The Road to Native Listening:

Language-General Perception, Language-Specific Input

Sho Tsuji

The Road to Native Listening:

Language-General Perception, Language-Specific Input

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op woensdag 9 juli 2014
om 14.30 uur precies

door
Sho Tsuji
geboren op 21 september 1984
in Dortmund, Duitsland

Acknowledgements

The road to my PhD lead over the same general path that many others have walked, a path leading from a blank piece of paper to the contents of this book. However, the specific ups and downs and lefts and rights of this road were determined by the many people I encountered along the way. I want to thank all those who have made this road a very special one to me.

First and foremost, I want to thank my supervisors Paula Fikkert, Anne Cutler, and Alex Cristia. The three of you have been and will be true role models for me, and I feel extraordinary lucky to have been able to work with you.

Paula - no matter how steep and marathonesque the road, you showed me that it is always possible to reach the goal. While you left me the freedom to explore many side-paths, you made sure I would always return to the main road. I couldn't have wished for a more accessible, caring, and motivating supervisor.

Anne - you are a true veteran on the road of native listening, and you paved the way for many of us with your research. Your knowledgeable, straightforward, and enthusiastic advice, in little doses along the way and condensed in the write-up phase, was always to the point and lifted up the quality of my thoughts and writing.

Alex - you crossed my road halfway through my PhD, and you inspired me to peak around yet another bend, and to challenge yet another hill. I am still deeply impressed by your seemingly bottomless resources and willingness to support and advise me whenever needed.

Next to my three official supervisors, I want to thank Reiko Mazuka, my initial mentor in first language acquisition research. You made sure I was wearing the right walking shoes for embarking onto this PhD, and without you I might not have found the entrance to this road. Thanks for your continuous trust and support.

I deeply appreciate that Carlos Gussenhoven, Barbara Höhle, and Daniel Swingley agreed to be on my reading committee. Thank you for

taking the time to read and comment on my thesis, and for participating in my public defense.

Elise de Bree, Claartje Levelt, René Kager, and Mits Ota, I am very grateful that you will be members of my public defense committee.

I thank the many babies and their families who participated in my studies at the Baby Research Center Nijmegen. I am especially grateful to Angela Khadar and Margret van Beuningen for making the BRC such a well-organized and fun place to work at, as well as Imme, Karin, Dirkje, Marjolijn, Diana, Nadine, Charlotte, Brenda, and Jose for their assistance in recruiting and testing infants.

Dear Radboud First Language Acquistion group, you were my wonderful little academic family throughout the three years in Nijmegen. Helen and Nienke, if I could, I would warp us all back into the same office every day at 4 o'clock for *thee en koekjes*. Thank you for being such lovely office-mates over these years, and for always finding occasions to bake cake and decorate the office with balloons. Christina, it has been the greatest fun to share German spleens with you, from reservations against wild Dutch bike rides to wintery warming sips of Feuerzangenbowle. Maarten, be it minimal techno nights out or research brainstormings over coffee, it was good to share PhD life with you. Evelien, Steffi, Titia, Antje, Laura, it was a pleasure to attend conferences and meetings together, and to share hotel rooms and drinks. Each and everyone of you made a great addition to the group. Steffi, special thanks for your help with my Dutch summary.

I visited the Laboratory for Language Development at Riken for four months during my PhD to test Japanese infants. My thanks go to all participating babies and their families. Yamane-san, after having taught me the basics of running infant experiments in the past, you continued to solve my thousand little technical problems during my stay. Thank you! Mari-san, Hatabou, Kijima-san, thank you for all your assistance with administrative issues, recruiting and testing infants. I am grateful to all lab members for making me feel at home every time I come for a visit, including *akachochin* tours and Karaoke adventures. Chie-chan, thanks for always being available for fun in- and outside of research!

Table of Contents

VI

General Introduction

Chapter 1

During their first year of life, infants growing up in a Japanese-speaking household become worse at discriminating between [l] and [r], while infants growing up in an English- (or Dutch-) speaking household become better at discriminating these two speech sounds (Kuhl et al., 2006; Tsushima et al., 1994). This situation is characteristic of a process know as *perceptual attunement.* The general assumption is that infants start out with certain, language-general perceptual abilities and their perception is attuned to their native language at a certain point (cf. Kuhl, 2004). The progress of this language-specific attunement can be modulated by speech sound characteristics such as their language-general acoustic salience or their language-specific distributional frequencies, but also by infants' developing vocabulary and their general cognitive development. These language-general and language-specific characteristics are assumed to interact over the course of phonological development, ultimately culminating in language-specific, robust, and abstract phonological representations. The current thesis puts a spotlight on many of these factors, as will be illustrated further below after a brief overview of studies on and models of early speech sound perception.

