrast elicited a robust, adult-like MMN in all age groups, thus corroborating the results regarding the 6-month-olds. In line with these results, a similar contrast (T2 vs. T3) resulted in a P-MMR at 12 and 18 months. However, this contrast did not elicit an MMR in the 24-month-olds. These results again demonstrate the influence of stimulus properties on the polarity of the MMR in infants. The acoustically more similar contrast showed a P-MMR, compared to the acoustically more distinct tone contrast. The null findings in the 24-month-olds may be due to the co-occurrence of P-MMR and adult-like MMN among participants, resulting in responses canceling each other out.

Very few studies have investigated the neural response to tones in infants learning a non-tone language (Liu et al., 2019). As previously mentioned, infants learning a non-tone language show diverse patterns of discrimination abilities for lexical tone contrasts on the behavioral level. Given these divergent results, ERPs can help answer the question of whether a general decrease in perceptual sensitivity is reflected in the neural correlates or a product of a reorganization process with maintained neural discriminability of non-native tone contrasts. Liu et al. (2019) investigated neural responses to Mandarin tone contrasts in bilingual (English with other languages) and monolingual English-learning infants who were 5 to 6 and 11 to 12 months old. The authors used the same acoustically manipulated and contracted Mandarin contrast of a high-level versus high-falling tone for which they had found a U-shaped developmental pattern in a preceding study (Liu & Kager, 2014). Their results revealed that bilingual infants in both age groups showed a P-MMR. However, a P-MMR was only elicited in the 5- to 6-month-old monolingual infants and not in the 11- to 12-month-old monolingual infants. The absence of a P-MMR in the 11- to 12-month-old monolingual infants might again be attributable to the mutual canceling out of P-MMRs and adult-like MMNs at the group level.

As previously mentioned, the findings regarding lexical tone processing in adults are diverse. Hence, if we want to draw clear conclusions about infants' perceptual reorganization process for lexical tones, it is essential to understand how adults process the contrasts used in this study. Similarly, if we compare the developmental trajectory of infants' responses to lexical tones and vowels, we must ask to what extent adults' neural responses to lexical tones differ from their responses to vowels.

The current study

As previously discussed, a few studies have addressed the question of how infants learning a non-tone language perceive lexical tones within the first year of life. Results from behavioral studies have shown different developmental trajectories including a decline, maintenance, and enhancement of perceptual sensitivity as well as a U-shaped developmental pattern. However, it remains an open question whether the initial decline in perceptual sensitivity between 6 and 9 months is also reflected in the neural correlates. To expand the behavioral findings, the present study used ERPs to systematically investigate the neural underpinnings of the perceptual reorganization process in infants learning a non-tone language. For this purpose, we used a double oddball paradigm to test German-learning 6- and 9-month-old infants' and German monolingual adults' processing of a vowel contrast (Cantonese /ɛ/ vs. /i/) and a tone contrast (Cantonese mid-level, T33 vs. high-rising, T25). A previous study demonstrated that German adults behaviorally discriminated the tone pair used in this study, although at a lower performance level than native speakers of Cantonese (Chapter 5, Götz et al., 2018). Additionally, the previous study demonstrated that German infants behaviorally discriminated the tone contrast at 6 and 18 months but not at 9 months.

Since both the /ɛ/ and /i/ vowels are part of the German vowel inventory, we expected assimilation to the German categories (Best, 1995). For this reason, we expected phonological processing even if a native speaker of Cantonese articulated the vowels. Based on the aforementioned meta-analysis showing an overall facilitation effect for native vowel discrimination (Tsuji & Cristia, 2014), we also expected to find facilitation in discrimination in German-learning infants for this specific contrast. The two contrasts selected for this study allowed us to further investigate the perceptual reorganization process by comparing a non-native and a native-like contrast. The aim of the present study is to investigate the neural underpinnings of the developmental trajectory of the perceptual reorganization process for a contrast where behaviorally no discrimination has been found compared to the discrimination of a native-like contrast where we assume maintained or enhanced discrimination ability with increasing age.[8]

To pursue this goal, we conducted two experiments. Experiment 1 investigated the neural underpinnings of the perceptual reorganization process in German-learning 6- and 9-month-old infants. With this experiment, we aimed to (1) further analyze the developmental trajectory in the ability to discriminate non-native lexical tone contrasts and (2) compare the ERP elicited by a non-native tone contrast to the ERP elicited by a native-like vowel contrast, where we expected the perception to be maintained or facilitated with increasing age. In Experiment 2, we explored adults' neural responses to the same lexical tone and vowel contrasts to verify the presence of an MMN given the diverging findings on lexical tone processing in non-tone language speakers. The verification of adults' neural responses allows us to further assess whether any potential differences in infants' neural responses to the speech contrasts are a result of the perceptual reorganization process or whether these differences can be attributed to the inherent properties of the stimuli.

---

[8] This assumption is driven by the findings from the meta-analysis by Tsuji & Cristia (2014), who found a positive slope for discriminating native vowel contrasts between 6 and 10 months.

6.2 Experiment 1: Neural correlates of lexical tone and vowel perception in 6-, and 9-month-old infants

This experiment investigated the neural correlates of the perceptual reorganization process in German-learning 6- and 9-month-old infants. With this experiment, we aimed to further investigate the developmental trajectory of the lexical tones discrimination and compare the ERPs elicited by a non-native tone contrast to the ERPs elicited by a native-like vowel contrast.

6.2.1 Method

6.2.1.1 Participants

In total, data from 50 monolingual German-learning infants were analyzed in this study: 25 9-month-olds (M_{age} = 274, *range* = 262 to 294 days, 11 females) and 25 6-month-olds (M_{age} = 185 days, *range* = 165 to 210 days, 12 females). Infants that participated in the study at 6 months, did not participate in the same study at 9 months. An additional 36 infants were tested but excluded from further analysis due to one of the following reasons: less than 35 artifact-free trials (14 6-month-olds: 10 female, 20 9-month-olds: 10 female), no toleration of the cap by the infant (one 6-month-old, one 9-month-old).

All infants were born full-term. According to parental report, infants did not suffer from repeated or acute ear infections, showed no indications of atypical development, and did not have any regular exposure to a tone language. This study was approved by the Ethics Committee of the University of Potsdam. Parents gave written informed consent in accordance with the Declaration of Helsinki.

6.2.1.2 Stimuli

The syllables [sɛː] and [siː] served as stimuli. Stimuli were meaningful CV syllables in Cantonese but nonsense words in German. Several exemplars of these syllables were recorded with either the high-rising (T25) or the mid-level (T33) tone by a female speaker of Cantonese in a sound-attenuated booth. All recordings were digitalized with a sampling rate of 44.1 kHz. For the ERP experiment, two different tokens of each syllable were selected. Acoustic measures on duration, pitch, and formant frequencies were conducted for the two tokens of each stimulus (see Table 9 for the results). The pitch contours of the stimuli are displayed in Figure 6. All stimuli were normalized in intensity and presented via loudspeaker at 60 dB SLP. In the vowel condition, the vowel changed between standard and deviant from [ɛ:] to [i:] with a consistent tone (T33). In the tone condition, the vowel remained constant ([ɛ:]) between deviant and standard, but the tone changed from T33 to T25. The syllable /sɛ33/ ([sɛ:] with a T33 tone contour) was always presented as standard, the syllable [sɛ:25] as tone deviant and [si:33] as vowel deviant.

Token	Vowel Duration	Syllable Duration (ms)	F1 (Hz)	F2 (Hz)	F3 (Hz)	F0 initial (min-max) (in Hz)	F0 middle (min-max) (in Hz)	F0 final (min-max) (in Hz)
sɛ25_1	461	586	726	1214	2485	208 (199-220)	196 (193-199)	230 (199-266)
sɛ25_2	454	579	728	1209	2479	211 (200-222)	197 (194-202)	237 (201-268)
si33_1	456	584	351	1755	2891	217 (207-227)	193 (188-208)	187 (183-190)
si33_2	451	578	374	1707	2835	202 (194-215)	192 (188-198)	190 (187-197)
sɛ33_1	458	583	688	1295	2478	203 (195-218)	192 (188-195)	192 (188-196)
sɛ33_2	457	584	703	1385	2673	211 (202-222)	196 (191-202)	188 (186-192)

Table 9. Results from the acoustic analysis of the different vowel and tone stimuli used in the present study. F0 was measured at three different time points in the vowel: at the beginning of the vowel (initial), at 50 % of the vowel (middle), and the end (final) position of the vowel. [Sɛ:25] served as standard in the infant study, token of [si:25] were used as vowel deviants, and tokens of [sɛ:33] were used as tone deviants.

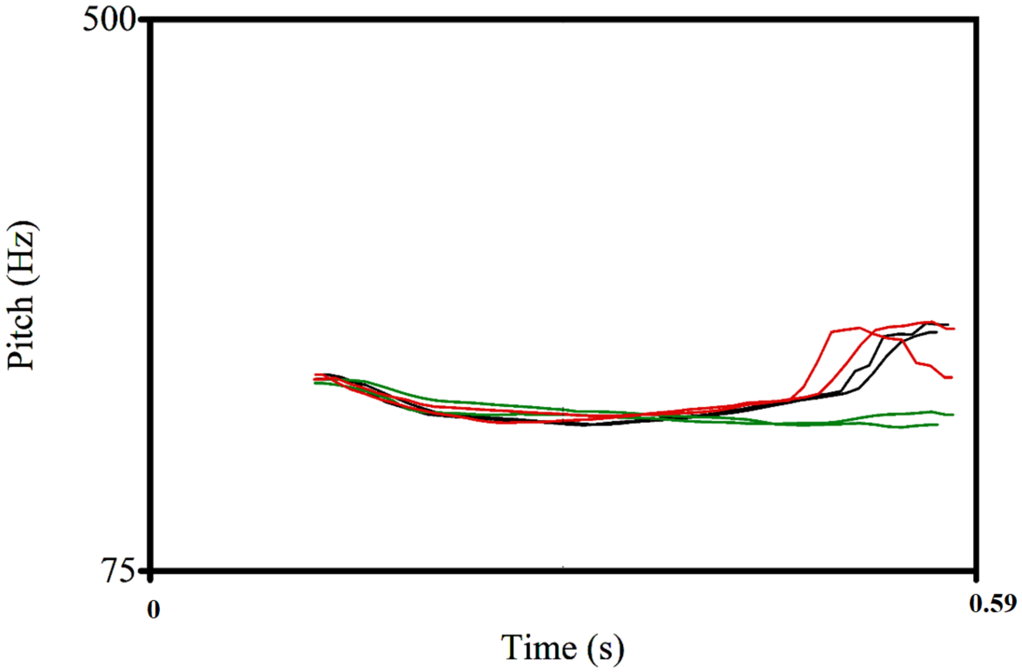

Figure 6. F0 contours of the three different syllables [sɛ:25] (black), [si:25] (red), and [sɛ:33] (green).

6.2.1.3 Procedure

Before the experiment started, caretakers were informed about the procedure and signed or handed in the signed consent form. Infants were seated on their caretakers' lap approximately 1 m away from a computer screen. The acoustic stimuli were presented by a loudspeaker positioned next to the screen. During stimulus presentation, infants watched an infant friendly movie (with muted sound) on the screen and/ or a second experimenter engaged the child with silent toys. All acoustic stimuli were presented with Presentation Software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) with stimulus durations ranging between 578 ms – 586 ms and interstimulus intervals (ISI) ranging between 800 – 900ms. The varying ISI prevents participants from building rhythmic patterns over the stimulus presentation (e.g., Kirmse, et al., 2008). Stimuli were presented in a double passive oddball paradigm. With the double oddball paradigm, it was possible to compare the MMN responses for a vowel (as native-like) and a tone (as non-native) speech contrast. Infants listened to 800 stimuli, of which 640 were standards, 80 were tone deviants and 80 were vowel deviants. The stimulus presentation was divided into four blocks. Each block started with the presentation of eight standards. Deviants were presented in a pseudo-random fashion: 3-8 standards were presented between two deviants. The first eight standards as well as standards directly following deviants were excluded from further analysis. The experiment was terminated if the infant became fussy and could not be calmed by the caretaker or if the maximum presentation time (20 min) was reached. A break was inserted automatically after each block (after approximately 5 minutes) or manually when the caretaker/ infant indicated the need for an additional break. Infants were excluded from the analysis if they contributed fewer than 35 artifact-free trials per condition. The mean trial number for the 6-month-old infants were 218 artifact-free trials for standards, 36 for the vowel deviant, 37 for tone deviants. The 9-month-olds had an average of 281 artifact-free trials for standards, 49 for vowel deviants, and 49 for tone deviants.

6.2.1.4 ERP recording and analysis

The electroencephalogram (EEG) was continuously recorded by 32 cap-mounted active Ag/AgCl electrodes (BrainProducts, Gilching, Germany) at a sampling rate of 1000 Hz. Electrodes (F3, F7, F9, F4, F8, F10, FC1, FC5, C3, FC2, FC6, C4, CP1, CP5, P3, P7, CP2, CP6, P4, P8, FCz, Fz, Cz, CPz, Pz, O1, O2) were positioned following the 10–20 system convention. The electrooculogram (EOG) was recorded from electrodes placed above the right and left eye. The ground electrode was placed at AFz position. Impedances were kept below 25 kΩ. The EEG data were analyzed using Brain Vision Analyzer (version 2.01; Brain Products, Gilching, Germany). The signal was filtered offline with a 0.1-30 Hz bandpass filter. Data were segmented in epochs of 1000 ms, starting 100 ms prior to the onset of the stimuli. The EEG recording was referenced online to the left mastoid and then re-referenced offline to the linked mastoid electrode, and baseline-corrected 100 ms before stimulus onset. Eye blinks and eye movements in the segments were corrected by a computer algorithm (Gratton, et al., 1983). All other artifacts were detected automatically (exceeding a range of ± 100 µV) and were excluded from further analysis. In infant research, the timing of the MMR is considered to be more flexible, ranging from 100 ms to 400 ms after the point of divergence (e.g., Marklund et al., 2019, Yu et al., 2019). In line with previous studies, we analyzed the data within two a priori selected time windows: an early window from 100-300 ms, and a later time window from 300-500 ms after the point of acoustic divergence of the stimuli[9] (e.g., Marklund et al., 2019; Yu et al., 2019). For the vowel contrast, the early time window ranged from 200 to 400 ms and the late time window from 400 to 600 ms after stimulus onset. For the tone deviant, the early time window ranged from 400 to 600 ms, and 600 to 800 ms for the late time window after stimulus onset. The different time windows were selected because the point of divergence is different for vowels (100 ms after onset) and tones (300 ms after onset), see Table 9 and Figure 6 in the stimuli section. Electrodes were clustered into three different regions: left (F4, F7, F9, FC1, C3, FC5, CP1, CP5, P3, P7), right (F4, F8, F10, FC2, C4, FC6, CP2, CP6, P4, P8), and central (FCz, Fz, Cz, CPz, Pz).

6.2.2 Results

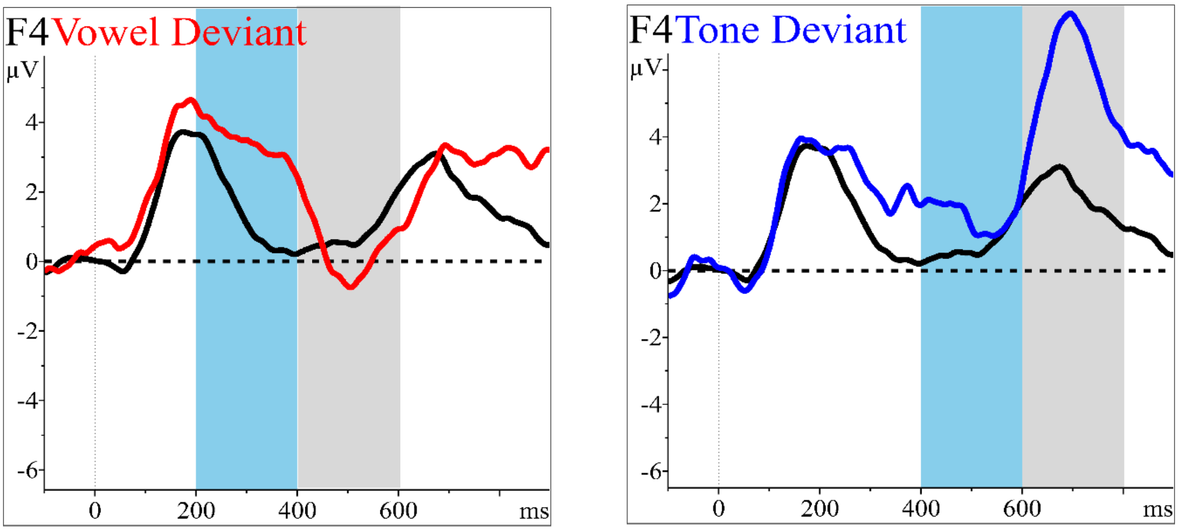

Figure 7. Grand-average of the MMRs to vowels (red line) and tones (blue line) in comparison to the standard (black line) in 6-month-olds for the F4 electrode. The blue bar represents the early time window; the grey bar represents the late time window.

---

[9] Note that the length of the initial consonant /s/ is approximately 125 ms, nevertheless we used 100 ms as starting point for the time window since formant transitions might already introduce the following vowel due to coarticulation.

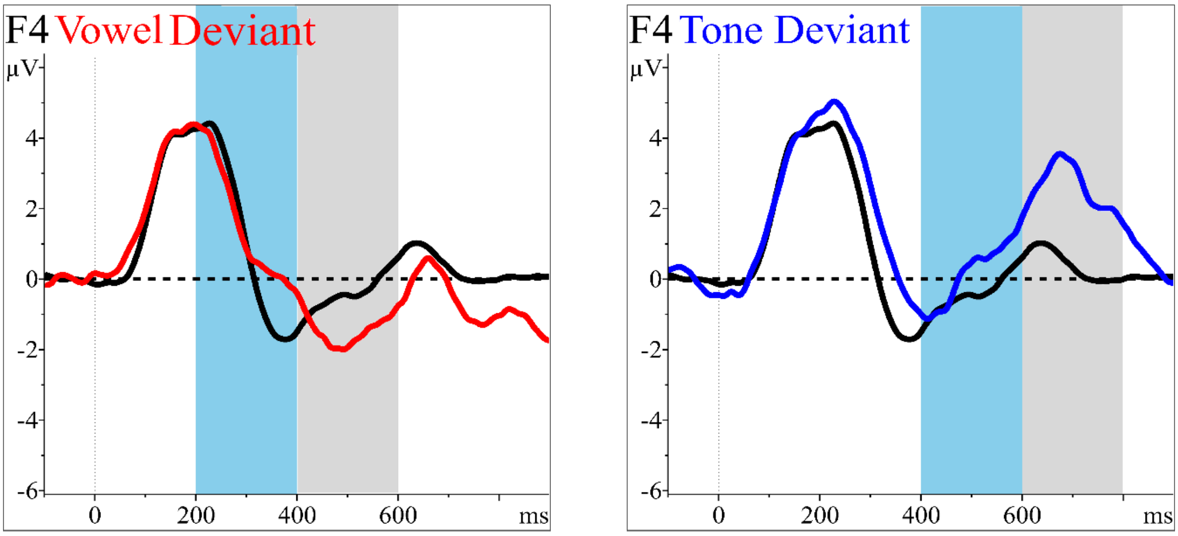

Figure 8. Grand-average of the MMRs to vowels (red line) and tones (blue line) in comparison to the standard (black line) in 9-month-olds for the F4 electrode. The blue bar represents the early time window; the grey bar represents the late time window.

For illustration purposes, Figures 7 and 8 depict the grand-averages for standards, vowel contrasts, and tone contrasts obtained from the F4 electrode at 6 and 9 months. All electrodes were included in the analysis. We calculated the MMR by the difference between the amplitude of the standard and vowel deviant, as well as between the standard and tone deviant. For statistical analysis, we used the values of the MMR amplitude as the dependent variable. As a first step, we compared the MMR amplitude in both conditions averaged across all electrodes in order to test whether the two deviant types elicited an MMR. We analyzed the difference between amplitudes over the early and the late time windows separately and calculated the statistics against zero. The MMR amplitude for the 6-month-olds (Figure 9) differed significantly from zero for the vowel contrast at the early but not at the late time window (early time window: (*t* (62) = 6.924, *p* < 0.001, *amplitude* = 1.425 µV; late time window: (*t* (62) = -1.0964, *p* = 0.2733, *amplitude* = -0.2967 µV). The MMR amplitude for the tone contrast was not significantly different from zero at the early time window but differed significantly at the late time window (early time window: (*t* (62) = -0.8044, *p* = 0.4215, *amplitude* = -0.204 µV; late time window: (*t* (62) = 2.7272, *p* = 0.007, *amplitude* = 0.8219 µV).

The analysis for the 9-month-olds revealed that the MMR amplitude for the vowel contrast did not differ from zero at the early time window (*t* (62) = 0.24201, *p* = 0.8089, *amplitude* = 0.0411 µV), but differed significantly from zero at the late time window (*t* (62) = -5.7643, *p* < 0.001, *amplitude* = -1.104 µV). For the tone contrast the MMR amplitude at both time windows differed significantly from zero (early time window: (*t* (62) = 3.383, *p* = < 0.001, *amplitude* = 0.6044 µV; late time window: (*t* (62) = 3.579, *p* < 0.001, *amplitude* = 0.7536 µV). Figure 9 displays the MMR amplitude for the vowel and tone conditions at both ages (6 and 9 months), as well as the corresponding time windows.

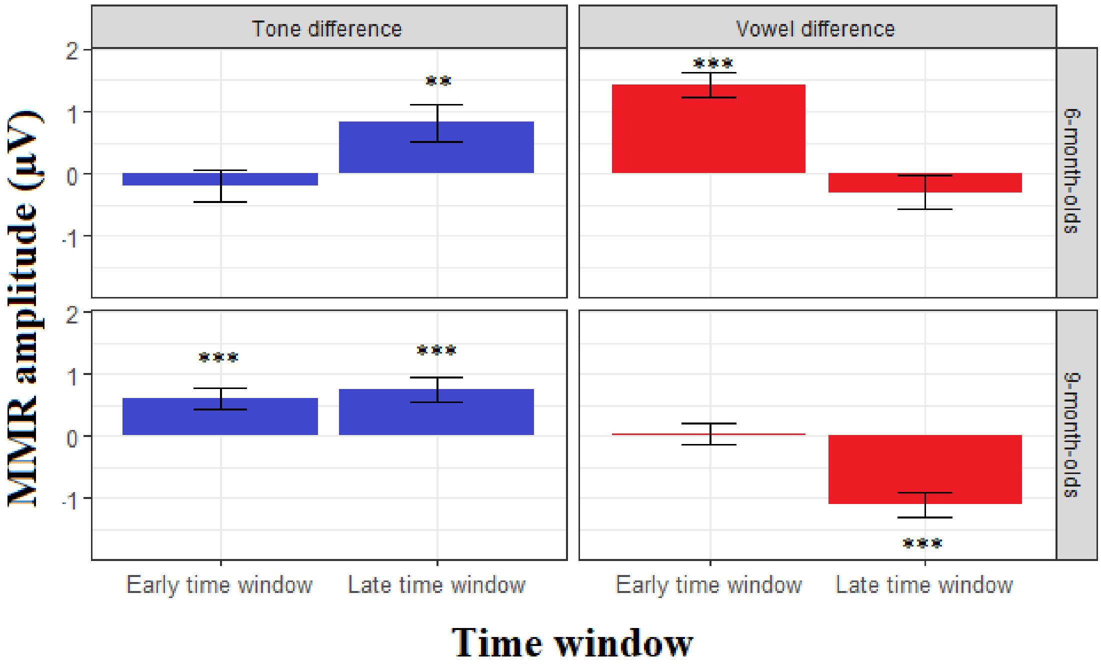

Figure 9. Barplots of the MMR amplitudes averaged across all electrodes of the lexical tone (blue) and vowel contrast (red) for the early (on the left side) and late time windows (on the right side) and the different age groups: 6-month-olds (upper side) and 9-month-olds (bottom side). Asterisks mark the significance from zero: ** p < .01, and *** p < .001, whiskers represent ± 1 standard error of the mean.

Subsequently, we computed different mixed models to test which model fits the data best and to exploratively analyze the relevance of different brain regions (left, central and right regions) and time windows (early and late time windows) for the MMR. The models varied systematically in the added factors (starting with deviant type and then adding brain region, and their interactions as fixed effects) and were compared against each other. The final model included the main effect of deviant type (tone vs. vowels), age (6- and 9-months), time window (early and late), region (left, right and central brain regions), as well as the interaction of all factors as fixed effects. Subject was included as random effect. Table 10 displays the results from the model fitting process.

Model		AIC	BIC	Chisq	Df	p (> Chisq)
1.	~ condition + (1 | subject)	31330	31356			
2.	~ condition + (condition*age) + (1 | subject)	31311	31350	22.94	2	< 0.001 ***
3.	~ condition + (condition*age) + (condition*age * time_window) + (1 | subject)	31263	31328	56.117	4	< 0.001 ***
4.	~ condition + (condition*age) + (condition*age*time_window) + condition*age* time_window * region) + (1 | subject)	31232	31401	62.81	16	< 0.001 ***

Table 10. Summary of the results from the model comparisons. Models were compared hierarchically: the first model was compared to the second model, the second model to the third, and so forth. Triple asterisks (***) indicate p < 0.001. The results indicate best fit to the data for the model, including the interaction of condition (vowel, tone), age (6, 9 months), time window (early, late), and region (left, right, central brain regions).

We compared models separated for each age group to resolve the interaction. The same models, as in the global fitting, were applied to the data from the 6- and 9-month-olds.[10] For the post-hoc analysis, we were only interested in the differences between regions and time windows for tones and vowels separately.

Effect of time window and region on the MMR in 6-month-olds

For the 6-month-olds, post-hoc analysis (with Tukey test with adjusted p-values for multiple comparisons) revealed that the processing of vowels and tones differed only at the late time window in the right region. Tones elicited a more positive MMR on the left and right region compared to central regions at the late time window, see Figure 10. All other comparisons revealed non-significant results. Table 11 displays detailed information about the resolved interaction of condition, region, and time window.

---

[10] The model MMN amplitude ~ condition + (condition * time_window) + (condition * time_window * region) + (1 | subject) was applied to the data for each of the age group (6, and 9 months)

Region	Early time window				Late time window			
Tones								
	β (SE)	df	t-value	p-value	β (SE)	df	t-value	p-value
left - right	-0.172 (0.556)	2486	-0.31	1.00	0.326 (0.556)	2486	0.587	1.00
central - left	-1.302 (0.680)	2486	-1.914	0.907	-3.477 (0.680)	2486	-5.110	< 0.001
central - right	-1.474 (0.680)	2486	-2.167	0.776	-3.151 (0.680)	2486	-4.631	< 0.001
Vowels								
Region	Early time window				Late time window			
	β (SE)	df	t-value	p-value	β (SE)	df	t-value	p-value
left - right	0.687 (0.556)	2486	1.237	0.99	1.268 (0.556)	2486	2.283	0.696
central - left	-0.913 (0.680)	2486	-1.342	0.997	-1.227 (0.68)	2486	-1.804	0.99
central - right	-0.226 (0.680)	2486	-0.332	1.00	0.041 (0.68)	2486	0.06	1.00

Table 11. Results from the resolved interaction of condition (vowel and lexical tones), time window (early and late) and region (left, right, central) at 6 months.

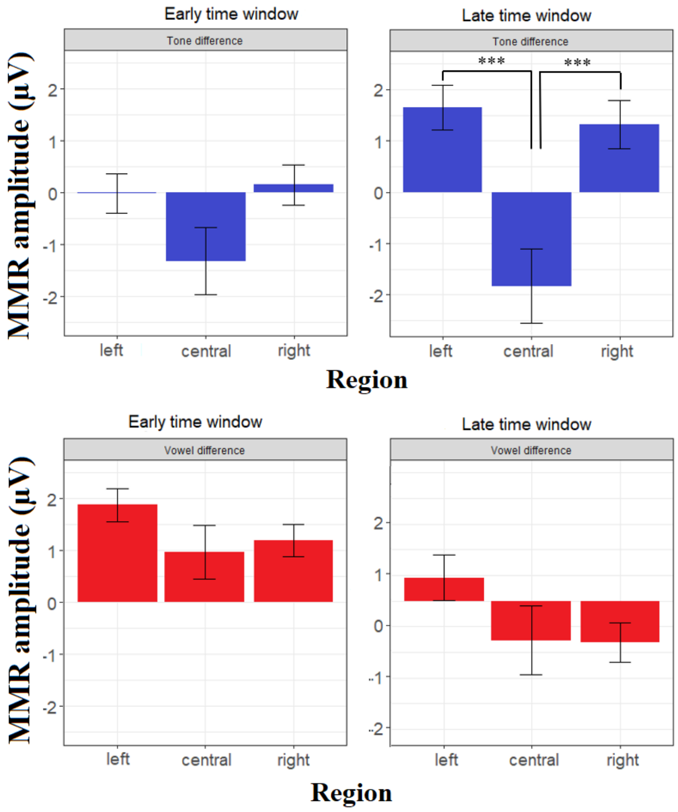

Figure 10. MMR amplitude in the different regions (left, central, right) and early and late time windows in the 6-month-olds for tones (blue) and vowels (red). Whiskers represent
± 1 standard error of the mean.

Effect of time window and region on the MMRs in 9-month-olds

For the 9-month-olds, we again applied the same model with condition (tone and vowels), region (left, right, central), and time window (early and late). The post-hoc analysis (Tukey test with adjusted p-values against multiple comparisons) revealed no statistically significant differences in processing between the different regions for lexical tones or vowels at 9 months, see Table 12 and Figure 11.

	Tones							
Region	Early time window				Late time window			
	β (SE)	df	t-value	p-value	β (SE)	df	t-value	p-value
left - right	0.505 (0.401)	2486	1.258	0.99	1.085 (0.401)	2486	2.702	0.379
central - left	0.102 (0.492)	2486	0.208	1.00	-0.312 (0.492)	2486	-0.635	1.00
central - right	0.607 (0.492)	2486	1.235	0.99	0.031 (0.492)	2486	0.063	1.00
	Vowels							
Region	Early time window				Late time window			
	β (SE)	df	t-value	p-value	β (SE)	df	t-value	p-value
left - right	-0.080 (0.401)	2486	-0.200	1.00	0.774 (0.401)	2486	1.928	0.901
central - left	-0.150 (0.492)	2486	-0.306	1.00	-0.574 (0.492)	2486	-1.169	0.99
central - right	-0.230 (0.492)	2486	-0.469	1.00	0.199 (0.492)	2486	0.405	1.00

Table 12. Results from the resolved interaction of condition (vowel and lexical tones), time window (early and late) and region (left, right, central) at 9 months.

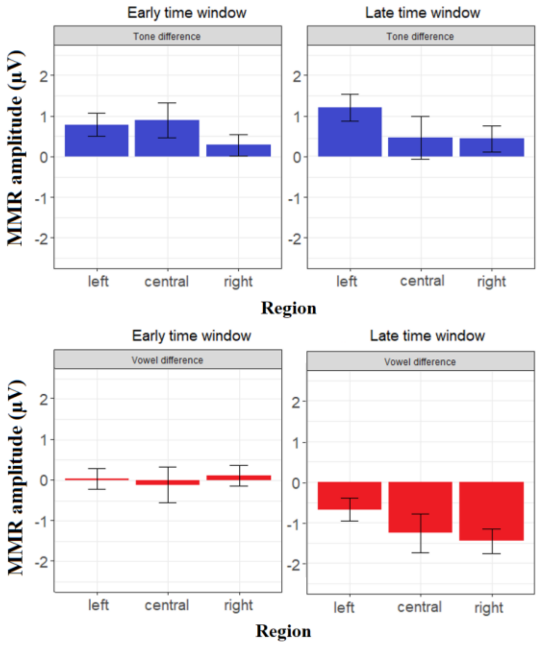

Figure 11. MMR amplitude in the different regions (left, central, right) and early and late time windows in the 6-month-olds for tones (blue) and vowels (red). Whiskers represent ± 1 standard error of the mean.

To summarize, the results showed that lexical tones elicited only P-MMRs in 6- and 9-month-olds, whereas for the vowel contrast we observed a P-MMR in 6-month-olds and an MMN in 9-month-olds. With respect to differences between regions, our results revealed that at 6 months, the lexical tone contrast elicited a more positive MMR in the left and right region compared to the central region during the late time window. At 9 months, the difference in the regions disappeared.

6.2.3 Discussion

This experiment was intended to investigate the neural correlates of the perceptual reorganization process of vowels and lexical tones in 6- and 9-month-old infants learning a non-tone language. Based on the assumptions of perceptual reorganization, we expected a decrease in the perceptual sensitivity to the (non-native) lexical tone contrast concurrent with a facilitation or maintenance effect for the (native-like) vowel contrast with increasing age. Our study revealed two main findings. First, the lexical tone contrast elicited only P-MMRs across the tested age groups, while the vowel contrast elicited a P-MMR in 6-month-olds and an adult-like MMN in 9-month-olds. The second finding relates to the differences between regions. At 6 months, tones elicited a more positive MMR in the left and right regions compared to the central region during the late time window, whereas the difference in regions disappeared at 9 months.

Differences between mismatch responses in the early and late time windows

The underlying processing mechanisms of the infants’ MMRs are not conclusively clarified (e.g., Bishop et al., 2011; Kushnerenko et al., 2013; Shafer et al., 2011); therefore, we analyzed two different time windows in this experiment – an early and a late response to the two tested contrasts – which is in line with previous studies (e.g., Garcia-Sierra et al., 2016; Marklund et al., 2019; Yu et al., 2019). The MMR occurring in the early time window may reflect the acoustic processing of sounds, while the MMR in the later time window can be seen as the precursor of the adult MMN (e.g., Ferjan Ramirez, 2017; Garcia-Sierra et al., 2016; Marklund et al., 2019; Shafer et al., 2011; Yu et al., 2019). For the neural processing of native vowel categories, Yu et al. (2019) reported a P-MMR in the early time window in 3- to 47-month-olds, and a P-MMR in the late time window in 3- to 12-month-old monolingual English-learning infants, and an MMN in the 13- to 47-month-olds for the English American vowel /ɛ/ versus /ɪ/ (Yu et al., 2019), whereas Marklund et al. (2019) reported only a P-MMR in the early time window but no P-MMR or MMN in the late time window in 4- to 8-month-old Swedish-learning infants for the Swedish vowels /e/ versus /i/ (Marklund et al., 2019). Marklund et al. (2019) explain the absence of an MMN in the late time window by suggesting that exposure to speech impacts the amplitude of the MMR in the late time window, leading to a shift from positive to negative amplitudes with high amount of language exposure, see also Garcia-Sierra et al. (2016). This switch in infancy might not follow the same temporal course in each infant during the developmental trajectory. Therefore, some infants might have already shown an MMN, whereas in other infants the contrast still elicited the P-MMR (e.g., Rivera-Gaxiola et al. 2005). In the current study, we might observe a similar pattern. The vowel contrast /ɛ/ versus /i/ elicited only a P-MMR in the early time window at 6 months and only an MMN in the late time window at 9 months. In line with the suggestion by Garcia-Sierra et al. (2016) and Marklund et al. (2019), the MMN at 9 months might reflect the amount of language exposure, whereas the absence of an MMR or MMN in 6-month-olds might indicate the heterogeneous transition from positive to negative amplitude between infants. However, we did not observe the same pattern for the tone contrast. Tones elicited a P-MMR in the late time window at 6 months and in both the early and late time windows at 9 months. This result may indicate that infants have less experience with tones, which might be reflected in the presence of the P-MMRs in 6- and 9-month-olds.

In 6-month-old infants, the MMR in the early time window to vowels as opposed to tones may indicate that infants are faster in encoding vowels compared to tones (e.g., Liu et al., 2014). The emergence of an additional P-MMR to tone contrast in the early time window at 9 months could indicate that processing becomes faster at 9 months. This faster processing may be due to facilitated acoustic processing of sounds at 9 months compared to 6 months. In contrast, the MMN in the late time window, which emerged for the vowel contrast at 9 months, might be a precursor of the adult MMN elicited in a later time window in infants (e.g., Shafer et al., 2011). However, the emergence of the MMN in a later time window might suggest that the infant brain needs more processing time compared to the adult brain.

Since both MMRs (P-MMR and MMN) indicate neural discrimination ability, we discuss the two time windows in conjunction to better understand to what extent the ERP results represent the perceptual reorganization process of vowels and tones. In the next section, we first separately discuss our lexical tone and vowel perception results in light of the perceptual reorganization process. Finally, we discuss the comparison between lexical tones and vowels.

Neural correlates of the perceptual reorganization process of lexical tones in 6- and 9-month-olds

The 6-month-olds in the current study showed a neural response to the non-native lexical tone contrast in the form of a P-MMR. This result is in line with the development of acoustic discrimination abilities (e.g., Garcia-Sierra et al., 2016). Other studies investigating infants' neural response to non-native and native tone contrasts have observed the initial neural discrimination demonstrated by a P-MMR (e.g., Cheng et al., 2013; Cheng & Lee, 2018; Liu et al., 2019). Hence, we interpret the P-MMR in the 6-month-olds as a response to the acoustic difference between the Cantonese high-rising and mid-level tones. However, our results regarding the 9-month-olds (P-MMRs in the early and late time windows) contrast with the findings by Liu et al. (2019) who tested infants learning a non-tone language. Liu et al. (2019) found that the Mandarin lexical tone contrast (T1 vs. T4) did not elicit an MMR in monolingual English-learning infants between 11 and 12 months of age. However, German-learning infants in our study exhibited a P-MMR at 9 months. A possible explanation for this discrepancy derives from the acoustic differences between the tone contrasts used in Liu et al.'s (2019) study and our study. The more salient a contrast is, the better infants (and adults) can detect it. More acoustically distinct contrasts elicit greater MMRs compared to more acoustically similar contrasts (e.g., Cheng et al., 2015; Cheng & Lee, 2018). The contrast Liu et al. (2019) used consisted of an acoustically reduced pitch contour of the Mandarin T1 versus T4 tone contrast with a small perceptual distance, while the tone contrast we used was a natural one without an artificially reduced pitch. The contrast we used might thus be more acoustically salient compared to the one Liu et al. (2019) used and might therefore be more readily detected by infants' brains. Another explanation for the different results between Liu et al. (2019) and our study derives from the morphology of the MMR in infants. As already mentioned, it is commonly observed in ERP studies that the polarity of the MMR switches from positive to negative with the infant's maturational status (e.g., Leppänen et al., 2004). Hence, it could be speculated that there is individual variability within the group of 11- to 12-month-old infants in Liu et al.'s (2019) study, with some infants showing a P-MMR and others an MMN, resulting in a canceling out of the MMR on the group level.

In summary, our results concerning the neural correlates of lexical tones show no decrease in perceptual sensitivity or change in the morphology of the MMR across the tested age groups. Hence, our results do not provide evidence of a perceptual change predicted by the perceptual reorganization process.

Neural correlates of perceptual reorganization of vowel processing in 6- and 9-month-olds

The vowel contrast elicited a P-MMR in 6-month-olds and an adult-like MMN in 9-month-olds. The P-MMR in the 6-month-olds in the early time window is in line with previous studies and show the neural response to the acoustic difference between the vowels /ɛ/ and /i/ (e.g., Marklund et al., 2019; Yu et al., 2019). Additionally, we observed an MMN at 9 months in the later time window which partly contrasts with Yu et al.'s (2019) results. Yu and colleagues found a P-MMR in the late time window in English-learning infants until 12 months of age and an MMN after 13 months for the native English /ɛ/ versus /ɪ/ vowel contrast. There might be several reasons why we observed an MMN instead of a P-MMR or no MMR. The first reason might be related to the difference between the acoustic distances in Yu et al.'s vowel contrast and our vowel contrast. Previous studies have shown that acoustic distance between standards and deviants modulates the polarity of the MMR in infants, where greater acoustic differences are associated with MMNs while smaller acoustic differences are associated with P-MMRs (e.g., Cheng et al., 2013; Cheng & Lee, 2018). Although the vowels tested are relatively similar (/ɛ/ vs. /ɪ/ in Yu et al., 2019 and /ɛ/ vs. /i/ in our study), the contrast we used in the present study is characterized by a greater acoustic distance compared to the contrast Yu et al. (2019) used. The contrast /ɛ/ versus /i/ differs in both F1 and F2 values compared to the formants of the contrast /ɛ/ versus /i/, which could lead to the elicitation of an MMN already at the age of 9 months. Another possible explanation of the different polarities in the 6- and 9-month-olds is that the MMN in the late time window in our results reflects initial formation of phonological categories. Cheour et al. (1998) demonstrated that native contrasts elicited a strong MMN in 12-month-olds while a non-native contrast elicited a reduced MMN, demonstrating the effect of language-specific processing on the MMN in infants. Kuhl et al. (2008) have provided further support for language-specific processing mechanisms. These authors suggest that the decreased MMR in the early time window reflects the decline of universal discrimination ability, whereas the increased MMN in the late time window reflects the transition to language-specific discrimination. We also found a P-MMR in the 6-month-olds in the early time window and an MMN in the 9-month-olds in the late time window. Thus, if we now consider the different polarities in combination with the emergence of the MMRs in the different time windows, it seems that our data concerning the 9-month-olds can be better explained by language-specific processing mechanisms than by acoustic processing of the vowel contrast. The phonological development in conjunction with the greater acoustic salience of the vowel contrast compared to the one used in Yu et al. (2019) might be responsible for the emergence of the MMN already in the 9-month-olds compared to the 13-month-olds in the study by Yu et al. (2019).

Comparison between lexical tones and vowels in 6- and 9-month-olds

Our results indicate different processing of vowels and tones at the ages of 6 and 9 months. The vowel contrast elicited a P-MMR in the early time window at 6 months and an MMN in the late time window at 9 months. The tone contrast elicited a P-MMR in the late time window at 6 months and P-MMRs in the early and late time windows at 9 months. The question is, why we observed a different pattern for the vowel contrast compared to the tone contrast. The occurrence of an MMN for the vowel contrast but not for the tone contrast in the late time window might indicate that at 9 months, infants have built internal phonological categories for the vowel contrast but not for the tone contrast. As previously mentioned, language-specific processing for the vowel contrast might not only be indicated by the MMN in 9-month-olds but also by the processing difference in the different time windows. In comparison to the tone contrast, which elicited a P-MMR in the late time window at 6 months, the vowel processing seems to be faster than tone processing and therefore present in the earlier time window. At 9 months, the infant’s brain also detected the acoustic differences between the tone contrast faster, which might be indicated by the additional emergence of the P-MMR in the early time window. Hence, our data suggest that the perceptual reorganization process for vowels is reflected in the emergence of an MMN in combination with the change from an early response to a late response. In contrast, we observed no comparable change in the processing of the tone contrast, which may indicate that, on the neural level, the perceptual reorganization process may begin later for tones than for vowels.

Difference between regions

Our last finding relates to the processing difference between regions. We observed that tones were processed differently in the central versus left and central versus right regions in 6-month-olds. However, this difference disappeared in the 9-month-olds. Since 9-month-old German-learning infants did not show any processing differences between the three regions for either vowels or tones, the difference between regions in the 6-month-olds cannot be explained by inherent processing differences between vowels and tones but is likely a product of neurological development (Dehaene-Lambertz & Gliga, 2004). Nevertheless, differences between regions in ERP studies must be treated with caution since ERPs have limited spatial resolution due to the difficulties in detecting cortical sources.