1 Perceptual attunement

The insight that characteristics of the input influence infants' developing speech sound perception is a central achievement of the past decades' work on early phonological acquisition. Two seminal studies, one on consonants (Werker & Tees, 1984), the other on vowels (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), illustrated that infants from different language backgrounds start out with a comparable ability to discriminate a wide variety of human speech sounds, but attune their perception to the contrasts relevant in their native language during their first year of life. Werker and Tees (1984) showed that English-learning infants' ability to discriminate two (Hindi and Nthlakampx) non-native

consonant contrasts declined between 6-8 and 10-12 months of age, whereas Hindi-learning and Nthlakampx-learning 10-12-month-old infants retained the ability to discriminate the contrasts that were part of their respective languages. Perceptual attunement takes place even earlier for vowels, as was demonstrated later by Kuhl et al. (1992). Centered around these two studies, subsequent research from a multitude of languages has replicated the process of perceptual attunement during the first year of life (for an overview, see Kuhl, 2004). Contrary to the early assumption that infants' sensitivity would always decline for non-native contrasts and be maintained for native contrasts, these studies document that infants prior to perceptual attunement do not necessarily have the ability to discriminate all human speech sounds, but are insensitive to some contrasts early on (e.g., Mazuka, Hasegawa, & Tsuji, 2013). In case such speech sounds are native, infants' discrimination has been shown to improve over the first year of life (e.g., Kuhl et al., 2006). It has also been documented that infants continue discriminating particular non-native contrasts without exposure (Best, McRoberts, & Sithole, 1988). Other studies have suggested that the ability to use phonetic detail in speech sound discrimination tasks is not the same as using phonological knowledge in lexical tasks (e.g., Stager & Werker, 1997).

To capture these different facts, a number of models of early infant speech perception have been proposed.

2 Models of infant speech perception

Several models of infant speech perception account for perceptual attunement. The Native Language Magnet model (NLM; Kuhl, 1994; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola, & Nelson, 2008) assumes that exposure to native speech sounds leads to the formation of prototypes. These prototypes, which have been described as the representations most often activated (Kuhl et al., 2008), or as the centers of a vowel category (cf. Feldman, Griffiths, & Morgan, 2009), act as magnets and warp the perceptual space such that it shrinks around prototypical phonemes. As a consequence, discrimination of tokens close to a prototype becomes worse than discrimination of tokens towards the category edge (for a similar mechanism, see also Word Recognition and

Phonetic Structure Acquisition, WRAPSA model, Jusczyk, 1993). Since warping depends on exposure to native speech sounds, no such magnet effect occurs for non-native speech sounds.

The developmental framework for Processing Rich Information from Multi-dimensional Interactive Representations (PRIMIR; Werker & Curtin, 2005) conceives perception as operating in multiple planes. In the General Perceptual plane, discrimination abilities are encoded that are initially independent of language exposure. Based on language experience, this plane is reorganized such that some boundaries are erased, enhanced, or shifted. PRIMIR does not rely on prototypes, but assumes infants' tracking of token distributions as a key process to reorganization (cf. Maye, Werker, & Gerken, 2002). The language-specific representations in this plane are neither robust nor abstract, and phonological categories will only emerge once the child begins to learn words and store them in the Word Form plane, at which point the Phoneme plane will begin to develop (compare this with the WRAPSA model, e.g. Jusczyk, 1993). Thus, another difference compared to NLM is that perceptual reorganization is not only influenced by frequency but also word learning.

Despite their differences, these models agree that speech sound perception becomes attuned to infants' native language during the first year of life, and that input frequencies are a central component in shaping infants' perception.

Another prominent model, the Perceptual Assimilation Model (PAM; Best, 1994), primarily focuses on how non-native sounds are processed once native perceptual categories have already been formed. PAM proposes that listeners tend to assimilate non-native speech sounds to native speech sound categories based on their similarity. In case two non-native speech sounds can be mapped onto two different native speech sounds, infants will continue discriminating them. However, in case they are perceived as equally acceptable exemplars of the same native sound, discrimination will decline. Similarity in PAM depends on the articulators involved in producing a speech sound, predicting that discrimination of two non-native speech sounds involving the same articulatory gesture is poorer than of those involving two or more different articulatory gestures. Unlike the previous two models, PAM does

not provide an account of how native and non-native categories come to be treated differently in the first place.