To summarize the current infant experiment, the results suggest that lexical tones and vowels are discriminated on the neural level by 6- and 9-month-old German-learning infants. However, potential language-specific processing arises between 6 and 9 months as demonstrated by the change in the polarity of the MMR to the vowel contrast, with additional changes from the early time window to the late time window. However, it remains an open question whether neural discrimination of tones is maintained in adulthood given neural discrimination in 9-month-olds and the diverging results of experiments concerning adults' lexical tone processing. To clarify whether the processing differences between lexical tones and vowels are only evident during the sensitive period for perceptual reorganization or whether they persist in adulthood, we tested German-speaking adults on their neural processing of lexical tones and vowels in a second experiment.

6.3 Experiment 2: Neural correlates of lexical tone and vowel processing in German-speaking adults

In this experiment, we explored adults’ neural responses to the same lexical tone and vowel contrasts used in the infant experiments in order to verify the presence of an MMN in German-speaking adults given the diverging findings on lexical tone processing in non-tonal speakers. The investigation of adults’ neural responses allowed us to further analyze whether the observed differences in infants’ neural responses to vowel and tone contrasts are based on perceptual reorganization or whether they are based on inherent properties of the stimuli.

Additionally, the chosen paradigm allowed us to further investigate whether German adults asymmetrically process vowels and tones and how asymmetric processing is related to acoustic and phonological processing. Different theoretical frameworks assign these processing asymmetries to different processing levels. For example, the Natural Referent Vowel Framework (NRV; Polka & Bohn, 2011; see Section 3.2.3) assigns these asymmetries to the acoustic level of processing. The NRV assumes that vowels with extreme articulatory-acoustic properties act as natural referent vowels. According to this framework, asymmetrical perception is caused by inherent acoustic characteristics of the vowels, such as focalization. Focalization refers to the formant convergence of two adjacent formants where the spectral energy is focused in a narrow space. Focal vowels act as anchor points, and the perceptual salience increases from the less to the more focal vowel. Hence, discrimination should be easier in discrimination tasks where vowels are tested from a less focal to a more focal vowel rather than in the opposite direction. On the other hand, the underspecification theory assigns asymmetries to the phonological level (Lahiri & Reetz, 2002, 2010). Lahiri and Reetz’s (2002, 2010) Featurally Underspecified Lexicon (FUL) combines theoretical approaches of underspecification to speech perception. The basic assumption of this model is that abstract phonological features are mapped onto lexical representations and hence those features are lexically specified. However, not all phonological features need to be specified. Underspecified features are not stored in the lexical representations. Standard stimuli in oddball paradigms create abstract representations of sounds in the brain that reflect the underlying phonological representations in the mental lexicon. In contrast, deviant stimuli create low-level representations that correspond to acoustic rather than phonological representations. Accordingly, if the representations are mutually exclusive (e.g., [CORONAL] and [DORSAL] vowels), a conflict occurs in the mental lexicon. A non-conflict between the two forms of representations occurs if the two representations have no mutually exclusive features. In conflicting conditions, the FUL model predicts asymmetrical MMNs, whereas symmetrical MMNs are predicted in non-conflicting conditions (e.g., Eulitz & Lahiri 2004). This means that discrimination abilities increase in the direction from specified speech sounds to underspecified speech sounds rather than vice versa. Accordingly, in oddball paradigms MMN responses are larger when the standard is specified and the deviant is underspecified than when the standard is underspecified and the deviant is specified (e.g., Schluter et al., 2016, 2017; Shafer et al., 2005).

With the vowel contrast used in this study (/ɛ/ vs. /i/), we were able to investigate the principles of asymmetrical vowel perception. Both vowels are front vowels, and the FUL framework assumes that front vowels belong to coronal speech sounds (Lahiri & Reetz, 2010). According to the FUL framework, coronal vowels are underspecified. In the lexicon. However, [i:] has the feature specification [HIGH], whereas [ɛ:] is underspecified with respect to tongue height (e.g., Scharinger et al., 2012). Regarding this feature, [ɛ:] can be characterized as the underspecified vowel, whereas [i:] can be characterized as the specified vowel. If the specified vowel [i:] is presented as standard, it activates the features of vowel height in the lexicon. This activation results in a conflict condition if [ɛ:] is presented as deviant. If [ɛ:] is presented as standard, no feature of vowel height is activated, which results in a non-conflict condition when [i:] is presented as deviant. Hence, this model would predict a larger MMN for the specified [i:] as standard and the unspecified [ɛ:] as deviant than vice versa.

In contrast, the NRV predicts the opposite direction of asymmetrical vowel perception. In general, vowels at the periphery of the vowel space have greater formant convergence than central vowels do. The most focal vowels are /a/, /i/, /u/, and /y/ (Polka & Bohn, 2011; Sanders & Padgett, 2008; Schwartz et al., 1997, 2005). According to the NRV, the vowel /i:/ is more focal than /ɛ:/. Easier discrimination is therefore expected in the direction from /ɛ:/ to /i:/. This should result in larger MMNs for /ɛ:/ as standard and /i:/ as deviant.

The predictions for lexical tones are not as straightforward as those for the vowel contrast. The application of the underspecification theory to lexical tones is still a matter of debate (e.g., Politzer-Ahles et al., 2016). According to the FUL approach, features are mapped onto the mental lexicon. Since tones are lexically irrelevant in non-tone languages, tones are treated as redundant information in the mental lexicon and therefore do not form abstract phonological representations (Fitzpatrick & Wheeldon, 2000). Hence, the FUL model does not predict any asymmetrical perception for non-tone language speakers. On the acoustic level, the NRV cannot explain asymmetrical lexical tone processing since the NRV mainly explains asymmetrical processing in vowels. However, asymmetrical tone perception may be attributed to differences in salience and the dynamics of the pitch contour (Masapollo et al., 2019; Wayland et al., 2019). The stimulus dynamic hypothesis predicts that a change from a less dynamic pitch pattern to a more dynamic pitch pattern is easier to detect than the opposite (e.g., Wayland et al., 2019). As a result, in the present study, high-rising deviants should elicit larger MMNs compared to mid-level tone deviants.

6.3.1 Method

6.3.1.1 Participants

Twenty-nine native German adults (age 18-31) participated in this study. They reported being monolingual with no previous knowledge of any tone or pitch accented language. All participants were right-handed following the Edinburgh handedness inventory (Oldfield, 1971), had no self-reported hearing deficits, had normal or corrected-to-normal vision, and reported no history of neurological or psychological disorders. All participants received course-credits as compensation for participating in the experiment. Five participants were excluded from the final data analysis because they contributed less than 50 artifact-free trials per condition. Each participant provided written informed consent according to the Declarations of Helsinki.

6.3.1.2 Stimuli

For this experiment, the same stimuli were used as in the infant experiment described in the stimuli section of the infant experiment. The only difference was that in the adult experiment each syllable (/sɛ25/, /sɛ33/, /si25/, /si33/[11]) served as standard and as vowel and tone deviant. In the vowel condition, the vowel changed between standard and deviant from /ɛ/ to /i/ or vice versa while the tone did not change between standard and deviant (either T25 or T33). In the tone condition, the vowel remained constant between deviant and standard, but the tone changed from T25 to T33 or vice versa. All stimuli were normalized in intensity and presented via headphones at 60 dB.

6.3.1.3 Procedure

Participants were tested individually. They were seated approximately 1.5 m from a computer screen and listened to the auditory stimuli via headphones (E-A-RTONE 3A Insert Earphones, Aearo Technologies Auditory Systems). During the presentation of the auditory stimuli, the participants watched a silent movie. All stimuli were presented with Presentation Software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). The duration of the syllables varied between 578 ms – 586 ms, they were presented with an ISI of 800 – 900ms. The varying ISI prevents participants from building rhythmic patterns over the stimulus presentation (e.g., Kirmse et al., 2008). Stimuli were presented in a double passive oddball paradigm. This paradigm allowed us to compare the simultaneous elicitation of the MMN for the vowel and the tone contrast. The adult experiment allowed for a longer experiment duration than the infant experiment, which is necessary to test possible asymmetric effects. In total, 3200 stimuli were presented. These stimuli were grouped into four blocks (800 stimuli per block, 80 % standards, 10 % vowel deviants, 10 % tone deviants). Overall, 640 standards and 80 deviants per condition (80 vowel deviants, 80 tone deviants) were presented. Each of the four syllables (/sɛ25/, /sɛ33/, /si25/, and /si33) were presented as standards as well as tone deviants and vowel deviants in separate blocks, e.g. /sɛ25/ as standard together with /sɛ33/ as tone deviant and /si25/ as vowel deviant in one block and in another block /sɛ33/ as standard, /sɛ25/ as tone deviant, and /si33/ as vowel deviant. The order of the blocks was counterbalanced across the participants. After each block (lasting approximately 20 minutes), participants were offered a short break. In total, the experiment took 80 minutes, and including preparation, the whole experimental session lasted around 3 hours. Each block started with the presentation of eight standards. Deviants were presented in a pseudo-random order: 3-8 standards were

---

[11] The syllable /si33/ was only presented in the adult experiment and included in order to counterbalance the presentation orders.

presented between two deviants. The first eight standards as well as standards directly following deviants were excluded from further analysis. Participants were excluded from the analysis when fewer than 50 deviants per condition were artifact-free.

6.3.1.4 ERP recording and analysis

The electroencephalogram (EEG) was continuously recorded by 30 cap-mounted active Ag/AgCl electrodes (BrainProducts, Gilching, Germany) at a sampling rate of 1000 Hz. Electrodes (Oz, POz, Pz, PO3, C5, P3, P7, CP5, CPz, Cz, FCz, FC3, C3, F3, F5, F7, FPz, FP2, AFz, Fz, C4, FC4, F4, F8, P8, P4, C6, PO4, F6, CP6) were positioned following the 10–20 system convention. The electrooculogram (EOG) was recorded from electrodes placed below and above the right eye. Impedances were kept below 25 kΩ. The ground electrode was placed at FP1 position. The EEG data were analyzed using Brain Vision Analyzer (version 2.01; Brain Products, Gilching, Germany). The signal was filtered offline with a 0.1-30 Hz bandpass filter, and a 50 Hz notch filter. Data were segmented in epochs of 1000 ms, starting 100 ms prior to the onset of the stimuli. The EEG recording was referenced online to the left mastoid and then re-referenced offline to the linked mastoid electrode, and baseline corrected 100 ms before stimulus onset. Eye blinks and eye movements in the segments were corrected by a computer algorithm (Gratton et al., 1983). All other artifacts were detected automatically (exceeding a range of ± 100 µV, lowest allowed activity of 0.5 µV, and a maximum allowed voltage step of 30 µV/ ms) and were excluded from further analysis.

Based on previous research on the timing of the adult MMN, the MMN is expected to occur in the narrow time window of 100-250 ms after the point of divergence. We therefore analyzed the amplitude difference (the MMN) in the time window from 100-250 ms after the point of divergence between the standard and the deviant stimulus. For the vowel contrast, this time window was 200 – 350 ms after syllable onset, and for the tone contrast 400 – 550 ms after the syllable onset. The time window for the tone contrast was later since the point of divergence of the two tones occurred in the middle of the vowel and not at the vowel onset (compare Figure 6). For analysis of the ERP signal, we used the factors deviant type (vowel vs. tone deviants) and region (left, right, and central). For the factor region, we clustered electrodes into three different regions: left (F3, F5, F7, FC3, C3, C5, CP5, P3, P7), right (F4, F6, F8, FC4, C4, C6, CP6, P4, P8), and central (AFz, FPz, Fz, FCz, Cz, CPz, Pz, POz, Oz). We performed two analyses. For both analyses, we compared the MMNs elicited by the vowel and the tone contrast by calculating the identity MMN. The identity MMNs were calculated by subtracting the response to the standard from the response to the deviant of the same vowel or the same tone (e.g., the amplitude difference between /si33/ as standard to /si33/ once presented as tone deviant and once presented as vowel deviant). With this method, the effect of pure physical, acoustic differences can be ruled out, but an observed effect can be attributed to the match or mismatch of the deviant stimulus to the preceding stimuli (e.g., Cornell et al., 2013). In the first analysis, we examined the difference between vowel and tone processing. With the second analysis we tested whether asymmetrical processing can be observed for the presented stimuli. This analysis compares the identity MMNs for the different change directions. In other words, for the vowel contrast, the MMNs elicited by the deviant /ɛ/ are compared with those of the deviant /i/. For the lexical tones, the MMNs are compared between T25 and T33 as deviants.

6.3.2 Results

Figure 12 depicts the grand-averages and corresponding difference waves for vowel and tone contrasts obtained from the F4 electrode.

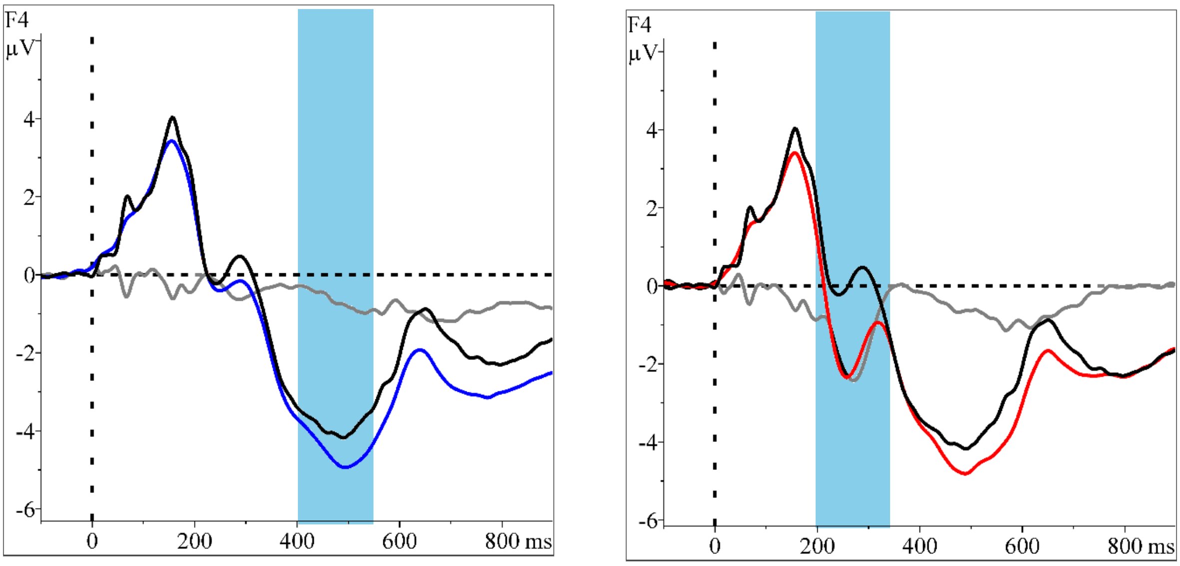

Figure 12. Grand-averages for standards (black line), tone deviant (blue line, on the left), vowel deviant (red line, on the right) and corresponding difference wave (grey line) between standards, and deviants. The blue bar represents the analyzed time window: 200-350 ms for the vowel contrast and 300-550 ms for the tone contrast (each 100-250 ms after the point of divergence).

We conducted statistical analyses with the statistical software R (R Core Team, 2019), and with the lme4 package (Bates et al., 2015). We performed two analyses calculating the identity MMN of the respective contrasts over all electrodes. In the first analysis, we compared the tone processing with the vowel processing. The second analysis investigated the effects of asymmetrical processing in the presented sound contrasts.

Identity MMN of lexical tone vs. vowel processing

In order to test whether the two deviant types (vowel and tone) elicited an MMN component, we compared the identity MMNs (iMMNs) between standards and the respective deviant (vowel and tone) averaged over all electrodes against zero. The iMMNs for the vowel contrast differed significantly from zero (/ɛ/ deviant: β *(SE)* = -1.1475 (0.1582), $df = 28$, $t = -7.254$, $p < 0.001$; /i/ deviant: β *(SE)* = -1.4407 (0.1582), $df = 28$, $t = -9.108$, $p < 0.001$). Similarly, the iMMNs for the tone contrast differed significantly from zero (T25 deviant: β *(SE)* = -0.8190 (0.167), $df = 29$, $t = -4.904$, $p < 0.001$; T33 deviant: β *(SE)* = -0.8653 (0.167), $df = 29$, $t = -5.181$, $p < 0.001$). Subsequently, we computed different mixed models (sum contrast coded) to test which model fits best to the data. The models varied systematically in the added factors (starting with deviant type (vowel vs. tone) and then adding region (left, right, and central brain regions), and their interactions as fixed effects) and were compared against each other (see Table 13 for detailed information about the model comparison). Overall, the results obtained from the model comparison revealed that adding the interaction of deviant type and region significantly improved the model fit, indicating that the two deviant types differ with respect to their influence on the MMN amplitude in the three regions.

Model	Df	AIC	BIC	logLik	deviance	Chisq	Df	Pr (>Chisq)
1. ~Deviant type+ (1+deviant type |subject)	6	20009	20048	-9998.5	19997			
2. ~ Deviant type* region + (1 + deviant type |subject)	10	19996	20061	-9987.8	19976	21.515	4	<0.001

Table 13. Results from the model comparisons with the difference amplitude between standard and deviant (MMN). The comparison is hierarchically organized. The first model was compared to the second model – which fits better to the data. The results indicate best fit to the data for the model, including the interaction of deviant type (vowel, tone deviant) and region (left, right, central).

Overall, the iMMN for the vowel contrast (*M* = -1.011 µV, *SD* = 1.604) was more negative compared to the iMMN for the tone deviant (*M* = -0.645 µV, *SD* = 1.862). This difference between the vowel and the tone contrast was significant (*F* (1) = 61.922, *p* < .001), see Figure 13 for a comparison between the iMMNs of the vowel and tone.

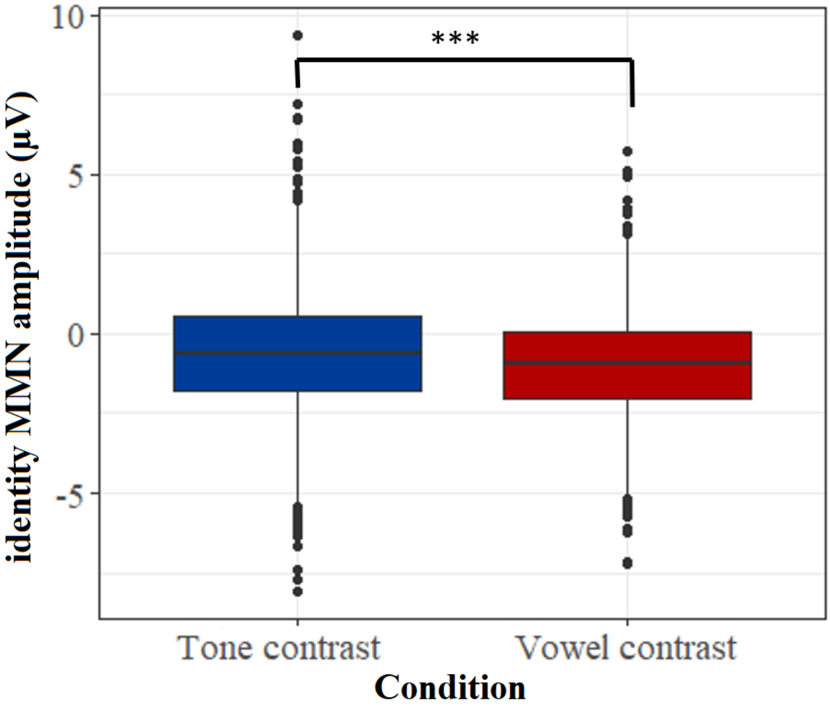

Figure 13. Comparison of the identity MMN amplitudes between vowels and tones for the time window 100-250 ms after the point of divergence. Triple asterisks mark p-values < 0.001.

Differences between regions

In order to resolve the statistically significant interaction of deviant type and region, we performed in an exploratory analysis a post-hoc analysis that compared the amplitude of the identity MMN between the regions for tones and vowels. The post-hoc analysis (Tukey for multiple comparison with adjusted p-values) for the different processing between the regions of vowels and tones revealed significant differences in all regions. Overall, the MMN for vowels had a larger amplitude in the right region compared to the left region. The detailed statistics of the vowel and tone processing in the different regions can be seen in Table 14, and in Figure 13.

Region	**Tone**				**Vowel**			
	β (SE)	df	t-value	p-value	β (SE)	df	t-value	p-value
central - left	-0.020 (0.079)	5140	-0.247	0.999	-0.127 (0.079)	5140	-1.603	0.597
central - right	-0.202 (0.079)	5140	-2.556	0.109	0.164 (0.079)	5140	2.074	0.301
left - right	-0.182 (0.079)	5140	-2.308	0.191	0.229 (0.079)	5140	3.677	0.003

Table 14. Detailed statistics of the post-hoc test (Tukey for multiple comparisons with adjusted p-values) testing for different processing between the regions for tones and vowels.

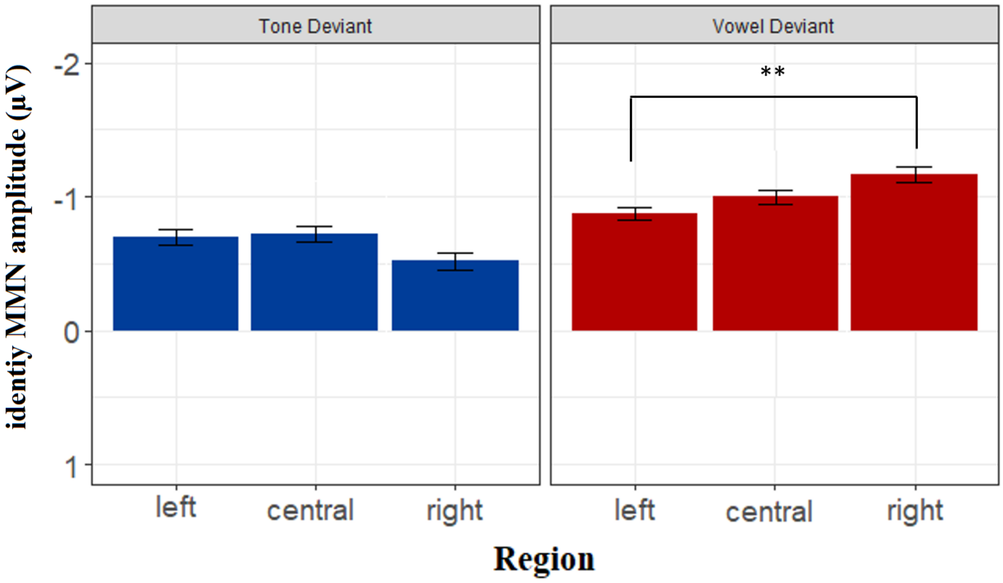

Figure 14. Comparison of the identity MMN amplitudes between the different regions (left, central, right) for tones (blue) and vowels (red). Double asterisks mark p-values < 0.01, whiskers represent ± 1 standard error of the mean.