3 Towards consolidating research findings

The process of perceptual attunement has been investigated with a variety of speech sound contrasts and in infants of many linguistic backgrounds. Moreover, infants have been tested with diverse behavioral and neural methods that assess discrimination based on measures ranging from sucking responses and looking times to brain activation. This variability renders it challenging to extract commonalities between study outcomes, for instance when measuring to what extent they fit a model's predictions on perceptual attunement introduced in Section 2 (an issue I will come back to in Section 4).

This variability can also be exploited to gain a better insight into the methodological and conceptual factors affecting infant speech sound perception. Indeed, qualitative reviews and quantitative meta-analyses are common and complementary tools to get a systematic overview of a given topic. A key problem they share, however, is that they reflect the state of the art at a certain point in time and can be outdated rapidly. In order to remedy this situation, we created a database of infant vowel discrimination studies, which can be accessed and updated online. This tool is introduced in **Chapter 2**, along with two examples to illustrate its usefulness in exploring both methodological and conceptual questions (whether effect sizes differ depending on the method used, and whether spectral or phonological distance is a better predictor of effect sizes). The studies on early vowel discrimination contained in the database are qualitatively reviewed in **Chapter 3**. The review is organized into four central topics, comprising methodological considerations, infants' vowel perception prior to, or independent from, language exposure, developmental changes in vowel perception, and comparison of monolingual and typically developing infants with other populations.

4 Evaluating perceptual attunement in vowels

To date the evidence for perceptual attunement derives predominantly from consonant studies. Perceptual attunement is by and large assumed to apply to all speech sounds. Given that vowels and consonants might not be entirely comparable (cf. Chapter 4 for a detailed discussion), it is of particular interest to assess to what extent perceptual attunement also holds for vowels. This is the central reason that the database introduced in Chapter 2 focuses on vowels. Moreover, although the critical age for perceptual attunement in vowels has been assumed to be around six months of age ever since Kuhl et al.'s (1992) seminal study, evidence for later attunement (Polka & Bohn, 2011; Pons, Albareda-Castellot, & Sebastián-Gallés, 2012), or a lack of attunement (Polka & Bohn, 1996) has also been reported. In **Chapter 4**, the available evidence for perceptual attunement in vowels is therefore investigated by means of a quantitative meta-analysis. Only the subset of studies assessing discrimination of the same contrast in two or more age-groups out of the database introduced in Chapter 2 is considered in this analysis.

A meta-analysis serves to quantitatively summarize results from single experiments on the same topic. In order to make the outcomes from different studies comparable, they are transformed into a common effect size metric and weighted in relation to their sample size. To assess evidence for perceptual attunement, we focused on whether the infants' age and whether the contrast under investigation was native or non-native were significant predictors of effect sizes. Based on the literature on perceptual attunement, we expected an interaction between these two factors such that effect sizes for the discrimination of native contrasts would increase, whereas effect sizes for the discrimination of non-native contrasts would decrease with age.

5 The influence of frequency on infant vowel perception

The models of infant speech perception introduced in Section 2 both ascribe the frequency of exposure to certain speech sounds a critical role in perceptual attunement. And yet, the meta-analysis presented in Chapter 4 was almost exclusively based on studies assessing exposure in a

categorical way, namely by comparing a zero-exposure (non-native) contrast to an above-zero-exposure (native) contrast. Little research has been conducted on the influence of differences in *relative* exposure on vowel discrimination. Only two studies so far have investigated the influence of frequency of exposure (in consonants: Anderson, Morgan, & White, 2003; in vowels: Pons, Albareda-Castellot, & Sebastián-Gallés, 2012; cf. Chapter 5 for details), and neither of these studies has focused on frequency-dependent improvements in native discrimination. In order to measure whether differences in the amount of exposure affect vowel discrimination, we compared infants' discrimination of a frequent vowel contrast to their discrimination of an infrequent vowel contrast in **Chapter 5**. Dutch infants around six months of age were assessed on their discrimination of two native contrasts, frequent [ɪ-eː] and infrequent [ʏ-øː]. To seek converging evidence, we measured infants in a behavioral paradigm as well as using near-infrared spectroscopy (NIRS), a neuroimaging technique which allows to get measures of discrimination as well as of lateralization as an index of increasingly linguistic processing (Minagawa-Kawai, Cristià, & Dupoux, 2011). If the frequency of the input, rather than its mere presence versus absence, was indeed influencing perceptual attunement, infants' response to the frequent contrast should be stronger compared to their response to the infrequent contrast.

6 Assessing the input: Phoneme frequencies in Japanese infant-directed speech

After focusing on perceptual attunement to native and non-native contrasts (Chapter 4) and the contributions of input frequency on this process (Chapter 5), we aimed at characterizing the input itself. That the input infants are exposed to (infant-directed speech, IDS) differs from adult-directed speech (ADS) has been demonstrated repeatedly, for instance at the semantic, syntactic, and phonetic level (cf. Soderstrom, 2007). Only few studies have looked at differences between IDS and ADS at the level of phoneme frequencies, however. These studies of differences between IDS and ADS in Korean and English suggested some systematic differences across registers, for instance a higher frequency of phonemes

that are produced early across languages in IDS (Lee & Davis, 2010; Lee, Davis, & MacNeilage, 2008).

In **Chapter 6**, we followed up on these findings by focusing on differences between phoneme frequencies in Japanese IDS and ADS, and evaluating these findings in comparison to the previous literature. In addition to assessing whether or not converging evidence for differences between IDS and ADS could be found in another language, Japanese provided an excellent test case for evaluating to what extent differences between IDS and ADS would indeed reflect language-general patterns of early phoneme production. For instance, unlike infants in a large number of languages, Japanese infants have been shown to produce a relatively high amount of dorsal phonemes early on (Boysson-Bardies & Vihman, 1991). Assessing whether differences between IDS and ADS would better reflect language-general of language-specific production tendencies may provide additional insight into early patterns in child language.

7 Initial biases and language-specific changes: the labial-coronal asymmetry

The experiments in **Chapter 7** and **Chapter 8** investigate a topic that requires, but to date lacks a cross-linguistic developmental perspective, namely the language-general origins and the language-specific development of the labial-coronal perceptual asymmetry. While Chapters 4 and 5 have concentrated on the impact of the input on infant perception by meta-analyzing studies on developing native and non-native perception and by investigating the impact of native speech sound frequency on perceptual attunement, Chapter 7 extends its focus to infants' speech sound perception prior to perceptual attunement.

The labial-coronal perceptual asymmetry refers to the finding that listeners are more sensitive to the change from a labial to coronal consonant than from a coronal to a labial consonant. This asymmetry has been suggested to be tied to the special status of coronals in the phonologies of the world as the default place of articulation (Paradis & Prunet, 1991). However, whether or not this asymmetry is indeed based on language-general properties or arises from language-specific experience has been a matter of debate (e.g., Lahiri & Reetz, 2010;

Mitterer, 2011). Since studies on this topic have predominantly been conducted with adult speakers from a similar language background, this debate has reached a deadlock. In Chapter 7, we therefore tested the discrimination of a labial-coronal contrast in four-to-six month-old infants of two markedly different language backgrounds, Dutch and Japanese. This builds a particularly strong test of the language-independent nature of the asymmetry: Not only can we assume that the perception of four-to-six month-old infants has not yet been attuned to native consonant categories (cf. Werker & Tees, 1984), but we also assess infants from two different language backgrounds, for whom predictions would go in opposite directions. While coronal place of articulation is assumed to be the default place of articulation in Dutch, coronals do not act as the default place of articulation in Japanese (Labrune, 2012), illustrated for instance by the higher frequency of dorsal compared to coronal plosives (cf. Chapter 8).

Anticipating the results, Chapter 7 demonstrates a language-general asymmetry in the perception of the labial-coronal contrast such that infants from both language backgrounds were sensitive to the contrast in the direction from labial to coronal place of articulation, but not the other way around. Given the differences between the two languages, however, we subsequently asked whether the insensitivity to the coronal-labial change would remain or change with language exposure. To that end, in Chapter 8 we investigated Dutch and Japanese 18-month-old children's perceptual sensitivities in a word learning task. Previous studies have shown that young Dutch children are indeed insensitive to the coronal-to-labial change (Altvater-Mackensen, Van der Feest, & Fikkert, 2013; Van der Feest & Fikkert, 2006), but no study has investigated the sensitivity to this change in children learning Japanese. Since accounts that derive their predictions from the special status of coronals would predict the same insensitivity towards the coronal-to-dorsal change, we added this change to get a better idea of the pattern of perceptual sensitivities.