Asymmetrical effects in the identity MMN

The post-hoc analysis (Tukey for multiple comparison with adjusted p-values) for the different vowels (/ɛ:/ and /i:/) revealed a significant difference in the responses between the two vowels (*β* = -0.219, *SE* = 0.059, *z* = -3.714, *p* < .001). The identity MMN of the vowel /i:/ as deviant elicited a stronger MMN (*M* = -1.121 µV, *SD* =1.678) component compared to when the vowel /ɛ:/ was the deviant (*M* = -0.902 µV, *SD* = 1.519). The post-hoc test (again Tukey for multiple comparison with adjusted p-values) for the two tones revealed no such asymmetry (*β* = -0.079, *SE* = 0069, *z* = 1.135, *p* = 0.256; Tone 33 as deviant: *M* = -0.607 µV, *SD* = 1.994; Tone25 as deviant: *M* = -0.685 µV, *SD* = 1.722). Separate results for the two tones and the two vowels are displayed in Figure 15.

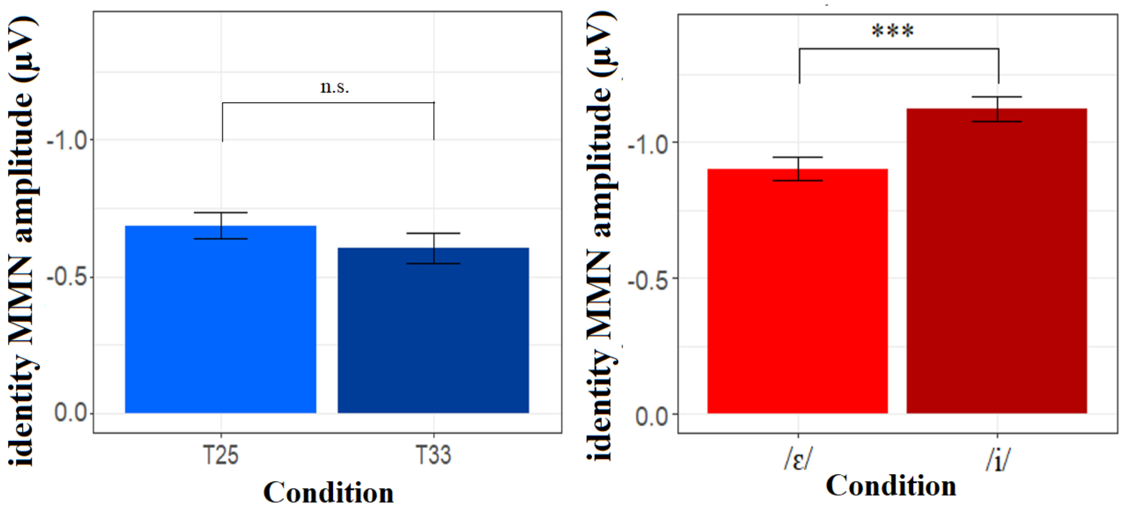

Figure 15. Comparison of the identity MMN amplitudes of the two vowels (/ɛ/ and /i/) and the two tones (T25 and T33). Asymmetrical processing is evident in vowels (right) but not in tones (left). Triple asterisks indicate p-values < 0.001, n.s. indicates p-value > 0.05, whiskers represent ± 1 standard error of the mean.

6.3.3 Discussion

The present study produced four different main findings. First, vowel and tone deviants elicited an MMN in German-speaking adults. Second, vowels elicited a higher-amplitude MMN than tones did. Third, vowels but not tones elicited stronger MMN in the right compared to the left region. Fourth, the MMN to vowels created an asymmetrical pattern with a higher amplitude for the vowel /i:/ as deviant compared to /ɛ:/ as deviant. No such asymmetry was observed for the tone contrast.

Our first result revealed that both types of deviants elicited a robust MMN. Hence, German adults successfully discriminated between the Cantonese high-rising and mid-level lexical tones and the Cantonese /ɛ/ and /i/ vowels. This tone discrimination is in line with several other behavioral and neuroscientific studies that tested tone discrimination in non-tone language speakers (Chandrasekaran et al., 2007; Chen et al., 2018; Gao et al., 2019; Liu et al., 2018; Kaan et al., 2008; Politzer-Ahles et al., 2016; Yu et al., 2017). However, not all of these studies found robust MMNs (Liu et al., 2018; Kaan et al., 2008; Yu et al., 2017), which may indicate that the neural reflections of tone discrimination are influenced by the acoustic salience of the tested contrast, by exposure time, or by the similarity of the tone contrast to specific pitch patterns in the intonation system of the native language. For example, Liu et al. (2018), who tested English-speaking adults on a Mandarin tone contrast with acoustically reduced pitch, found that the MMN emerged during the second half of their experiment. In contrast to the study by Liu and colleagues (2018), the contrast used in the present study was a natural tone contrast without artificially reduced pitch. Therefore, the contrast in the present study might have greater acoustic salience compared to that used by Liu et al. (2018) and therefore elicited an MMN even without much exposure time. Concerning the similarity of the tested tone contrast to the native intonation system, rising tones mark question intonation in German (Grice & Baumann, 2002). Hence, the similarity of lexical tones to higher-order sentential pitch patterns of the native language might influence the discrimination ability and lead to more pronounced MMN responses in German-speaking adults. However, whether the neural responses reflect a possible assimilation to the native language intonation categories or whether the acoustic salience facilitates the processing of tones cannot be deduced from considering the tone contrast without comparison to another speech contrast. The MMN to the Cantonese vowel contrast can be attributed to the fact that German has equivalents to these vowels in its phonological vowel system, which suggests an assimilation process. Previous studies have shown that native and native-like vowels can elicit an MMN in adults (e.g., Näätänen et al., 2007; Peltola et al., 2003; Winkler et al., 1999).

The second finding demonstrated that tone changes elicited a weaker MMN compared to vowel changes in German speakers. This finding indicates that German-speaking adults process these two sound categories differently. A potential source of these differences could be the status of the contrasts in relation to the native language system: the vowels might be assimilated to the phonological vowel system by the German listeners while the tones might not be assimilated to the intonation patterns. The differences in MMN amplitude could therefore stem from a processing difference where vowels are processed on the phonological level while tones are processed acoustically. Another possibility is that differences in the strength of the MMN are related to the acoustic differences between the vowel contrasts and the tone contrasts themselves, with generally higher acoustic salience for vowels over tones. This assumption is supported by findings from behavioral studies that have shown that native and non-native speakers detected segmental changes faster compared to tone differences (Cutler & Chen, 1997; Ye & Connine, 1999).

The only study we are aware of that investigated MMNs for vowel and tone contrasts in the same non-tone speaking adults is that by Yu et al. (2017). This study focused on tone contrasts, but the researcher included stimuli with different vowels to enhance phonetic variability. Their results showed that the vowel deviant (/gy3pa/ vs. /gu3pa/) elicited larger MMN amplitudes compared to the tone deviant (/gu2pa/ vs. /gu3pa/) in both the English- and Mandarin-speaking adults. So far, the results of the present study are consistent with those of Yu et al. (2017). Vowel deviants elicited larger MMN responses compared to lexical tone deviants in non-tone language speakers. However, since we did not test native Cantonese speakers, the present findings do not allow us to conclude whether an overall larger vowel MMN is also present in native Cantonese speakers. Further cross-linguistic studies are needed to evaluate potential differences and similarities between vowel and tone discrimination in native and non-native listeners.

The third result revealed that only the processing of vowels differed significantly in the different regions. The vowel contrast elicited a stronger MMN in the right compared to the left region. This result contradicts other findings that indicate a left-hemispheric preference for linguistic stimuli and a right-hemispheric preference for non-linguistic stimuli (e.g., Zatorre et al., 2002). Nevertheless, especially for vowel processing, other results have been found. Britton et al. (2009) discovered a right-hemispheric preference for the processing of vowels and pure tones for longer stimulus durations. The left hemisphere, in contrast, showed stronger brain activation for fast and short acoustic changes (e.g., Belin et al., 1998; Zatorre & Belin, 2001). In light of the results of the present study, this suggests that the stronger MMN for vowel contrast on the right side might be due to the long stimulus duration of the vowels' steady-state spectral properties. In contrast, the processing of the tone contrast shows no processing difference in brain regions. The symmetrical processing of the pitch, however, could indicate that the right hemisphere initially reacts to the steady-state of the vowel and the pitch, which remain unchanged until the middle of the vowel. The left hemisphere might then react more to the fast modulation of the pitch at the end of the syllable, which results at the end in a consistent processing within the regions.

The fourth main result of this study is that vowels are perceived asymmetrically while tones are not. The MMN in vowels was larger for /ɛ/ as standard and /i/ as deviant compared to /i/ as standard and /ɛ/ as deviant. The acoustic features of the stimuli themselves cannot explain the asymmetry in vowel perception since we used the identity MMN, which is already controlled for the effects of acoustic differences on the MMN. Similar asymmetries have previously been found (e.g., Politzer-Ahles et al., 2016, Scharinger et al., 2012). In the following section we discuss whether the observed asymmetries can be related to phonological (FUL approach) or acoustic (NRV) features. Lahiri and Reetz's (2002, 2010) FUL approach predicts larger MMN components for standards that are specified in the mental lexicon and for deviants that are underspecified compared to underspecified standards and specified deviants (Cornell et al., 2013). Within the FUL approach, [ɛ:] is treated as the underspecified vowel since this vowel has no specification for vowel height, whereas [i:] has the specification [HIGH] (e.g., Scharinger et al., 2012). Hence, this approach predicts a larger MMN for the specified [i:] as standard and the underspecified [ɛ:] as deviant compared to the opposite. However, the results of the present study disagree. We found a larger MMN for [ɛ:] as standard and [i:] as deviant compared to [i:] as standard and [ɛ:] as deviant. Therefore, the FUL approach cannot account for the asymmetrical vowel perception found in the present study.

The other framework that attempts to explain asymmetries is the NRV, which predicts better discrimination from a less focal to a more focal vowel compared to the opposite. Our data support the NRV framework's prediction: /i/ is the more focal vowel compared to /ɛ/, and the MMN component is larger with /ɛ/ as standard and /i/ as deviant compared to /i/ as standard and /ɛ/ as deviant. Since the NRV framework predicts a universal bias, we should find the same effect for the processing of native and non-native vowels. In summary, the asymmetry in the vowel condition is best explained by the NRV framework, which suggests that asymmetrical perception is based on acoustic rather than phonological processing.

We did not observe asymmetrical processing for the tone contrast. This finding contradicts the results by Politzer-Ahles et al. (2016), who found asymmetrical processing in tones for native and non-native speakers of Mandarin. They argue that asymmetrical processing in non-native speakers is an effect of general perceptual biases. In two experiments, they found larger MMNs in the presentation of the Mandarin T3 (falling-rising contour) as deviant and a different tone (rising, T2 or falling, T4) as standard compared to the presentation of T3 as standard and T2 or T4 as deviants. They argue that the difference in dynamic pitch patterns caused the asymmetry. The change from a less dynamic pitch pattern to a more dynamic pitch pattern might be easier to discriminate than the switch from a more dynamic pitch pattern to a less dynamic pitch pattern. Thus, T3 is easier to discriminate when it is presented as deviant because T3 has a more dynamic pitch pattern than the other tones (i.e., rising and falling), and therefore the MMN is more pronounced (e.g., Masapollo et al., 2019; Wayland et al., 2019). According to this explanation, the tone contrast used in the present study (rising and mid-level tones) might have similar dynamic pitch patterns that did not lead to asymmetrical processing. According to the FUL framework, no asymmetrical perception is predicted for non-tone langue speakers. In the FUL framework, tones are treated as redundant information since they are lexically irrelevant in non-tone language speakers, and therefore no conflicting conditions in the mental lexicon and thus no asymmetries occur.

6.4 General discussion: Relating the findings regarding infants and adults

The present study was intended to further investigate the perceptual reorganization in tone and vowel perception in German-learning infants between 6 and 9 months, thus augmenting previous behavioral findings from the neural perspective. To verify that our infant findings reflect the different stages of the perceptual reorganization rather than the inherent properties of the stimuli, we also tested adult participants.

All age groups discriminated both tones and vowels on the neural level. However, we found differences between the developmental trajectories of tone and vowel discrimination. For vowels, we found a P-MMR at 6 months and an MMN at 9 months, whereas for tones we observed P-MMRs in both age groups. Adults exhibited an MMN for both contrasts, but with greater amplitude for vowels compared to tones. Furthermore, our results revealed that adults perceived the vowel contrast, but not the tone contrast, asymmetrically.

To better understand what these results communicate about the perceptual reorganization of tones and vowels, we now compare the infant and adult ERP results with our behavioral data (see Chapter 5) and previous studies.

Neural correlates and perceptual reorganization of lexical tone processing

Previous studies have mostly investigated the perceptual reorganization process of lexical tones on either the behavioral or the neural level (e.g., Liu & Kager, 2014; Liu et al., 2019; Mattock & Burnham, 2006; Yeung et al., 2013). With the current study, we aimed to add to the field of lexical tone discrimination by extending our previous results on behavioral discrimination of the Cantonese high-rising versus mid-level tones (see Chapter 5) with data on the neural processing of this contrast.

In contrast to our behavioral data (Chapter 5), our ERP results reflected a maintained discrimination of tones: both the 6-month-old infants and the 9-month-old infants showed a P-MMR to the tone contrast. Furthermore, we observed that in the 9-month-old infants, tone contrast elicited P-MMR in the early and late time windows, whereas in the 6-month-old infants, a P-MMR was elicited only in the late time window. In addition, the occurrence of an MMN in response to the tone contrast in adults suggests that a general decline in the ability to discriminate this contrast is not to be expected. However, the neural responses of infants and adults differed in their polarity: the MMR in adults showed a negative polarity (which is the typical polarity in sound discrimination by adults), while the polarity of the infants’ MMR was positive. This means that even if we were to test older children, we might not expect a decrease in the amplitude of the MMR at a later stage (e.g., 11 to 12 months) but only a shift in polarity of the MMR. The neural discrimination of lexical tones in both ages in form of the P-MMR do not provide evidence for a perceptual change predicted by the perceptual reorganization process for non-native speech contrasts. Moreover, we observed that with the appearance of an additional P-MMR in the early time window in the 9-month-olds, that tones might be processed faster at 9 months than at 6 months.

However, the lack of evidence for the perceptual reorganization is not in line with the results of our behavioral study (see Chapter 5). At the behavioral level, we showed that German-learning 9-month-old infants could not discriminate the Cantonese high-rising versus mid-level tone contrast, whereas the 6- and 18-month-olds demonstrated perceptual sensitivity to the contrast. Accordingly, our behavioral and neurophysiological results show a different trajectory for the perceptual reorganization process of tones. These diverging results on the behavioral and neural levels may be linked to the development of word learning. Studies have shown that infants already have a rudimentary understanding of simple and high-frequency words at the age of 6 months (e.g., Bergelson & Swingley, 2012). The authors showed that word learning and phonological development take place in parallel and affect each other. The null finding on the behavioral level may therefore suggest that at 9 months, infants are aware that pitch is not a phonological category that carries lexical meaning in German like segmental contrasts do. This assumption is further supported by a word-learning study where English-learning toddlers and adults failed to recognize word-object relations that differed only in pitch contour but recognized word-object relations as soon as a phonemic change was involved in the word-learning task (Quam & Swingley, 2010). Quam and Swingley (2010) provide further evidence that children can already weight acoustic cues during word learning according to whether the acoustic information corresponds to the phonology of the native language. Non-phonological acoustic information (e.g., pitch variations on the lexical level in English) is disregarded during word learning. Further evidence of different pitch interpretations comes from more recent word-learning (Burnham et al., 2018) and word-segmentation (Zahner et al., 2016) studies. Burnham et al. (2018) tested different groups of 17-month-old infants (monolingual English, monolingual Mandarin, and bilingual Mandarin-English-learning) on their ability to integrate tones (Mandarin or Thai) or English intonation contrasts to learning novel words. Their results showed that English-learning 17-month-olds did not integrate any of the pitch patterns (neither native intonation nor non-native lexical tones) when learning novel words. In contrast, monolingual Mandarin learning infants were able to map a native tone contrast to novel words. However, they failed to integrate non-native tones to novel words. The results suggest that even if infants show auditory discrimination for tone contrasts, they do not map native pitch patterns to words, which results in a functional rather than perceptual reorganization process. An interpretation that is supported by the present results. We suggest that infants might already shift their attention to sound categories that are lexically contrastive at the age of 9 months. This decrease in attention to non-phonological acoustic information is likely reflected in our data by a decrease in tone perception on the behavioral level – however infants are still sensitive to pitch differences demonstrated by the persistent neural discrimination. The re-emergence of behavioral discrimination at 18 months and behavioral and neuronal discrimination among German-speaking adults might then be explained by an increase of attention to tonal information at later ages due to the use of pitch in intonation languages such as German (e.g., on the phrasal level).

Neural correlates and perceptual reorganization of vowel processing

As with tones, we found neural discrimination of vowels in 6- and 9-month-old infants and adults. Unlike the tone contrast, German has equivalent phonological categories for the Cantonese vowels /ɛ/ and /i/. Accordingly, we expected discrimination in both infants and adults. On the behavioral level, the ability to discriminate native-language vowels has been shown to become stronger from the age of 6 months on (Tsuji & Cristia et al., 2014). Indeed, we observed neural discrimination of the vowel contrast in infants at 6 and 9 months and adults. However, our neural data did not reveal a stronger MMR with increasing age. Instead, we observed a change in the MMR’s polarity in combination with a change in the time window of the MMR to the vowel contrast between 6 and 9 months. Infants showed a P-MMR at 6 months in the early time window and an MMN at 9 months in the late time window. Previous research has suggested that the P-MMR in the early time window may reflect the less mature speech discrimination process, while the MMN in the late time window may reflect the precursors of the adult MMN, the strength of which is modified by the native phonological system (e.g., Garcia-Sierra et al., 2016; Rivera-Gaxiola et al., 2005). We suggest that for the vowel contrast, the perceptual reorganization process is reflected by the combination of the change in the MMR’s polarity in conjunction with the change in the time window (e.g., Kuhl et al., 2008; Rivera-Gaxiola et al., 2005, 2012). The emergence of the MMN for native-like vowel contrast at 9 months may be interpreted as a more mature response to language-specific phonological information.

Perceptual processing differences between vowels and lexical tones

We now compare the processing of tones and vowels to discuss the trajectory of the perceptual reorganization process on the neural level for the non-native tone contrast compared to the native-like vowel contrast. At the age of 6 months, we found P-MMRs to tones and vowels. However, we found a difference between the time windows. At 6 months, the P-MMR for tone processing was found in the later time window, whereas the P-MMR for the vowel contrast was found in the early time window. This pattern suggests that infants discriminated both vowels and tones but that infants might encode the vowel contrast more readily compared to the tone contrast. In 9-month-old infants, we observed an MMN for the vowel contrast in the later time window and P-MMRs for the tone contrast in the early and late time windows. Three different factors could explain the presence of an MMN in 9-month-olds for the vowel contrast but not for the tone contrast. First, the adult-like MMN may be a reflection of (neurophysiologically) more mature processing (e.g., Leppänen et al., 2004; Yu et al., 2019). However, since we found only an MMN for the vowel contrast but not for the tone contrast, and the infants were tested with both contrasts at the same time, the (neurophysiological) maturation cannot be considered the sole explanation. Second, the acoustic distance in the vowel contrast is greater compared to the tone contrast (Cheng et al., 2013, 2015; Cheng & Lee, 2018). Finally, the emergence of the MMN might reflect the beginning of language-specific phonological processing between the ages of 6 and 9 months for the vowel contrast (e.g., Kuhl et al., 2008; Rivera-Gaxiola et al., 2012). In light of our data, we consider the conjunction of the second and third possibility the most likely because the vowel contrast also induced a stronger MMN than the tone contrast in German-speaking adults and because the MMN in the later time window can be interpreted as the precursor of the adult MMN, the emergence and strength of which is influenced by the native phonological system (e.g., Garcia-Sierra et al., 2016; Marklund et al., 2019; Shafer et al., 2011; Yu et al., 2019). Further cross-linguistic studies are needed to investigate how different acoustic distances in lexical tones and vowels modulate MMRs in tone and non-tone language-learning infants during development.

Conclusion

In the present study, we investigated infants’ and adults’ neural discrimination in a double oddball paradigm consisting of two speech contrasts: a non-native lexical tone and a native-like vocalic contrast. The results indicate no evidence of a perceptual change from 6 to 9 months for the lexical tone contrast. Infants at both ages showed maintained neural discrimination for the tone contrast in form of P-MMRs. However, we observe a perceptual change in infants from 6 to 9 months for the vowel contrast. The vowel contrast elicited a P-MMR in infants at 6 months and an MMN in infants at 9 months. The tone and vowel contrast elicited MMNs in German-speaking adults, with greater amplitude for the vowel contrast than for the tone contrast. We suggest that the emergence of the MMN in 9-month-olds for the vowel contrast is a result of greater acoustic salience for vowels compared to tone and the formation of phonological categories for the vowels. Thus, our results demonstrate evidence for the perceptual reorganization in vowels but not for tones.

CHAPTER 7

Asymmetries in infants' vowel perception: German learning 6- and 9-month-olds' native vs. non-native vowel discrimination

Abstract

Infants' speech perception is characterized by substantial changes during the first year of life that attune the processing mechanisms to the specific properties of the ambient language. The current paper focuses on these developmental changes in non-native compared to native vowel perception. More specifically, the emergence and potential cause of perceptual asymmetries in vowel perception are investigated by an experimental study on German 6- and 9-month-olds' discrimination of a German and a Polish vowel. Results show discrimination without any asymmetry in the 6-month-olds but an asymmetrical pattern with better performance when the vowel changes from the more central Polish vowel to the more peripheral German vowel than vice versa by the 9-month-olds. The results support the native language magnet model (Kuhl, 1991; Kuhl & Iverson, 1995) that assumes that perceptual asymmetries emerge via the acquisition of the internal organization of native language vowel categories. Based on these findings it is argued that the native language vowel system also impacts the perception of non-native vowels.

7.1 Introduction

During early language acquisition infants need to discover which sound differences are relevant and which sound differences are not relevant in the linguistic system of the language that they are learning. One indicator for the formation of native speech sound categories is a process called perceptual reorganization or perceptual attunement. During this process, infants' initial ability to discriminate native sounds is maintained or even enhanced while the ability to discriminate between non-native speech sounds gets weakened with growing age (Aslin & Pisoni, 1980). On the sound level, this developmental change has been shown to occur within the first year of life for consonants (Rivera-Gaxiola et al., 2005; Werker & Tees, 1984), lexical tones (Götz et al., 2018; Liu & Kager, 2014; Mattock et al., 2008; Mattock & Burnham, 2006; Yeung et al., 2013), lexical stress (Höhle et al., 2009; Skoruppa et al., 2009) and vowels (e.g., Polka & Bohn, 1996, 2011; Polka & Werker, 1994; Tsuji & Cristia, 2014).

The current study focuses on the perceptual reorganization of vowels, especially on the question when potential asymmetries, for example, varying discrimination abilities depending on the direction of change of the vowels, occur. To this end, German-learning 6- and 9-month-olds were tested on their discrimination of a German and a Polish vowel that falls within one vowel category in German. Our main question was whether a perceptual change would happen across these ages and whether perceptual asymmetries in the perception of the native and the non-native vowel would already be present in the younger children or whether they would only appear with growing age.

A meta-analysis of 22 different studies on infant vowel discrimination that were published until 2012 revealed that the discrimination of native and non-native vowels develops into different directions from the age of 6 months on: the effect sizes of measures for infants' discrimination of native vowel contrasts increases significantly between 6 and 10 months while the effect sizes of measures for non-native vowel contrasts decrease (Tsuji & Cristia, 2014). This pattern suggests the typical perceptual reorganization with enhancement in the perception of native vowels but weakening in the perception of non-native vowels. This has also been confirmed in studies investigating the neural underpinnings of sound processing: With increasing age, the neurophysiological (MMN) as well as the hemodynamic responses were found to diminish for non-native vowels, and within category length contrasts, whereas the response for native vowel contrasts and for across category (phonemic) changes in vowel length was maintained or enhanced with increasing age (Jansson-Verkasalo et al., 2010; Minagawa-Kawai, Mori, et al., 2007; Minagawa-Kawai, Naoi, et al., 2007). The decline of perceptual sensitivity for non-native vowels was not confirmed in other studies (de Klerk et al., 2019; Mazuka et al., 2014; Polka & Bohn, 1996, 2011). However, the previous findings on non-discrimination might be related to a phenomenon observed in language perception: asymmetrical perceptual biases (e.g., Polka & Werker, 1994).

The status of nativeness or non-nativeness is not sufficient to predict infants' ability to discriminate vowels and their potential developmental changes. One factor that has repeatedly been shown to modulate vowel discrimination is the direction of discrimination, for example, whether for a given vowel pair X-Y, X is chosen as the background stimulus and a response to the change to Y is measured or vice versa. A first study reporting such asymmetrical effects in non-native vowel discriminations was conducted by Polka and Werker (1994). They found a decline in English infants' ability to discriminate the German front-back vowel contrast (/u/-/y/) between the ages of 4 and 10 months. In addition, their results showed better discrimination when tested from /y/ to /u/ than vice versa with this asymmetry being stronger in 6 to 8-month-olds than in 10 to 12-month-olds. As the /u/ was more similar to an English /u/ than the /y/ Polka and Werker (1994) explained this asymmetry as resulting from the perceptual magnet effect (see below). In a further study, Polka and Bohn (1996) tested 6 to 8, and 10 to 12 month-old English- and German-learning infants on their discrimination between the German vowels /u/ and /y/ and on their discrimination of the English /æ/-/ɛ/ contrast. They found neither effects of age nor of language but only of testing direction: discrimination was overall better when testing the change from /y/ to /u/ and from /ɛ/ to /æ/ than in the other direction. Polka and Bohn (1996) suggest that this language and age-independent pattern suggest a universal acoustically based bias that favors the discrimination from a more central to a more peripheral vowel (as determined by vowel height and the front-back dimension) compared to the opposite direction. In another study, Polka and Bohn (2011) showed that Danish-learning infants between 6 months and 12 months discriminated the non-native vowel English contrast /ʌ/-/ɒ/ better when a change occurred from /ʌ/ to /ɒ/ than vice versa. A study by Pons, et al. (2012)

complemented the findings of an asymmetrical perception for native vowels. They tested Catalan and Spanish infants at 4, 6, and 12 months on the /e/-/i/ contrast which is present in both languages. Independent of native language, the 4- and 6-month-olds showed enhanced discrimination when the stimuli changed from /e/ to /i/ than vice versa. The same directional effect was found for the Catalan 12-month-olds, but the Spanish 12-month-olds discriminated the contrast better when the change was in the opposite direction. The authors concluded that after the perceptual reorganization process language specific effects like frequency of vowel occurrence (which is different for the two vowels in Catalan and Spanish) restructures asymmetrical vowel perception. However, Tsuji et al. (2017) could not confirm effects of frequency on native vowel discrimination in Dutch-learning infants – infants' discrimination ability was not modulated by the frequency of the vowel.

To account for the observed asymmetries, Polka and Bohn (2003) proposed the first version of their Natural Reference Vowel framework (NRV). In this initial version vowels that occupy the peripheral positions in the acoustic vowel space (as defined by the relation of F1 and F2 and therefore by the vowel height and the front/back dimension) act as natural referents. Hence, the discrimination of vowels is enhanced in the direction from a central to a more peripheral vowel, e.g., better discrimination occurs from the more central /ɛ/ to the more peripheral /æ/ than vice versa (Polka & Bohn, 1996). However, this proposal could not elucidate all observed patterns of asymmetries. For instance, Danish-learning infants between 6 and 12 months showed better discrimination of a vowel change from /e/ to /ø/ than in the reverse direction despite /e/ being the more peripheral vowel than /ø/ (Polka & Bohn, 2011). Consequently, they considered that it is not the peripherality of vowels that account for asymmetrical vowel perception but rather the inherent acoustic qualities of the vowel, like focalization. Vowel focalization refers to the formant convergence of two adjacent formants, the spectral energy between the two formants is reinforced and the acoustic energy is concentrated on a narrower field. In general, the closer the formants are the more focused is the spectral energy in specific regions and the more focal is the vowel. A focal vowel acts as an anchor point and the perceptual salience increases towards the more focal vowel. The most focal vowels are /a/, /i/, /u/, and /y/ (Polka & Bohn, 2011;

Sanders & Padgett, 2008; Schwartz et al., 1997, 2005)[12], hence, the most focal vowels are also the most peripheral ones but there is no one-to-one mapping between these two dimensions. Previous infant studies confirmed that infants' discrimination abilities are better from less focal to more focal vowels than vice versa. (Polka & Bohn, 1996, 2003, 2011; Polka & Werker, 1994).

Irrespective of the specific acoustic basis for the asymmetrical discrimination, the asymmetries in vowel perception predicted by the NRV framework are considered to be universal and thereby initially language-independent as they are grounded in the general phonetic properties of the vowels. However, these perceptual asymmetries can be modulated by native speaker experience, as they are more evident in non-native vowel contrasts than in native ones (Masapollo et al., 2017; Polka & Bohn, 2011). Thus, it is argued that influences of the native language phonological system may override the perceptual bias.

A second account to explain asymmetries in vowel perception is the Native Language Magnet model (Kuhl, 1991; Kuhl et al., 2006; Kuhl et al., 2008; Kuhl & Iverson, 1995). This framework assumes that asymmetries in vowel perception are not based on universal acoustic biases but on the language-specific internal structure of a vowel category. Most importantly, according to this approach native language sound categories have prototypical exemplars that act as perceptual magnets by attracting other members of the category. Grieser and Kuhl (1989) and Kuhl (1991) tested American English-speaking adults' and 6-month-old infants' discrimination of synthesized sets of variants of the vowel /i/. One set consisted of variants of the prototypical and the other of non-prototypical exemplars from the vowel category. Both infants and adults showed better performance in detecting the change from the non-prototypical to the prototypical vowel compared to a change in the opposite direction.

In a subsequent study, Kuhl et al. (1992) investigated this asymmetry effect cross-linguistically. They tested 6-month-old American English- and Swedish-learning infants on their ability to discriminate exemplars of the vowel categories /y/ and /i/. The exemplar used as a referent vowel in the /y/ category was a prototypical exemplar for

---

[12] Focalization is calculated according to the following equation:

$$E_F = \alpha \sum_{i=1}^{N} \left(\frac{-1}{(\mathrm{F1}_i - \mathrm{F2}_i)^2} + \frac{-1}{(\mathrm{F2}_i - \mathrm{F3}_i)^2} + \frac{-1}{(\mathrm{F3}_i - \mathrm{F4}_i)^2} \right)$$

The lower E_F the more focal the vowel is (Sanders & Padgett, 2008).

Swedish while it should be an untypical exemplar from the category /i/ for a native speaker of English (as English does not have the category /y/). The reverse applied to the /i/ category with a non-prototypical exemplar for Swedish but a prototypical one for English. The results showed language-specific effects with the infants showing the magnet effect only for the category with the prototypical member of their respective native language. Based on these findings Kuhl and colleagues (1992) argued that asymmetries in vowel perception are related to the specific organization of the vowel space in the native language and only emerge as an effect of language acquisition around the age of 6 months. An important critical aspect of these early studies was that the stimuli used in the original study tested a category boundary effect by spanning different phonetic categories rather than examining the perceptual magnet effect itself. Iverson and Kuhl (1995) then conducted another study where they explicitly tested the relation between category goodness, identification, and perceptual sensitivity of the /i/ contrast. Their and later results confirmed that the perceptual magnet effect is not a result of discrimination sensitivity based on phonetic boundaries (Iverson & Kuhl, 2000; Kuhl & Iverson, 1995).

The present study

The goal of the present study was to further investigate the developmental trajectory of asymmetries in vowel perception. According to the revised NLM model (NLM-e; Kuhl et al., 2008), asymmetrical vowel perception can occur at different stages in development. In the first phase, asymmetrical vowel perception is more related to general acoustic phenomena, whereas at later stages asymmetries are related to the language-specific internal organization of native vowel categories in early infancy. In particular, the magnet model makes assumptions regarding within category native language speech contrasts. On the other hand, the asymmetries predicted by the NRV framework refer at any time to inherent phonetic qualities of the vowels but the asymmetries also interact with listeners' experience of their language system. For native vowel contrasts, the asymmetries may be reduced or erased but maintained for non-native vowel contrasts. In this respect, NLM and NRV are not mutually exclusive. In order to investigate the influence of language experience on the vowel discrimination abilities, we tested 6- and 9-month-old German infants' discrimination of the German vowel /ɪ/ and the Polish vowel /ɨ/, see Figure 16 for a phonetic comparison between

German and Polish vowels. The ages were selected based on the meta-analysis by Tsuji and Cristia (2014) that suggests that from 6 months of age nativeness starts to show divergent effects on vowel perception, with increased discrimination abilities for native vowels. Hence, we expect that the 6-month-olds are at the beginning of their language-specific discrimination and not showing the asymmetrical vowel perception on the basis of language-independent phonetic properties. In contrast, the 9-month-olds are expected to show a perceptual sensitivity related to language-specific characteristics. The contrast chosen is expected to fall into a vowel category in German, but German listeners will perceive /ɪ/ as a more typical example of the German vowel category than /ɨ/.

This contrast allows us to make model-specific assumptions. Following the NLM model, we would expect an asymmetrical perception with better discrimination from the Polish /ɨ/ to the German /ɪ/ than in the opposite direction since the German vowel should be the more prototypical instance of the German vowel category than the Polish vowel. This language-specific asymmetry should emerge during language acquisition, i.e., it may be evident in the older age group but not in the younger infants. As the German /ɪ/ is the more focal vowel (Sanders & Padgett, 2008), better discrimination performance from /ɨ/ to /ɪ/ is also expected according to the NRV. However, due to its language-independent phonetic basis, this asymmetry may be evident only to younger infants and may be reduced or erased for the native contrast. The NRV proposed the loss or reduction of these asymmetries for native vowel contrasts but makes no assumptions for within-category discrimination. Thus, according to the NRV model, two alternative predictions for a potential change between the different developmental states of infants on the perception of a native vowel contrast are possible: maintained asymmetrical perception across the different developmental states or reduced asymmetrical perception in infants with more language experience.

7.2 Methods

7.2.1 Participants

In total, 80 German-learning infants participated in this study, 40 6-month-olds (20 female, M = 181 days, r*ange* = 162 – 205 days) and 40 9-month-olds (20 female, M_{age} = 272 days, *range* = 261 – 288 days). Infants were recruited from a database of parents who had expressed interest to participate with their child in research. An additional 21 infants were tested but were excluded from further analyses for the following reasons: crying (6: 3 in the 6month-olds, and 3 in the 9-month-olds), fussiness (1: 9-month-old), not reaching the habituation criterion (11: 6 in the 6month-olds, and 5 in the 9-month-olds), and technical error (3: 3 in the 6month-olds).

According to parental report, infants did not suffer from repeated or acute ear infections, and there were no indications of atypical development. All infants were born full-term and had no hearing deficits. This study was carried out in accordance with the recommendations of the Ethics Committees of the University with written informed consent given by the parents.

7.2.2 Stimuli

A comparison of German and Polish vowels is provided in Figure 16. In order to construct the stimuli for the present study, we recorded two female native monolingual speakers of Polish and German, as well as one balanced bilingual speaker of Polish and German. In a language questionnaire, the bilingual speaker indicated a balanced amount of written and oral communication using Polish and German in her daily life. Six tokens of the vowels (/ɪ/ and /ɨ/, respectively) were recorded in CVC syllables (/pVk/, and /tVk/) in isolation as well as embedded in either Polish or German sentences. The bilingual speaker produced the stimuli in both languages, the monolinguals in their respective native language. The stimuli were presented to the speakers in written form. In Polish, the vowel /ɨ/ was orthographically presented as <y> (e.g., <pyk>, and <tyk>), whereas the German vowel was presented as <pick>, and <tick>, which marks orthographically the vowel /ɪ/.

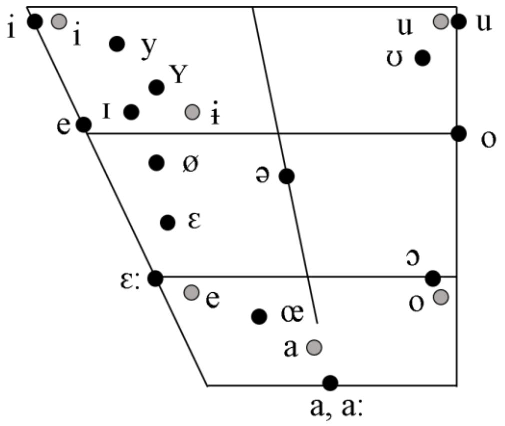

Figure 16. Vowel chart with the comparison of the Polish (grey) and the German vowel (black) system. Modified according to Kohler (1999, p. 199) for German, and Jassem (2003) for Polish vowels.

The stimuli were recorded and digitalized at 44.1 kHz with Audacity in a sound-proof booth. The crucial syllables were then spliced from the utterances and acoustic measures of the vowel formats (F1 and F2 frequencies) were obtained via PRAAT. The stimuli used in the present study were selected from the productions of the bilingual speaker on the basis of the frequencies of the first two formants which should not differ more than 1 SD from the productions of the respective monolingual speaker. The acoustic measures of the two selected tokens from each language are displayed in Table 15.

The four tokens selected for the use in the experiment were then submitted to a rating task, in which 15 German-monolingual adults were asked to identify the selected vowels and rate their goodness as an instance of a German vowel. The stimuli were presented three times mixed with instances of three other German vowels (/u/, /y/, and /i/). In the identification task, participants were asked to decide whether they heard /ɪ/, /ə/, /i/, /u/, or /y/. The corresponding vowels were orthographically displayed on the screen. Before the experiment started, the match between the orthographical representation of the vowels and the corresponding sound was demonstrated to the participant. For example, <i> was presenting the /ɪ/ vowel, and <ie> the /i/ vowel, which is analogue to the German orthography. Participants had to indicate their response by pressing a button on the keyboard. Listeners identified both vowels to an almost identical percentage as belonging to the German vowel category (88 % of the /ɨ/ productions and 89 % of the /ɪ/ productions). Directly after the identification task, the category goodness rating was conducted. The participants were asked to indicate how typical/untypical the vowels sounded for a German vowel on a scale ranging from 1 (no fit) to 7 (very good fit) by pressing the corresponding button on the keyboard. The results revealed that the German adults rated the Polish vowel significantly less typical than the German vowel (*means* = 4.2 vs. 5.6; $\beta = -0.47$, $SE = 0.09$, $t = -4.93$, $p < .001$). Thus, the control study confirms that the selected German and Polish vowel exemplars are identified as instances of the same vowel category but that the German vowel is a better representative of this category for German listeners.

For the use in the infant experiment, the two tokens of a syllable with the same vowel then were concatenated in a random order to a 40 s speech string that included 30 instances of the syllables with an interstimulus interval of 1 s.

Stimuli	Duration (ms)	Vowel Duration (ms)	F0 (Hz)	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
/t ɨ k/ token1	363	68	233 (224-245)	461	2184	2952	3523
/t ɨ k/ token2	361	88	246 (245-251)	474	2195	2922	3523
/t ɪ k/ token1	357	75	219 (212-235)	433	2296	2844	3501
/t ɪ k/ token2	348	81	224 (222-229)	445	2295	2804	3647

Table 15. Acoustic measurements of the vowels in the four syllables used as stimuli in the experiment.

7.2.3 Procedure

Infants were tested in a visual fixation paradigm. During the experiment, they sat on their caretakers' lap facing a monitor at a distance of approximately 1.2 meters. Infants' looking behavior was monitored and recorded by a camera positioned above the screen. Listening time was online coded via a button-box by an experimenter who sat behind a partition. The speech stimuli were presented using Habit2 (Version 2.1.25, Oakes et al., 2015) with an intensity of 65 dB over loudspeakers positioned behind the screen. During the presentation of the speech strings, a black and white checkerboard was displayed on the screen. A silent bouncing ball appeared on the screen during the intertrial interval between the presentation of the speech strings. As soon as the infant fixated the screen, the experimenter pressed a key and thereby started the presentation of the speech string. Trials ended when infants either looked away for more than 2s or when the end of the speech string was reached.

The experiment consisted of three phases: habituation, test, and post-test phase. During the habituation phase, the infants were exposed to one of the strings with the syllables containing one of the vowels until a habituation criterion was reached. The habituation criterion was fulfilled when infants' mean listening time across three consecutive trials had decreased to 50% of the mean listening time of the first three habituation trials. The maximum number of trials within the habituation phase was 18 trials. Half of the infants were habituated with the /tVk/ syllables containing the vowel /ɨ/ and the other half with the syllable containing the vowel /ɪ/. The test phase started immediately after the infant had reached the habituation criterion or after the maximum number of habituation trials had been presented. The test phase contained four trials: two trials presented the two exemplars of the habituated vowel; the other two trials presented the exemplars of the non-habituated novel vowel. The trial order was counterbalanced across infants. If infants discriminate between the two vowels, longer listening times to the novel compared to the habituated vowel are expected. The post-test phase followed directly after the test phase. During the post-test phase, a novel stimulus, with a consonantal change from /tVk/ to /pVk/, was presented to verify infants' attention to the auditory stimuli. In total, the video recordings of 25 % of the participants (randomly selected) were offline re-coded (frame by frame, 25 fps) with ELAN by a second coder. The inter-coder reliability was $r = 0.98$, $p < 0.001$.

7.3 Results

Only data from infants who had reached the habituation criterion within the 18 habituation trials were included in the analysis. Listening times were logarithmically transformed to achieve a normal distribution (see, Csibra et al., 2016). Listening times in the habituation phase did not differ significantly in both age groups (6-month-olds: t (39) = -1.11, p = 0.27; 9-month-olds: t (39) = -1.26, p = 0.21). To compare the predictions made by the NLM and NRV (emergences or presence of asymmetrical vowel perception within the first year of life), we compared different models that were computed with the lmer function from the lme4 package (Bates et al., 2015). The log-transformed listening times of the test trials were used as the dependent variable, and we included stepwise fixed factors (first Condition, with the two levels of habituated vs. novel vowel), then Age (6 and 9 months) and finally the Habituation vowel (/ɨ/ vs. /ɪ/)) to the models and subsequently compared the models to their fit to the data. Subject was included as random factor. The best fitting model included the interaction of Condition (habituated vs. novel vowel), Age (6 and 9 months) and Habituation (/ɨ/ vs. /ɪ/), which had the lowest Akaike Information Criterion (AIC, Akaike, 1998) (*AIC* = 611.87), and was significantly different in Chi-square (χ^2 (6) = 14.09, *p* = 0.029) compared to the model including only condition (with an AIC = 613.96) or the model including the interaction of Condition and Age (with an of *AIC* = 614.62). The output of the model comparisons is provided in Table 16. The precise model was lmer <- log(LT) ~ Condition * Age * Habituation + (1|Subject).