If the language-general bias observed in Chapter 7 continues influencing perception, we expected both Dutch and Japanese children to be insensitive to the coronal-to-labial change. Further, if this insensitivity is related to coronals in general rather than to the coronal-to-labial

change in particular, we would expect a comparable insensitivity towards the coronal-to-dorsal change. If, however, language-specific factors influence perception of these changes, we would expect a different pattern of results. Since children at 18 months of age have not only been exposed to the frequency distributions of their language, but also have started building a perceptive and productive vocabulary, which in turn could affect their perceptual sensitivities, the predictions for this case are multifold as is elaborated on in Chapter 8.

8 Outline

This thesis takes a multi-angled view on the process of perceptual attunement. After introducing and reviewing a database on infant vowel discrimination studies in Chapters 2 and 3, a meta-analysis on perceptual attunement is reported in Chapter 4. In particular, the critical age for attunement as well as statistical evidence for the decline in non-native discrimination and the improvement in native discrimination are sought. Chapter 5 assesses the impact of frequency of exposure on perceptual attunement by comparing Dutch infants' discrimination of highly frequent and highly infrequent native vowel contrasts.

Chapter 6 focuses on characteristics of infants' input, more precisely on how phoneme frequencies in Japanese IDS differ from phoneme frequencies in ADS. Finally, the last two chapters focus on the role of early language-general biases (Chapter 7), on phonological development (Chapter 8) by exploring the origins and further development of the labial-coronal perceptual asymmetry. Chapter 9 provides a summary and general discussion of the main results of this thesis.

References

Altvater-Mackensen, N., Van der Feest, S. V. H., & Fikkert, P. (2013). Asymmetries in Early Word Recognition: The Case of Stops and Fricatives. *Language Learning and Development*, 1–30. doi:10.1080/15475441.2013.808954

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. doi:10.1177/00238309030460020601

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

Boysson-Bardies, B. de, & Vihman, M. M. (1991). Adaptation to language: evidence from babbling and first words in four languages. *Language*, *67*, 297–319.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews. Neuroscience*, *5*(11), 831–43. doi:10.1038/nrn1533

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

Kuhl, P. 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*, *31*(255(5044)), 606–608.

Labrune, L. (2012). *The Phonology of Japanese.* Oxford: Oxford University Press.

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

Lee, S. A. S., & Davis, B. L. (2010). Segmental Distribution Patterns of English Infant- and Adult-Directed Speech. *Journal of Child Language*, *37*, 767–791. doi:10.1017/S0305000909009568

Lee, S., Davis, B., & MacNeilage, P. (2008). Segmental properties of input to infants: A study of Korean. *Journal of Child Language*, *35*, 591–617. doi:10.1017/S0305000908008684

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–11. doi:10.1016/S0010-0277(01)00157-3

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

Minagawa-Kawai, Y., Cristià, A., & Dupoux, E. (2011). Cerebral lateralization and early speech acquisition: a developmental scenario. *Developmental Cognitive Neuroscience*, *1*(3), 217–232. doi:10.1016/j.dcn.2011.03.005

Mitterer, H. (2011). The Mental Lexicon Is Fully Specified: Evidence From Eye-Tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 469–513. doi:10.1037/a0020989

Paradis, C., & Prunet, J.-F. (Eds.) . (1991). *The special status of coronals: internal and external evidence. (Phonetics and Phonology 2)*. San Diego: Academic Press.

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

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

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), 965–76. doi:10.1111/j.1467-8624.2012.01740.x

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), 874–87. doi:10.1111/j.1467-7687.2009.00829.x

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. doi:10.1016/j.dr.2007.06.002

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. doi:10.1038/41102

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., & Kohno, M. (1994). Developmental changes in perceptual discrimination of non-native speech contrasts by Japanese infants -- on discrimination of American English /r, l/ and /w, y/ --. *Technical Report of IEICE, SP-94-31*.

Van der Feest, S. V. H., & Fikkert, P. (2006). *Learning to represent phonological contrasts*. Manuscript submitted for publication.