Model	Df	AIC	BIC	logLik	deviance	Chisq	Chi df	Pr (>Chisq)
~ Condition + (1 | Subject)	4	613.96	628.98	-302.98	605.96			
~ Condition * Age + (1 | Subject)	6	614.62	637.16	-301.31	602.62	3.33	2	0.188
~ Condition * Age * Habituation + (1 | Subject)	10	611.87	649.42	-295.93	591.87	14.09	6	0.029 *

Table 16. Results from the model comparison. The comparison is hierarchically organized. The first model was compared to the second model – which fit not better to the data. The better model (the first model) was then compared to the third. Results from Chisquare test and AIC score revealed the best model fit for the model which includes the interaction of Condition, Age, and Habituation as fixed effect and subject as random effect.

With respect to our research question of whether infants at the single age groups show asymmetrical vowel perception, we run separate (planned) models for the different habituation and age groups. The 6-month-old infants showed significant longer listening times during the trials that presented the novel vowel compared to the trials that presented the habituated vowel independent of the vowel presented during habituation (habituation with /ɨ/ : $\beta = 0.3708$, $SE = 0.1241$, $df = 60$, $t = 2.988$, $p = 0.00407$, effect size Cohen´s $d = 0.555$; habituation with /ɪ/: $\beta = 0.2835$, $SE = 0.1248$, $df = 60$, $t = 2.272$, $p = 0.0267$, Cohen´s $d = 0.464$), hence indicating no asymmetrical vowel perception, see Figure 17, and Table 17 for mean listening times across conditions for the different habituation and age groups. In addition, the model comparison shows that the model with the interaction of Condition and Habituation did not lead to a better fit of the model (model with Condition AIC = -386.5 compared to the model with Condition and Habituation AIC = -387.18, $\chi^2(2) = 4.68$, $p = 0.096$).

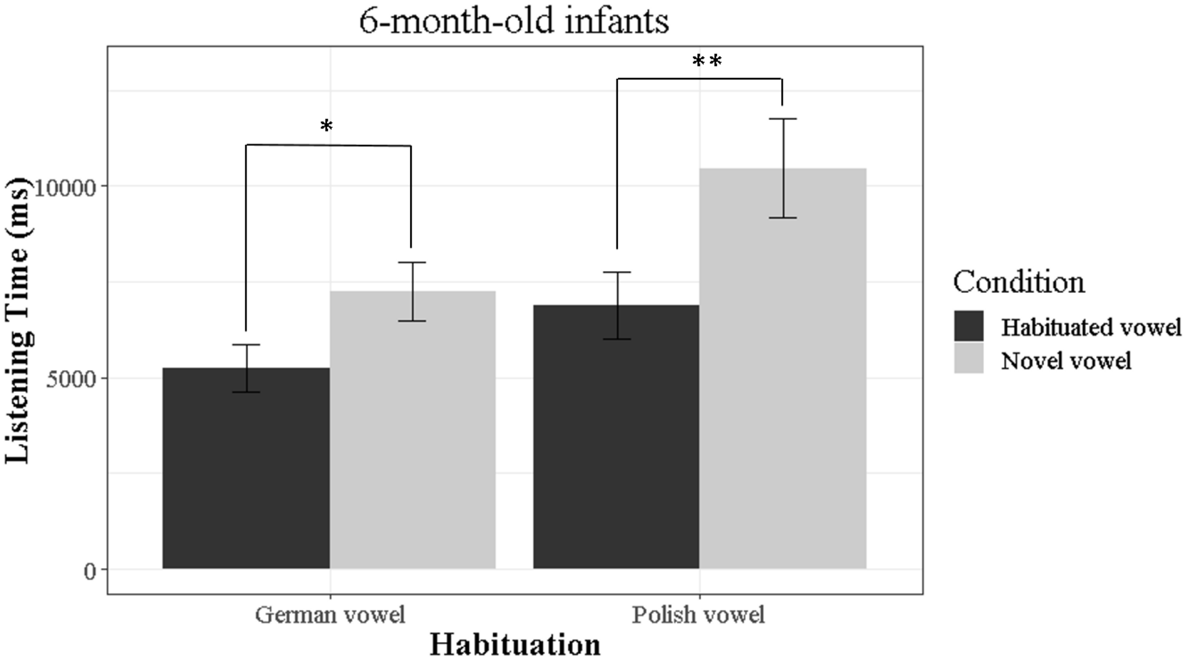

Figure 17. Mean listening times to the novel and habituated vowel of the 6-month-olds per habituation group. Simple asterisk (*) represents p < .05, double asterisks (**) indicates p < .01. Error bars represent standard errors of the mean.

	Habituation to /ɪ/		Habituation to /ɨ/	
	LT (SD) Habituated vowel	LT (SD) Novel vowel	LT (SD) Habituated vowel	LT (SD) Novel vowel
6 months	5258.65 ms (3935.23)	7255.05 ms (4895.30)	6886.25 ms (5464.13)	10464.25 ms (8157.41)
9 months	5074.68 ms (3916.47)	5850.90 ms (5178.98)	5595.48 ms (2896.08)	8901.30 ms (6909.52)

Table 17. Mean listening times (LT) and standard deviation for the habituated and novel vowels separated by habituation and age groups.

The 9-month-old infants showed significantly longer listening times to the novel vowel compared to the habituated vowel only when they were habituated with /ɨ/ (β = 0.3358, *SE* = 0.1129, *df* = 60, *t* = 2.973, *p* = .00424, Cohen´s[13] *d* = 0.510) but not when they were habituated with /ɪ/ (β = 0.0711, *SE* = 0.1335, *df* = 60, *t* = 0.533, *p* = 0.596, Cohen´s *d* = 0.105), see Figure 18, and Table 17. Hence asymmetrical vowel perception was evident, with a significant difference in listening times between the novel and habituated vowel only in the group that was habituated with the Polish vowel.

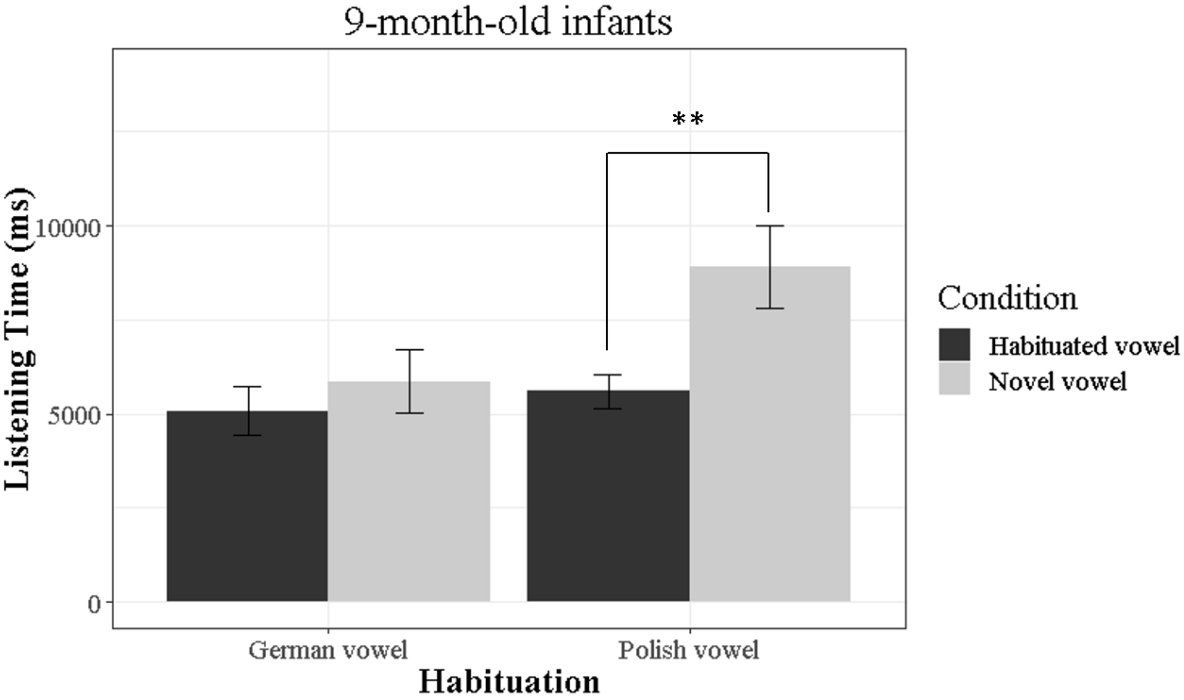

Figure 18. Mean listening times to the novel and habituated vowel of the 9-month-olds per habituation group. Double asterisks (**) indicates p < .01. Error bars represent standard errors of the mean.

7.4 Discussion

The present study aimed to investigate the developmental trajectory of vowel perception during the first year of life. Most importantly, we were interested in asymmetries in vowel discrimination and their potential developmental changes between the ages of 6 and 9 months and compared predictions made by the two central accounts to the sources of asymmetrical vowel perception: the Natural Reference Vowel model (NRV; Polka & Bohn, 1996, 2003, 2011) and the Natural Language Magnet model (NLM; Grieser &

---

[13] The calculation of Cohen's d was based on the following formula: $\frac{mean\,\log(LT)familiar - mean\,\log(LT)novel}{\sqrt{(SD\,\log(LT)familiar)^2 - (SD\,\log(LT)novel)^2}}$

Kuhl, 1989; Kuhl et al., 1992; Kuhl, 1991; Kuhl & Iverson, 1995; Kuhl et al., 2006). To this end, we tested potential developmental changes in vowel discrimination in German infants between 6 and 9 months and whether asymmetries in the discrimination of a native (more focal) and a non-native (less focal) vowel emerge across these ages as suggested by the NLM or whether they are already present at 6 months as suggested by the NRV.

Three main findings arise from our study. First, we did not observe an overall decline in perceptual sensitivity between 6 and 9 months, both age groups showed the ability to discriminate between the two vowels, when they were habituated with the Polish vowel. Second, at 6-months infants showed no asymmetrical performance: their discrimination of the two vowels was not modulated by the type of vowel used during habituation. Third, 9-month-olds showed an asymmetrical performance pattern with clear indications of discrimination only when habituated with /ɨ/ but not when habituated with the vowel /ɪ/. Overall, this result pattern supports the predictions made by the NLM model. We will discuss these three findings and their implications in more detail in the following.

The general result that both age groups showed longer listening times to the novel compared to the habituated vowel requires some modulation of the perceptual attunement account. Comparable to previous findings, German-learning 6-month-olds showed a rather robust ability to discriminate between the two vowels even though the Polish vowel is not part of their native language inventory. At 9 months the ability to discriminate the non-native from the native vowel was not lost, however some perceptual change was evidenced by the emergence of the observed asymmetry in discrimination. To our knowledge, this is the first evidence that demonstrates the relevance of considering potential perceptual asymmetries in studying perceptual reorganization. Our results show two developmental patterns for the same vowel contrast, namely a decline in discrimination ability when only the group habituated to the Polish vowel is considered, while the results support no developmental change when only the group habituated to the German vowel is considered. These findings may support our assumption that some of the seemingly inconsistent results on the perceptual reorganization of vowels may be caused by these kinds of asymmetries in perception. Further research thus needs to consider these potential effects in their experimental settings.

What do our results tell us concerning the two models about the underlying causes of asymmetries in vowel perception? Remember that the NRV considers the source of these asymmetries in the acoustic properties of the vowels with easier discrimination from a central to a peripheral vowel than in the reverse direction. This perceptual asymmetry is not assumed as being based on language experience, but language experience can cause its disappearance for native vowel contrasts. In contrast, the NLM interprets asymmetrical perception as only emerging with the establishment of language-specific vowel categories and their internal structure. According to this model changes from a more prototypical exemplar to a less prototypical exemplar within a native vowel category are harder to detect than changes in the opposite direction.

Coming first to the results of the German 6-month-olds, who did not show any evidence for asymmetries in their perception of the two vowels: they discriminated the two vowels independently of whether habituated to the Polish or to the German one. There may be two potential reasons why the asymmetry that is expected according to the NRV did not show up in the 6-month-olds. First, it could be the case that the two vowels are not sufficiently distinct in their acoustic parameters to create asymmetrical effects. However, asymmetrical perception has been found for numerous vowel contrasts, with some of them being rather distant in the F1/F2 vowel space while others being closer (for an overview, see Polka & Bohn, 2011). Hence, there are no indications that acoustic distance per se would modulate asymmetrical perception. Our failure to find asymmetrical perception aligns with other findings from previous research: de Klerk et al. (2019) also did not find asymmetrical effects in any of the age groups of Dutch infants’ (6-, 8-, and 10-month-olds) discrimination of neither the native /aː/ - /eː/ nor the non-native /æ/ - /ɛ/ vowel contrasts. Mazuka and colleagues’ (2014) findings present a complex picture with type of vowel contrast and age modulating the asymmetries: their Japanese 4.5-month-old infants indicated asymmetrical perception in the German /uː/ vs. /yː/ and the German /oː/ vs. /uː/ contrast but not for the German /iː/ - /eː/ contrast. Interestingly, this pattern reversed completely for the 10-month-olds who showed an asymmetry in their discrimination only for the /iː/ - /eː/. With respect to the predictions of the NRV, especially the failures to find perceptual asymmetries in the youngest age groups is crucial since their vowel perception should not have yet fully attuned to the native language. However, methodological differences between the studies and potential cross-linguistic differences in the speed of attunement to the language-specific vowel system may contribute to these partly inconsistent findings.

In the light of the NLM, the results of the 6-month-olds allow for an easier explanation as the NLM assumes asymmetries as only emerging with the development of a language-specific internal structure of vowel categories that organize the vowel space in terms of prototypicality within a category. However, the original studies by Grieser and Kuhl (1989), Kuhl (1991), and Kuhl et al. (1992) observed language-specific asymmetrical perception already in 6-month-old infants. One could argue that the language-specific vowel organization is not yet present in German 6-month-olds, which cannot be ruled out, since the development of vowel perception in German infants has hardly been studied so far. A second objective could be that we tested the discrimination of a native to a non-native vowel while the NLM only makes predictions for asymmetries within native vowel categories. In fact, it may be the case that – even though the German-speaking adults identified the Polish vowel as an instance of the German category - the acoustic distance between the two exemplars was so large that the younger infants did not map it onto the German vowel categories and therefore no asymmetries could arise. We will come back to this point at the end of this discussion.

In contrast to the 6-month-olds, the 9-month-olds showed an asymmetrical discrimination pattern for the two vowels. The direction of this asymmetry is compatible with both theoretical accounts: the NVR would predict this direction of asymmetry since the Polish vowel is more central than the German vowel. This holds also for the direction of asymmetry observed for the 9-month-olds. As shown by our control study, German adults identify the Polish vowel as an instance of the German /ɪ/ category, but they also rate the Polish vowel as a less good exemplar of this category as the German vowel. This suggests that the German vowel is closer to the prototype of this category than the Polish vowel. Accordingly, the resulting pattern of the 9-month-olds reflects the classical assumption of the NLM with better discrimination from the less prototypical instance to the more prototypical instance. Considered in this way the results of the 9-month-olds also suggest that at that age the native language vowel system is established in a sufficient way to impact the perception and thus adds to the existing evidence that across languages an important time span for the establishment of native language vowel categories is between 6 and 9 months.

The developmental shift that we observed between 6 and 9 months is better compatible with the predictions of NLM compared to NRV. Remember that the NRV assumes that asymmetric perception is based on universal acoustic dimensions, so it should be evident in perception from the youngest age. Developmental changes should only occur as attenuation of this asymmetry to guarantee the efficient processing of the native language. However, our results show an emerging asymmetry between 6 and 9 months which is fully coherent with the NLM view that perceptual asymmetries are a by-product of the internal organization of a native language vowel category and therefore can only occur when these vowel categories have been established. Thus, the overall picture of our findings is better with the prediction of NLM compared to NRV.

One general objection against our study could be that we tested the discrimination of a native and a non-native vowel while studies on non-native vowel discrimination typically test the discrimination of two non-native vowels. However, we do not think that this is a valid plea. First, studies that demonstrated differences in neurophysiological responses to native and non-native vowels always presented contrasts between a native and a non-native vowel (Cheour et al., 1998; Jansson-Verkasalo et al., 2010; Minagawa-Kawai, et al., 2007). Second, and more importantly, it is questionable whether a categorical distinction between native and non-native vowel contrasts is perceptually tenable. German adults identified the two vowels used in our experiments as members of the same vowel category, which is in line with the perceptual assimilation model by Best (1994) suggesting that listeners map encountered sounds to their native language phonological system. Vowels are less categorically perceived than consonants and listeners can discriminate between vowel exemplars even if they belong to the same phonological category in their native language (Fry et al., 1962). Universally, languages make use of the same acoustic space in their vowel repertoire, and the language-specific vowel system is mainly characterized by the way in which this space is divided into categories (in addition to some features like for example rounding or nasalization). Further, given the variability that characterizes vowel production within and across speakers in adults (for a recent demonstration on English, see Whalen et al., 2018) it is hard to argue that listeners encounter vowel productions that they cannot map onto their native language vowel system. Vowel production can vary substantially across infant and adult-directed speech and the vowel space may be extended (e.g., Cristia & Seidl, 2014), shrunken (e.g., Benders, 2013) or even shifted (e.g., Englund & Behne, 2005) such that infants may encounter highly different exemplars that they need to map onto one category from early on. These considerations underline that studies should not just compare native and non-native vowel discrimination but that in the case of non-native vowel perception it is strongly required to consider the relation of the non-native vowel to the native language vowel system of the participant.

To summarize, our results show restricted perceptual reorganization in vowel discrimination between 6 and 9 months of age which is characterized by the emergence of a perceptual asymmetry at 9 months which is not yet present at 6 months. In line with assumptions of an early universal discrimination ability 6-month-olds discriminated the vowels independent of any presentation direction while the 9-month-olds showed a perceptual asymmetry. The direction of this asymmetry is compatible with both models of asymmetric vowel perception, but the finding that it emerges only between the ages of 6 and 9 months favors NLM. We have argued that the mapping of perceived vowels to the native vowel system is relevant for the emergence of these perceptual asymmetries. Future research on infants' development of vowel perception needs to control for these potential perceptual asymmetries and for the relation of potential non-native vowels to the native language vowel system of the population studied.

CHAPTER 8

8.1 General discussion

In this dissertation, I investigated the development of infants' speech perception abilities during the first year of life, focusing on their discrimination of lexical tones (Studies 1 and 2) and vowels (Studies 2 and 3). Lexical tones are not lexically relevant for learning German, the native language of the tested infants. Nevertheless, in the prosodic system of German, pitch information has, for example, pragmatic functions and serves as a cue for phrasing on the sentential level (Gussenhoven, 2004).

The goal of Study 1 was to examine infants' discrimination of Cantonese high-rising versus mid-level tones on the behavioral level to further analyze a U-shaped developmental pattern previously found in Dutch children when tested with Mandarin tones (Liu & Kager, 2014). The presents study augments the findings with another native language (German) and another language contrast (Cantonese). To this end, discrimination of a Cantonese tone contrast was tested with German-learning infants between 6 and 18 months of age, as well as with a group of German and Cantonese adults. The results from the adult perception experiment showed that German adults discriminated Cantonese lexical tone contrasts. However, native Cantonese listeners outperformed German listeners in their discrimination ability. The discrimination performance in German adults might be attributed to the acoustic salience of lexical tones and/ or to the assimilation of lexical tones to native intonation categories. The results from the infant experiments revealed that 6- and 18-month-old German-learning infants discriminated the Cantonese lexical tone contrast, whereas 9-month-olds showed no perceptual sensitivity to this contrast, which supports the idea of a U-shaped developmental pattern for discriminating lexical tones. Furthermore, the study emphasizes the role of the experimental design in assessing perceptual sensitivity to speech contrasts: we showed that only 6-month-old infants who were tested with an infant-controlled initial exposure phase (habituation procedure) showed evidence of discrimination by longer listening times to the novel tone compared to the habituated tone. In contrast, infants of the same age who were presented with fixed initial exposure time (familiarization paradigm) did not show any indication of discrimination. Notably, even within the more sensitive habituation procedure, infants at 9 months did not discriminate the contrast, which demonstrates the reliability of this non-discrimination.

The effect of varying discrimination abilities depending on different experimental procedures was also found in other studies (Cristia et al., 2016; Sundara et al., 2018). The test-retest reliability study (Cristia et al., 2016) reported that larger effect sizes at the group level were generated with infant-controlled initial exposure phases (e.g., in habituation procedures) compared to fixed initial exposure phases (e.g., in familiarization procedures; Cristia et al., 2016) independent of whether the procedure was implemented in head turn or central fixation paradigms. The impact of the choice of the experimental procedure on infants' discrimination abilities is also relevant for theoretical implications. Narayan and colleagues (2010), for example, presented English- and Filipino-learning infants across different age groups with a Filipino alveolar-velar contrast using a modified version of the habituation procedure. In contrast to fully infant-controlled procedures, auditory presentation continued even when infants looked away. The results by Narayan et al. (2010) reveal that only the group of 10- to 12-month-old native Filipino infants discriminated the contrast, while the younger groups failed to discriminate it. English-learning infants did not discriminate the Filipino contrast at any age. Sundara et al. (2018), on the other hand, used a more sensitive infant-controlled procedure (e.g., auditory presentation entirely determined by the infants' behavior) and provided evidence of 4- and 6-month-old English infants discriminating the Filipino contrast. In terms of theoretical perspective, these diverging results support contradictory approaches: Narayan et al.'s (2010) findings favor the perceptual learning account, which predicts enhanced perception of contrasts due to induction effects, whereas Sundara et al.'s (2018) results support the attunement process, which postulates maintenance, reduction, or facilitation effects of perceptual sensitivity with language experience. It is therefore crucial to critically evaluate experimental procedures before drawing theoretical conclusions.

Study 2 of this dissertation further addressed the question of lexical tone perception during the first year of life. The specific goal was to complement the previous findings from behavioral studies by using neural correlates of tone processing. The same Cantonese tone contrast as in Study 1 was examined, but neural data rather than listening times were collected. Furthermore, we tested a native-like vowel contrast in this experiment to compare the trajectory of the perceptual reorganization process of an almost-native (vowel) contrast with a non-native (tone) contrast on the neural level. Accordingly, we expected facilitated or maintained discrimination for the native-like contrast, and we anticipated that the perceptual sensitivity to the non-native lexical tone contrast would decrease within the first year of life (e.g., Cheour et al., 1998; Garcia-Sierra et al., 2016). We included adult participants in this study to compare their neural responses to the same lexical tone and vowel contrasts to verify an overall presence of an MMN given the diverging findings on lexical tone processing in non-tone speakers. The verification of adults’ neural responses allowed us to further assess whether any potential differences in infants’ neural responses to the speech contrasts result from the perceptual reorganization process or whether these differences can be attributed to the inherent properties of the stimuli themselves.

The results of Study 2 revealed a P-MMR to the lexical tone contrast in 6- and 9-month-old infants. The native-like vocalic speech sounds elicited a P-MMR at 6 months and an adult-like MMN at 9 months. In infants, the polarity of the MMR is, among other factors, related to maturation (e.g., age, later language skills), acoustics, and experience to the native language (e.g., Ferjan Ramirez et al., 2017; Friedrich et al., 2009; Leppänen et al., 2004; Yu et al., 2019). In German-speaking adults, both contrasts elicited an MMN, which was greater for the vowel contrast compared to the lexical tone contrast. The results of this study suggest that the perception of tones and vowels in German-learning infants diverges between 6 and 9 months. While vowel processing already showed initial indications of more mature processing by 9 months, lexical tones still show a less mature (positive) response at the same age. The results are in line with general acoustic discrimination abilities at the younger age as predicted by the perceptual reorganization process: the 6-month-olds in Study 2 showed neural discrimination in the form of a P-MMR for the native vowel as well as for the non-native lexical tone contrast. This finding is in line with several other studies investigating neural discrimination in infancy (e.g., Cheng et al., 2013; Ferjan Ramirez et al., 2017; Garcia-Sierra et al., 2016; Marklund et al., 2019; Yu et al., 2019). However, neural discrimination in form of an MMN could only be observed in the 9-month-olds for the vowel contrast but not for the tone contrast. In light of our data, we consider the conjunction of acoustic and language-specific factors the most likely because the vowel contrast also induced a stronger MMN than the tone contrast in German-speaking adults and the MMN in infants can be interpreted as the precursor of the adult MMN, the emergence and strength of which is influenced by the native phonological system (e.g., Garcia-Sierra et al., 2016; Marklund et al., 2019; Shafer et al., 2011; Yu et al., 2019). Our data suggest that perceptual reorganization is evident for the vowel contrast but not for the tone contrast at 9 months.

The change in vowel processing from 6 to 9 months was also further investigated in Study 3. The focus of this study was to investigate how asymmetric vowel processing develops within the first year of life. In particular, we tested whether asymmetric processing is dependent on the experience to the native language or refers to general acoustic perception abilities. For this purpose, the two predominant models (NLM, Kuhl, 1991; Kuhl et al., 1992; Kuhl et al., 2008 and NRV, Polka & Bohn, 2003, 2011) of asymmetric vowel processing were contrasted with each other. The main question was whether asymmetrical perception is evident across age groups, indicating general acoustic processing, or whether the perceptual change is only evident with increasing age, indicating language-specific processing. To answer this question, we compared perceptual sensitivity to a native and a non-native vowel in 6- and 9-month-old German-learning infants. The results showed that infants at 6 and 9 months successfully discriminated the tested contrast, when habituated with the Polish vowel. However, asymmetric vowel processing was evident in 9-month-olds whose discrimination ability depended on which of the two vowels they were habituated with. Early discrimination is in line with the predicted universal sensitivity to speech contrasts. Language-specific processing became evident at 9 months. The developmental shift in perception from 6 to 9 months found in our data supports the language-dependent model (i.e., NLM; Kuhl, 1991; Kuhl et al., 2006; Kuhl & Iverson, 1995) for describing asymmetries. The asymmetric processing had only developed with increasing exposure to infants’ native language and is a result of the internal restructuring process of speech categories.

In line with the assumptions of the perceptual attunement account, we observed that infants show initial sensitivity to a wide range of speech contrasts by 6 months. It is widely assumed that this initial sensitivity is based on general acoustic perception abilities, as is also stated in the different theoretical frameworks as described in the introduction of this dissertation (Section 3.1). Nevertheless, this general acoustic discrimination ability depends on the salience of the contrasts (e.g., Best et al., 1988;

Liu & Kager, 2014; Narayan, 2019, 2020) and the sensitivity of the experimental procedure (e.g., Study 1 in this dissertation; Cristia et al., 2016; Sundara et al., 2018). By being exposed to native language speech sounds, infants begin to build internal phonological categories, which, according to the perceptual reorganization account, lead to a decline in perceptual sensitivity for non-native speech contrasts. However, we did not observe an overall decrease in the perceptual sensitivity of the different tested contrasts. Perceptual changes between 6 and 9 months were observed in multiple patterns. First, we observed the decline of non-native speech perception from 6 to 9 months. Second, the more native-like vowel contrast elicited an MMN in 9-month-olds compared to the 6-month-olds, where the contrast elicited a P-MMR. Furthermore, infants demonstrated maintained neural discrimination for the tone contrast in 6- and 9-month-olds. Third, asymmetrical vowel perception emerged between 6 and 9 months of age. These results shed new light on the different underlying mechanisms of the perceptual reorganization process that cannot be straightforwardly explained by the decrease in perceptual sensitivity to speech sounds that are not present in the infants' environment. Supporting evidence that the decline of non-native (vowel) contrasts, as predicted by the perceptual reorganization, does not occur so consistently, derives from a meta-analysis of infants' vowel discrimination abilities (Tsuji & Cristia, 2014). The authors showed facilitated discrimination for native vowel perception, but in contrast, the anticipated overall decline in perceptual sensitivity to non-native contrasts was not found.

In the next section, the results of the three studies are discussed in more detail with regard to the models described in the introduction (i.e., PRIMIR; Curtin & Werker, 2018; Werker & Curtin, 2005, PAM; Best, 1994, 1995, the NLM, or NLM-extended; Kuhl, 1991; Kuhl et al., 2006; Kuhl et al., 2008; Kuhl & Iverson, 1995, the NRV; Polka & Bohn, 2011). In particular, the ability to discriminate lexical tones, the effects of the experimental procedures, and the emergence of asymmetric vowel processing are discussed.

How do the models explain discrimination abilities of non-native lexical tones in infants, toddlers, and adults?

First, we must distinguish between the discrimination ability of 6-month-old infants and that of 18-month-olds and adults. The three models (the NLM, PAM, and PRIMIR) are based on the perceptual attunement account and the initial perceptual sensitivity to non-native lexical tones is attributed to the general acoustic discrimination ability, but the models differ in how they explain later discrimination abilities. PAM (Best, 1994, 1995) made the most detailed predictions of how well sound contrasts are discriminable. One assumption of the model is that unfamiliar non-native sounds can be perceptually assimilated to native speech categories. As previously mentioned, the intonation contour of internal boundaries of prosodic phrases, and Yes-No questions in German have a similar F0 contour as the high-rising lexical tone in Cantonese, whereas the mid-level tone has a similar F0 contour for declarative sentences. Therefore, it could be supposed that older children and adults assimilate non-native lexical tones to the intonation contour of their native language. In this case, according to the different assimilation types, the tones would be assimilated to two different types of intonation contours (e.g., Yes-No questions and declarative sentences), resulting in the Two-Category assimilation type. This type predicts successful discrimination, which is evident in Studies 1 and 2. In an fMRI study, Chien et al. (2020) showed that non-tone speakers process lexical tones in brain region responsible for the general phonological processing of linguistic pitch. However, we cannot draw clear conclusions about whether infants and adults assimilate the lexical tone contrast to native intonation contours. Our ERP results showed that, despite the evidence of neural discrimination of vowels and tones, the native-like vowel contrast was processed differently compared to the non-native tone contrast. This suggests that the underlying discrimination mechanisms for vowels and tones are different. The processing difference between tones and vowels could be attributed either to the different processing on the acoustic (by general greater perceptual salience of vowels over tones) or the phonological level (by assimilating vowel but not tones to native categories).

PRIMIR (Curtin & Werker, 2007; Werker & Curtin, 2005) assumes, as does PAM, that infants have an innate sensitivity to speech contrasts. PRIMIR explains speech discrimination abilities as a continuous interaction of three different planes (i.e., the General Perceptual, Word Form, and Phoneme planes), which are influenced by dynamic filters (the initial biases, the developmental level, and task demands). The filters, especially the initial biases (e.g., preference for speech and IDS), help to focus the infant’s attention on particular features. According to PRIMIR, the developmental changes observed in Study 1 (initial perceptual sensitivity at 6 months, followed by a decline at 9 months and with a subsequent perceptual rebound effect at 18 months) and the maintained discrimination of lexical tones in Study 2 in 6- and 9-month-olds can be explained by the non-hierarchical order of linguistic development. The formation and restructuring of the different planes (i.e., the General Perceptual plane, Phoneme plane, and Word planes) can occur in parallel. In the beginning, infants show universal abilities to discriminate speech sounds. Their perception is then sharpened by the filters and the General Perceptual plane, which includes phonetic and indexical information, in such a way that representations of the first words are formed as children's segmentation abilities advance. The parallel mechanisms of phonological development and early word learning are supported by experimental data. For example, Bergelson and Swingley (2012) have shown that children already recognize certain words at an early age (between 6 and 9 months), at which point a shift from universal to language-specific speech perception can typically be observed. Concerning the present studies, this means that lexical tone discrimination at 6 months is initially considered subject to the universal discrimination ability. Infants then enter the word-learning stage. The lack of pitch differences on the lexical level in German may lead to a decrease in perceptual sensitivity since infants might suppress pitch information for word learning. Study 1 provides empirical evidence for the decrease in perceptual sensitivity. The finding that infants suppress pitch information for word learning is further supported by a segmentation study with German-learning infants (Zahner et al., 2016). The authors have shown that 9-month-old German-learning infants extracted strong-weak (SW) part words from trisyllabic strong-weak-strong (SWS) carrier words only in situations where high pitch was aligned with the stressed syllable, regardless of intonation patterns. This supports the evidence that at 9 months infants are already aware that pitch is not lexically contrastive and use pitch to segment and recognize units.

The renewed increase in discrimination ability in 18-month-olds and maintained neural discrimination can be explained by the interaction between the filters and the General Perceptual plane. Children perceive pitch differences on the basis of general access to the acoustic information of the speech signal (Curtin & Werker, 2018). Thus, the acoustic salience of the contrast rather than the assimilation process to native intonation patterns contributes to the discrimination ability.

How do the models explain the different results due to the methodological changes?

PRIMIR (Curtin & Werker, 2007, 2018; Werker & Curtin, 2005) has made substantiated predictions regarding the modulation of the sensitivity to discriminate lexical tones through experimental procedures. Within the dynamic filters, task demands are explicitly mentioned. According to PRIMIR, the ability to discriminate sounds is a priori dependent on the experimental procedure. This prediction is reflected in the results from the present studies. As Study 1 demonstrates, 6-month-old infants discriminated the Cantonese tone contrast only in habituation but not in familiarization procedures. Furthermore, Study 2 revealed neural discrimination in 6- and 9-month-olds for the Cantonese tone contrast. Hence, discrimination performance varies as a function of experimental procedures. With procedures addressing other forms of discrimination than looking times (e.g., EEG or pupillometry studies), more detailed and complementary information about the infants' abilities becomes available. In contrast, neither PAM nor NLM-e do make explicit predictions of how the different experimental procedures affect the ability to discriminate between speech sounds.

What explanations do the models provide regarding the asymmetric processing of vowels?

The NRV (Polka & Bohn, 2011) and NLM (Kuhl et al., 2008), which were compared in Study 3, are (partially) in contrast in terms of their assumptions about asymmetric processing. In the NRV, asymmetries are mainly based on the acoustic-phonetic properties of the speech signal and are therefore universal and language independent. Discrimination ability is increased when the more focal vowel is tested against the habituated, less focal vowel. The native language can alter the asymmetric processing. In the case of native vowel contrasts, the asymmetry can be reduced or resolved, while it remains for non-native sounds. The NLM, on the other hand, explains asymmetries with the principle of prototypicality. The vowel space is structured through the formation of phonological categories. The internal vowel structures ensure that certain vowels are considered prototypical, while others are considered non-prototypical. According to the NLM, prototypic sounds attract the other sounds like a magnet; the discrimination is therefore better when tested in the direction from a non-prototypical speech sound to a prototypical speech sound. However, the asymmetry is not apparent until phonological categories have been established; consequently, the asymmetry is language specific. The data from Study 3 showed that 9-month-old German-learning infants showed better discrimination when they were habituated with the non-native vowel and tested with the native vowel. However, 6-month-old infants did not show any difference in discrimination depending on the habituated vowel. Regarding the direction of the asymmetrical perception, both theoretical accounts predict better discrimination abilities in the direction from the non-native to the native vowel. However, they differ in the underlying mechanisms: the NVR predicts this direction because the non-native vowel chosen for our study is more focal than the native vowel, while the NLM predicts this direction because German adults identify the native vowel as a prototype and the non-native vowel as a non-prototype. The asymmetries found in the present study emerged during the developmental trajectory and are consequently a by-product of the internal vowel reorganization process. Hence, the results show more evidence for the NLM or the language-specific account of asymmetrical vowel perception.

8.2 Conclusion

In conclusion, emerging phonological representations lead to a restructuring of discrimination abilities away from general speech discrimination to lexically contrastive speech discrimination. In other words, as soon as (rudimental) speech categories are formed, infants may start to shift their attention on learning words (e.g., Bergelson & Swingley, 2012; Mani & Plunkett, 2010). By recording gaze shifts toward named pictures, Bergelson and Swingley (2012) have shown that from 6 months on infants recognize meaningful words. The word learning may then again lead to a restructuring of sound discrimination abilities resulting in further shaping of phonological categories.

The possible ability to access multiple sources of information (e.g., pitch used for phrasal information) might have caused the successful re-discrimination evidence for lexical tones in the 18-month-olds. Previous studies have already suggested a possible effect of the native prosodic experience on perceptual sensitivity to lexical tones (e.g., Hallé et al., 2004; Liu & Kager, 2014; Ramachers et al., 2018; Shi et al., 2017). In German, for example, the intonation contour of Yes-No questions shows a similar contour of F0 as the tone contrast used in the first two studies at hand (Grice and Baumann, 2002; Braun & Johnson, 2011; Gussenhoven, 2004; Petrone et al., 2017). For this reason, the lexical tone contrast used may be particularly well discriminated at 18 months. As already described in the introduction, non-native sounds can be assimilated to native sounds (Best, 1994). Our and other studies (e.g., Hallé et al., 2004; Liu & Kager, 2014) suggest that assimilation may not only takes place at the segmental level but may also involve the suprasegmental level of prosodic processing. However, the present studies cannot rule out whether the discrimination is due to the assimilation process or whether experience to pitch on the sentential intonation level might support the discrimination performance on the acoustic level.

Further studies are needed to disentangle the possible assimilation and acoustic salience of lexical tones. These studies could be integrated into more natural settings, such as tone changes on the sentence level embedded in natural stories with the comparison of real words and nonsense words. More research is also needed to further investigate the asymmetric processing of vowels. For example, the contrast used in the present dissertation could also be used with Polish-learning infants. The non-native vowel used in this study corresponds to a Polish vowel; consequently, the Polish vowel forms the prototype for Polish-learning infants, whereas the German vowel is the non-prototype. Accordingly, if asymmetric processing is indeed language specific, Polish-learning infants should show asymmetrical perception in the opposite direction compared to German-learning infants, namely from the German to the Polish vowel. If, on the other hand, asymmetric processing is based on acoustic-phonetic properties, the direction of better discrimination should be the opposite, from the Polish to the German vowels. Furthermore, it would be beneficial to investigate a broader age range to determine how linguistic experience further shapes vowel discrimination.

There are two central concluding remarks regarding further studies. First, individual variability should be given greater consideration in infant research. As Study 1 demonstrated, the use of different experimental procedures can lead to diverging findings. Specifically, experimental designs that better capture infants' individual variabilities might be able to display more robust results. The Wordbank database for children's vocabulary development (Frank et al., 2017), demonstrates that the variations in the vocabulary development of children are immense. From this reality it is possible to derive the question of to what extent this individual variation can already be detected in the first year of life. In addition, by capturing individual variation and comparing the variation to later language outcome, developmental scientist might be able to find early predictors of language disorders. The second remark for further studies is that research should include broader age ranges when testing discrimination abilities concerning native and non-native speech contrasts. Up to the present, discrimination abilities have mostly been tested within the first year of life. By testing a broader age range, we can assess a larger picture of infants' and toddlers' ability to perceive native and non-native speech sounds. A broader age range might be especially relevant for testing whether the U-shaped developmental pattern is also evident for other speech contrasts.

References

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

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike* (pp. 199–213). Springer.

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, *46*(2–3), 155–182. https://doi.org/10.1177/00238309030460020601

Aslin, R. N., Jusczyk, P. W., & Pisoni, D. B. (1998). Speech and auditory processing during infancy: Constraints on and precursors to language. In *Handbook of child psychology: Cognition, perception, and language* (pp. 147–254). Wiley.

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology, Vol. 2: Perception* (pp. 67–96). Academic Press. https://ci.nii.ac.jp/naid/10018504644/

Bates, D. M., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.-C., & Samson, Y. (1998). Lateralization of speech and auditory temporal processing. *Journal of Cognitive Neuroscience*, *10*(4), 536–540. https://doi.org/10.1162/089892998562834

Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development*, *36*(4), 847–862. https://doi.org/10.1016/j.infbeh.2013.09.001

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, *46*(2–3), 183–216. https://doi.org/10.1177/00238309030460020701

Best, C. T. (1995). A direct realist view of crosslanguage speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in crosslanguage research* (pp. 171–204). York Press.

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

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, *167*(224), 233–277.

Best, C. T. (2019). The diversity of tone languages and the roles of pitch variation in non-tone languages: Considerations for tone perception research. *Frontiers in Psychology*, *10*, Article 364. https://doi.org/10.3389/fpsyg.2019.00364

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775–794. https://doi.org/10.1121/1.1332378

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360. https://doi.org/10.1037/0096-1523.14.3.345

Bijeljac-Babic, R., Serres, J., Höhle, B., & Nazzi, T. (2012). Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PLoS ONE*, *7*(2), Article e30843. https://doi.org/10.1371/journal.pone.0030843

Bishop, D. V., Hardiman, M. J., & Barry, J. G. (2011). Is auditory discrimination mature by middle childhood? A study using time-frequency analysis of mismatch responses from 7 years to adulthood. *Developmental Science*, *14*(2), 402–416. https://doi.org/10.1111/j.1467-7687.2010.00990.x

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer* (6.0.14) [Computer software]. http://www.praat.org/

Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T., & Höhle, B. (2016). Effects of experience with L2 and music on rhythmic grouping by French listeners. *Bilingualism: Language and Cognition*, *19*(5), 971–986. https://doi.org/10.1017/S1366728915000425

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

Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, *39*(4), 585-594. https://doi.org/10.1016/j.wocn.2011.06.002

Britton, B., Blumstein, S. E., Myers, E. B., & Grindrod, C. (2009). The role of spectral and durational properties on hemispheric asymmetries in vowel perception. *Neuropsychologia*, *47*(4), 1096–1106. https://doi.org/10.1016/j.neuropsychologia.2008.12.033

Burnham, D., & Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology: The Journal of the International Society for*

Developmental Psychobiology, *45*(4), 204–220. https://doi.org/10.1002/dev.20032

Burnham, D., & Francis, E. (1997). The role of linguistic experience in the perception of Thai tones. *Southeast Asian Linguistic Studies in Honour of Vichin Panupong*, 29–47.

Burnham, D., Singh, L., Mattock, K., Woo, P. J., & Kalashnikova, M. (2018). Constraints on Tone Sensitivity in Novel Word Learning by Monolingual and Bilingual Infants: Tone Properties Are More Influential than Tone Familiarity. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02190

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*(03), 455–474. https://doi.org/10.1017/S0142716407070257

Butler, J., Vigário, M., & Frota, S. (2016). Infants' perception of the intonation of broad and narrow focus. *Language Learning and Development*, *12*(1), 1–13. https://doi.org/10.1080/15475441.2015.1020376

Caldwell-Harris, C. L., Lancaster, A., Ladd, D. R., Dediu, D., & Christiansen, M. H. (2015). Factors influencing sensitivitiy to lexical tone in an artificial language: Implications for second language learning. *Studies in Second Language Acquisition*, *37*(2), 335–357. https://doi.org/10.1017/S0272263114000849

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, *1128*(1), 148–156. https://doi.org/10.1016/j.brainres.2006.10.064

Chen, A., & Kager, R. (2016). Discrimination of lexical tones in the first year of life. *Infant and Child Development*, *25*(5), 426–439. https://doi.org/10.1002/icd.1944

Chen, A., Peter, V., Wijnen, F., Schnack, H., & Burnham, D. (2018). Are lexical tones musical? Native language's influence on neural response to pitch in different domains. *Brain and Language*, *180–182*, 31–41. https://doi.org/10.1016/j.bandl.2018.04.006

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

Cheng, Y.-Y., & Lee, C.-Y. (2018). The development of mismatch responses to Mandarin lexical tone in 12- to 24-month-old infants. *Frontiers in Psychology*, *9*. Article 448. https://doi.org/10.3389/fpsyg.2018.00448

Cheng, Y.-Y., Wu, H.-C., Tzeng, Y.-L., Yang, M.-T., Zhao, L.-L., & Lee, C.-Y. (2013). The development of mismatch responses to Mandarin lexical tones in early infancy. *Developmental Neuropsychology*, *38*(5), 281–300. https://doi.org/10.1080/87565641.2013.799672

Cheng, Y.-Y., Wu, H.-C., Tzeng, Y.-L., Yang, M.-T., Zhao, L.-L., & Lee, C.-Y. (2015). Feature-specific transition from positive mismatch response to mismatch negativity in early infancy: Mismatch responses to vowels and initial consonants. *International Journal of Psychophysiology*, *96*(2), 84–94. https://doi.org/10.1016/j.ijpsycho.2015.03.007

Cheour, M., Alho, K., Čeponiené, R., Reinikainen, K., Sainio, K., Pohjavuori, M., Aaltonen, O., & Näätänen, R. (1998). Maturation of mismatch negativity in infants. *International Journal of Psychophysiology*, *29*(2), 217–226. https://doi.org/10.1016/S0167-8760(98)00017-8

Chien, P.-J., Friederici, A. D., Hartwigsen, G., & Sammler, D. (2020). Neural correlates of intonation and lexical tone in tonal and non-tonal language speakers. *Human Brain Mapping*, *41*(7), 1842–1858. https://doi.org/10.1002/hbm.24916

Cornell, S. A., Lahiri, A., & Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 757–772. https://doi.org/10.1037/a0030862

Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, *41*(4), 913–934. https://doi.org/10.1017/S0305000912000669

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

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology*, *52*(4), 521–536. https://doi.org/10.1037/dev0000083

Curtin, S., & Werker, J. F. (2007). The perceptual foundations of phonological development. In M. G. Gareth Gaskell (Ed.) *The Oxford handbook of psycholinguistics*. https://doi.org/10.1093/oxfordhb/9780198568971.013.0035

Curtin, S., & Werker, J. F. (2018). PRIMIR on Tone. *Frontiers in Psychology*, *9*, Article 1007. https://doi.org/10.3389/fpsyg.2018.01007

Cutler, A. (2005). Lexical stress. In D. B. Pisoni & R. Remez (Eds.), *The Handbook of Speech Perception* (pp. 264–289). Blackwell.

Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, *59*(2), 165–179. https://doi.org/10.3758/BF03211886

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141–201. https://doi.org/10.1177/002383099704000203

Dehaene-Lambertz, G., & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience*, *16*(8), 1375–1387. https://doi.org/10.1162/0898929042304714

de Klerk, M., Bree, E. de, 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

Duncan, C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., Polich, J., Reinvang, I., & Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*(11), 1886–1908. https://doi.org/10.1016/j.clinph.2009.07.045

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r - l] distinction by young infants. *Perception & Psychophysics*, *18*(5), 341–347. https://doi.org/10.3758/BF03211210

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

Englund, K. T., & Behne, D. M. (2005). Infant directed speech in natural interaction—Norwegian vowel quantity and quality. *Journal of Psycholinguistic Research*, *34*(3), 259–280. https://doi.org/10.1007/s10936-005-3640-7

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience, 16*(4), 577–583. https://doi.org/10.1162/089892904323057308

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*(3644), 668–670. https://doi.org/10.1126/science.146.3644.668

Ferjan Ramirez, N., Ramirez, R.R., Clarke, M., Taulu, S., & Kuhl, P.K. (2017). Speech discrimination in 11-month-old bilingual and monolingual infants: a magnetoencephalography study. *Developmental Science*, 20, Article e12427. doi:10.1111./dessc.12427

Fitzpatrick, J., & Wheeldon, L. R. (2000). Phonology and phonetics in psycholinguistic models of speech perception. In N. Burton-Roberts, P. Carr, & G. Docherty (Eds.), *Phonological Knowledge: Conceptual and Empirical Issues* (pp.131–160). Oxford University Press.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*(2), 268–294. https://doi.org/10.1016/j.wocn.2007.06.005

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/S0305000916000209

Friederici, A. D., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology, 17*(14), 1208–1211. https://doi.org/10.1016/j.cub.2007.06.011

Friederici, A. D., & Thierry, G. (2008). *Early language development: Bridging brain and behaviour* (Vol. 5). John Benjamins Publishing.

Friedrich, M., Herold, B., & Friederici, A. D. (2009). ERP correlates of processing native and non-native language word stress in infants with different language outcomes. *Cortex*, *45*(5), 662–676. https://doi.org/10.1016/j.cortex.2008.06.014

Frota, S., Butler, J., & Vigário, M. (2014). Infants' perception of intonation: Is it a statement or a question? *Infancy*, *19*(2), 194–213. https://doi.org/10.1111/infa.12037

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*(4), 171–189. https://doi.org/10.1177/002383096200500401

Gao, S., Hu, J., Gong, D., Chen, S., Kendrick, K. M., & Yao, D. (2012). Integration of consonant and pitch processing as revealed by the absence of additivity in mismatch negativity. *PLoS ONE*, *7*(5), Article e38289. https://doi.org/10.1371/journal.pone.0038289

Gao, Y. A., Toscano, J. C., Shih, C., & Tanner, D. (2019). Reassessing the electrophysiological evidence for categorical perception of Mandarin lexical tone: ERP evidence from native and naïve non-native Mandarin listeners. *Attention, Perception, & Psychophysics*, *81*(2), 543–557. https://doi.org/10.3758/s13414-018-1614-8

Garcia-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2016). Relationships between quantity of language input and brain responses in bilingual and monolingual infants. *International Journal of Psychophysiology*, *110*, 1–17. https://doi.org/10.1016/j.ijpsycho.2016.10.004

Geffen, S., & Mintz, T. (2011). Seven-month-olds' discrimination of statements and questions. *Proceedings of the 36th Boston University Conference on Language Development*.

Götz, A., Yeung, H. H., Krasotkina, A., Schwarzer, G., & Höhle, B. (2018). Perceptual reorganization of lexical tones: Effects of age and experimental procedure. *Frontiers in Psychology*, *9*, Article 477. https://doi.org/10.3389/fpsyg.2018.00477

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

Grice, M., & Baumann, S. (2002). Deutsche Intonation und GToBl. *Linguistische Berichte*, *191*, 267–298.

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, *25*(4), 577–588. https://doi.org/10.1037/0012-1649.25.4.577

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421. https://doi.org/10.1016/S0095-4470(03)00016-0

Hay, J. F., Graf Estes, K., Wang, T., & Saffran, J. R. (2015). From flexibility to constraint: The contrastive use of lexical tone in early word learning. *Child Development*, *86*(1), 10–22. https://doi.org/10.1111/cdev.12269

Hirsh-Pasek, K., Nelson, D. G. K., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*(3), 269–286. https://doi.org/10.1016/S0010-0277(87)80002-1

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, *32*(3), 262–274. https://doi.org/10.1016/j.infbeh.2009.03.004

Horowitz, F. D. (1974). Infant attention and discrimination: Methodological and substantive issues. *Monographs of the Society for Research in Child Development*, *39*(5/6), 1–15. https://doi.org/10.2307/1165968

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, *5*, 69–95.

Ingram, J. C., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, *25*(3), 343–370. https://doi.org/10.1006/jpho.1997.0048

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics*, *62*(4), 874–886. https://doi.org/10.3758/BF03206929

Jansson-Verkasalo, E., Ruusuvirta, T., Huotilainen, M., Alku, P., Kushnerenko, E., Suominen, K., Rytky, S., Luotonen, M., Kaukola, T., Tolonen, U., & Hallman, M. (2010). Atypical perceptual narrowing in prematurely born infants is associated with compromised language acquisition at 2 years of age. *BMC Neuroscience*, *11*, Article 88. https://doi.org/10.1186/1471-2202-11-88

Jassem, W. (2003). Polish. *Journal of the International Phonetic Association*, *33*(1), 103–107. https://doi.org/10.1017/S0025100303001191

Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1991). Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience*, *3*(4), 335–344.

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

Kaan, E., Barkley, C. M., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: An event-related potentials training study. *BMC Neuroscience*, *9*, Article 53. https://doi.org/10.1186/1471-2202-9-53

Kavšek, M., & Bornstein, M. H. (2010). Visual habituation and dishabituation in preterm infants: A review and meta-analysis. Research in developmental disabilities, 31(5), 951-975. https://doi.org/10.1016/j.ridd.2010.04.016.

Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, E., & Jacobsen, T. (2008). Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *International Journal of Psychophysiology*, *67*(2), 131–143. https://doi.org/10.1016/j.ijpsycho.2007.10.012

Kohler, K. (1999). *German. International Phonetic Association (ed.) Handbook of the IPA*. Cambridge: Cambridge University Press.

Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107. https://doi.org/10.3758/BF03212211

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 B: Biological Sciences*, *363*(1493), 979–1000. https://doi.org/10.1098/rstb.2007.2154

Kuhl, P. K, & Iverson, P. (1995). Chapter 4: Linguistic experience and the “Perceptual Magnet Effect.” In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (121–154). York Press.

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., 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

Kushnerenko, E. V., Van Den Bergh, B. R. H., & Winkler, I. (2013). Separating acoustic deviance from novelty during the first year of life: A review of event-related potential evidence. *Frontiers in Psychology*, *4*, Article 595. https://doi.org/10.3389/fpsyg.2013.00595

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.) *Laboratory Phonology*, *7*, 637–675. Mouton de Gruyter.

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, *38*(1), 44–59. https://doi.org/10.1016/j.wocn.2010.01.002

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*(5), 527–542. https://doi.org/10.1007/BF01758181

Leppänen, P. H., Guttorm, T. K., Pihko, E., Takkinen, S., Eklund, K. M., & Lyytinen, H. (2004). Maturational effects on newborn ERPs measured in the mismatch negativity paradigm. *Experimental Neurology*, *190* (Supplement 1), 91–101. https://doi.org/10.1016/j.expneurol.2004.06.002

Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance, 22*(5), 1094–1106. https://doi.org/10.1037/0096-1523.22.5.1094

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431–1436. https://doi.org/10.1073/pnas.1114783109

Lidji, P., Jolicœur, P., Kolinsky, R., Moreau, P., Connolly, J. F., & Peretz, I. (2010). Early integration of vowel and pitch processing: A mismatch negativity study. *Clinical Neurophysiology, 121*(4), 533–541. https://doi.org/10.1016/j.clinph.2009.12.018

Liu, H.-M., Chen, Y., & Tsao, F.-M. (2014). Developmental changes in mismatch responses to Mandarin consonants and lexical tones from early to middle childhood. *PLoS ONE, 9*(4), Article e95587. https://doi.org/10.1371/journal.pone.0095587

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. (2016a). Perception of tones by bilingual infants learning non-tone languages. *Bilingualism: Language and Cognition,* 1–15. https://doi.org/10.1017/S1366728916000183

Liu, L., & Kager, R. (2016b). 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., Ong, J. H., Tuninetti, A., & Escudero, P. (2018). One way or another: Evidence for perceptual asymmetry in pre-attentive learning of non-native contrasts. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00162

Liu, L., Peter, V., & Weidemann, G. (2019). A bilingual advantage in infant pitch processing. *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, 1397–1401.

Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neuroscience & Biobehavioral Reviews*, *34*(3), 269–284. https://doi.org/10.1016/j.neubiorev.2009.07.008

Mani, N., & Plunkett, K. (2010). Twelve-month-olds know their cups from their keps and tups. *Infancy, 15*(5), 445–470. https://doi.org/10.1111/j.1532-7078.2009.00027.x

Marklund, E., Schwarz, I.-C., & Lacerda, F. (2019). Amount of speech exposure predicts vowel perception in four- to eight-month-olds. *Developmental Cognitive Neuroscience*, *36*, Article 100622. https://doi.org/10.1016/j.dcn.2019.100622

Masapollo, M., Polka, L., Molnar, M., & Ménard, L. (2017). Directional asymmetries reveal a universal bias in adult vowel perception. *The Journal of the Acoustical Society of America*, *141*(4), 2857–2869. https://doi.org/10.1121/1.4981006

Masapollo, M., Zhao, T. C., Franklin, L., & Morgan, J. L. (2019). Asymmetric discrimination of nonspeech tonal analogues of vowels. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(2), 285–300. https://doi.org/10.1037/xhp0000603

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, *10*(3), 241–265. https://doi.org/10.1207/s15327078in1003_3

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

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

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

Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *Journal of Neuroscience*, *27*(2), 315–321. https://doi.org/10.1523/JNEUROSCI.1984-06.2007

Minagawa-Kawai, Y., Naoi, N., Nishijima, N., Kojima, S., & Dupoux, E. (2007). Developmental changes in cerebral responses to native and non-native vowels: A NIRS Study. *Proceedings of the International Conference of Phonetic Sciences XVI. Saarbrücken, Germany*, 1877–1880.

Morr, M. L., Shafer, V. L., Kreuzer, J. A., & Kurtzberg, D. (2002). Maturation of mismatch negativity in typically developing infants and preschool children. *Ear and Hearing*, *23*(2), 118–136.

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

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*(1), 1–21.

Näätänen, R., Paavilainen, P., Rinne, Teemu, & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review.

Clinical Neurophysiology, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

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

Narayan, C. R. (2020). An acoustic perspective on 45 years of infant speech perception. II. Vowels and suprasegmentals. *Language and Linguistics Compass*, *14*(5), e12369. https://doi.org/10.1111/lnc3.12369

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407–420. https://doi.org/10.1111/j.1467-7687.2009.00898.x

Nashida, T., Yabe, H., Sato, Y., Hiruma, T., Sutoh, T., Shinozaki, N., & Kaneko, S. (2000). Automatic auditory information processing in sleep. *Sleep*, *23*(6), 821–828.

Oakes, L. M., Sperka, D. J., & Cantrell, L. (2015). *Habit 2*. http://habit.ucdavis.edu/about.html

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Peltola, M. S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters*, *352*(1), 25–28. https://doi.org/10.1016/j.neulet.2003.08.013

Peña, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier speech exposure does not accelerate speech acquisition. *Journal of Neuroscience*, *32*(33), 11159–11163. https://doi.org/10.1523/JNEUROSCI.6516-11.2012

Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgrefe-Lang, J., Wartenburger, I., & Höhle, B. (2017). Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics*, *61*(Supplement C), 71–92. https://doi.org/10.1016/j.wocn.2017.01.002

Politzer-Ahles, S., Schluter, K., Wu, K., & Almeida, D. (2016). Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance, 43(10),* 1547–1570. http://dx.doi.org/10.1037/xhp0000242

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

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, *52*(1), 37–52.

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

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221–231. https://doi.org/10.1016/S0167-6393(02)00105-X

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

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/–/ḍ/ 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), 421–435. https://doi.org/10.1037/0096-1523.20.2.421

Pons, F., Albareda-Castellot, B., & Sebastián-Gallés, N. (2012). The interplay between input and initial biases: Asymmetries in vowel perception during the first year of life. *Child Development*, *83*(3), 965–976. https://doi.org/10.1111/j.1467-8624.2012.01740.x

Quam, C., & Swingley, D. (2010). Phonological knowledge guides 2-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language*, *62*(2), 135–150. https://doi.org/10.1016/j.jml.2009.09.003

R Core Team. (2017, 2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Ramachers, S., Brouwer, S., & Fikkert, P. (2018). No perceptual reorganization for Limburgian tones? A cross-linguistic investigation with 6- to 12-month-old infants. *Journal of Child Language*, *45*(2), 290.

Rivera-Gaxiola, M., García-Sierra, A., Lara-Ayala, L., Cadena, C., Jackson-Maldona, D., & Kuhl, P. K. (2012). Event-related potentials to an English/Spanish syllabic contrast in Mexican 10–13-month-old infants. *ISRN Neurology*, *2012*, Article 702986.

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

Sallinen, M., Kaartinen, J., & Lyytinen, H. (1996). Processing of auditory stimuli during tonic and phasic periods of REM sleep as revealed by event-related brain potentials. *Journal of Sleep Research*, *5*(4), 220–228. https://doi.org/10.1111/j.1365-2869.1996.00220.x

Sanders, N., & Padgett, J. (2008). Predicting vowel inventories from a dispersion-focalization model: New results. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, *44*(2), 293–307.

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

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2012). Asymmetries in the processing of vowel height. *Journal of Speech, Language, and Hearing Research*, *55*(3), 903–918. https://doi.org/10.1044/1092-4388(2011/11-0065)

Schluter, K., Politzer-Ahles, S., Al Kaabi, M., & Almeida, D. (2017). Laryngeal features are phonetically abstract: Mismatch negativity evidence from Arabic, English, and Russian. *Frontiers in Psychology*, *8*, Article 746. https://doi.org/10.3389/fpsyg.2017.00746

Schluter, K., Politzer-Ahles, S., & Almeida, D. (2016). No place for /h/: An ERP investigation of English fricative place features. *Language, Cognition and Neuroscience*, *31*(6), 728–740. https://doi.org/10.1080/23273798.2016.1151058

Schwartz, J.-L., Abry, C., Boë, L.-J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion–Focalisation Theory. *Speech Communication*, *45*(4), 425–434. https://doi.org/10.1016/j.specom.2004.12.001

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, *25*(3), 255–286.

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

Shafer, V. L., Morr, M. L., Datta, H., Kurtzberg, D., & Schwartz, R. G. (2005). Neurophysiological indexes of speech processing deficits in children with specific language impairment. *Journal of Cognitive Neuroscience*, *17*(7), 1168–1180. https://doi.org/10.1162/0898929054475217

Shafer, V. L., Yu, Y. H., & Datta, H. (2011). The development of English vowel perception in monolingual and bilingual infants: Neurophysiological correlates. *Journal of Phonetics*, *39*(4), 527–545. https://doi.org/10.1016/j.wocn.2010.11.010

Shi, R., Gao, J., Achim, A., & Li, A. (2017). Perception and representation of lexical tones in native Mandarin-learning infants and toddlers. *Frontiers in Psychology*, *8*. Article 1117. https://doi.org/10.3389/fpsyg.2017.01117

Shi, R., Santos, E., Gao, J., & Li, A. (2017). Perception of similar and dissimilar lexical tones by non-tone-learning infants. *Infancy*, *22*(6), 790–800. https://doi.org/10.1111/infa.12191

Singh, L., & Fu, C. S. L. (2016). A new view of language development: The acquisition of lexical tone. *Child Development*, *87*(3), 834–854. https://doi.org/10.1111/cdev.12512

Singh, L., Fu, C. S., Seet, X. H., Tong, A. P., Wang, J. L., & Best, C. T. (2018). Developmental change in tone perception in Mandarin monolingual, English monolingual, and Mandarin–English bilingual infants: Divergences between monolingual and bilingual learners. *Journal of Experimental Child Psychology, 173*, 59–77.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., Limissuri, R. A., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, *12*(6), 914–919. https://doi.org/10.1111/j.1467-7687.2009.00835.x

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293. https://doi.org/10.1177/0023830909357156

Soderstrom, M., Ko, E.-S., & Nevzorova, U. (2011). It’s a question? Infants attend differently to yes/no questions and declaratives. *Infant Behavior and Development*, *34*(1), 107–110. https://doi.org/10.1016/j.infbeh.2010.10.003

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381–382. https://doi.org/10.1038/41102

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: Re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153–191). John Benjamins.

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., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-ð/ in monolingual and bilingual acquisition of English. *Cognition*, *100*(2), 369–388. https://doi.org/10.1016/j.cognition.2005.04.007

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, *47*(2), 466–472. https://doi.org/10.2307/1128803

Tsao, F.-M. (2017). Perceptual improvement of lexical tones in infants: Effects of tone language experience. *Frontiers in Psychology*, *8*. Article 558. https://doi.org/10.3389/fpsyg.2017.00558

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America*, *120*(4), 2285–2294. https://doi.org/10.1121/1.2338290

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

Tsuji, S., Fikkert, P., Minagawa, Y., Dupoux, E., Filippin, L., Versteegh, M., Hagoort, P., & Cristia, A. (2017). The more, the better? Behavioral and neural correlates of frequent and infrequent vowel exposure. *Developmental Psychobiology*, *59*(5), 603–612. https://doi.org/10.1002/dev.21534

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

Wayland, R., Chen, S., Hong, Y., & Fang, Z. (2019). Perceptual asymmetry in lexical tone perception. *The Journal of the Acoustical Society of America*, *146*(4), 2760–2760. https://doi.org/10.1121/1.5136557

Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, *36*(2), 250–267. https://doi.org/10.1016/j.wocn.2007.06.004

Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (2012). How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology*, *3*, Article 580. https://doi.org/10.3389/fpsyg.2012.00580

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., & Gervain, J. (2013). Speech perception in infancy: A foundation for language acquisition. *The Oxford Handbook of Developmental Psychology*, *1*, 909–925.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35–44. https://doi.org/10.3758/BF03207136

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

Werth, A. (2011). *Perzeptionsphonologische Grundlagen der Prosodie*. Franz Steiner Verlag.

Whalen, D. H., Chen, W.-R., Tiede, M. K., & Nam, H. (2018). Variability of articulator positions and formants across nine English vowels. *Journal of Phonetics*, *68*, 1–14. https://doi.org/10.1016/j.wocn.2018.01.003

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, H., Iivonen, A., & Näätänen, R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research*, *7*(3), 357–369. https://doi.org/10.1016/S0926-6410(98)00039-1

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. *5th International Conference on Language Resources and Evaluation*, 1556–1559.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, *14*(5–6), 609–630. https://doi.org/10.1080/016909699386202

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

Yip, M. (2002). *Tone*. Cambridge University Press.

Yu, H. Y., Tessel, C., Han, H., Campanelli, L., Vidal, N., Gerometta, J., Garrido-Nag, K., Datta, H., & Shafer, V. L. (2019). Neural indices of vowel discrimination in monolingual and bilingual infants and children. *Ear and Hearing*, *40*(6), 1376–1390. https://doi.org/10.1097/AUD.0000000000000726

Yu, Y. H., Shafer, V. L., & Sussman, E. S. (2017). Neurophysiological and behavioral responses of Mandarin lexical tone processing. *Frontiers in Neuroscience*, *11*. Article 95. https://doi.org/10.3389/fnins.2017.00095

Zahner, K., Schönhuber, M., & Braun, B. (2016). The limits of metrical segmentation: Intonation modulates infants' extraction of embedded trochees. *Journal of Child Language*, *43*(6), 1338–1364. https://doi.org/10.1017/S0305000915000744

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*(10), 946–953. https://doi.org/10.1093/cercor/11.10.946

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46. https://doi.org/10.1016/S1364-6613(00)01816-7