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. doi:10.1016/S0163-6383(84)80022-3

Community-augmented meta-analyses: The example of infant vowel discrimination

Chapter 2

Based on:

Tsuji, S., & Cristia, A. (2014). *Community-augmented meta-analyses: The example of infant vowel discrimination*. Manuscript in preparation.

Abstract

While the importance of going beyond individual reports is increasingly discussed in the psychological sciences, open repositories remain rare. We present a new concept: community-augmented meta-analyses, which profit from online resources and a distribution of work in order to facilitate cumulative knowledge. We exemplify this approach through InPhonDB, a database of infant vowel discrimination. A meta-analysis of extant behavioral data suggests that there may be some bias in reporting, and that results gathered with different methods may not be comparable, two empirical considerations that are crucial when trying to understand replicability in a psychological subfield. Targeted analyses confirm a theoretical assumption for which no direct evidence had been gathered: linguistic, and not acoustic, distance predicts effect size. We conclude by highlighting some strengths and limitations of community-augmented meta-analyses for psychological research in general.

1 Introduction

It has been suggested that a proportion of scientific findings reported are false positives due to data fabrication, data selection, or even statistical oversights (Vul, Harris, Winkielman, & Pashler, 2009; Simonsohn, 2013). Moreover, under-reporting of null findings could lead to the overestimation of the strength and presence of certain effects (e.g., Rosenthal, 1979). Finally, there is growing discomfort with exclusive reliance on null hypothesis testing resulting in potentially arbitrary dichotomous decisions as to whether an effect is "present" or "absent" (cf. Hentschke & Stüttgen, 2011). These questions have recently attracted considerable attention particularly within the field of psychological science, resulting in a number of publications attempting to raise awareness and propose solutions to these problems (for instance, in addition to those above Simmons, Nelson, & Simonsohn, 2011). The present manuscript falls in the latter category. We present a simple, yet novel, tool that allows researchers within a scientific domain to accumulate and evaluate the state of the field in an efficient and transparent way.

Today there are two common tools for knowledge accumulation, namely open repositories and meta-analyses. The former include method-specific (e.g., BrainMap.org; Gibbons, 1992) and replication repositories (e.g., PsychFileDrawer.org; Spellmann, 2012), both of which can remain up to date since they are open and updated by users. Such repositories, however, tend to be overly broad in coverage, and often insufficiently detailed along cognitively relevant dimensions. In contrast, meta-analyses more readily speak to our psychological interests, and have the further advantage that results across diverse methodologies are expressed in a common metric, standardized effect sizes (Lipsey & Wilson, 2001). Unlike repositories, meta-analyses are private, static endeavors. The load of the work is done by the meta-analysts, who hold all decision power in terms of inclusion and mediator coding. Moreover, the resulting table is crystalized at publication, and it ages thereafter.

We propose that these two tools can be merged to create topic-oriented, community-augmented meta-analyses, which are accessible to the whole research community, and can become independent from the

originators. The optimal implementation for augmentation and maintenance would be then decided for the specific community, taking into account the potential contributors and users. For example, entry of new data points could be limited to registered contributors, open to everyone, or – an intermediate option – a small group of rotating volunteers could look over anonymous submissions to ensure the stability of the database. The database itself can be hosted at a private institution, on semi-open platforms, and on decentralized systems such as github.

We exemplify this concept with one such database, InPhonDB (sites.google.com/site/InPhonDB) which is currently focused on "infant vowel discrimination". The last 50 years have seen a multitude of work on how infants process sound contrasts, including how they are represented in the young brain and how these representations are shaped by experience and development (e.g., Gervain & Mehler, 2010). The topic of infant speech perception is not only interesting in itself, but also an ideal example for what psychological science stands to gain by incorporating topic-oriented, community-augmented meta-analyses. Indeed, infants are notoriously difficult to study; the manipulations that can be used are limited and indirect; and the resulting data are noisy. Faced with inherently "messy" data, specialists (authors, reviewers, editors) could be tempted to "clean up" the panorama through selective reporting and publication. Moreover, in search of improved signal-to-noise ratios to study a psychologically defined phenomenon, researchers may therefore explore variations of methods and paradigms, sometimes forgetting that different methodologies might pick up on different signals. Finally, they may zoom in on specific questions without checking underlying theoretical assumptions. Thus, infant vowel discrimination provided us with an ideal opportunity to exemplify the potential of a meta-analysis for revealing potential biases, provide an overview across methods, and illuminate areas for further conceptual work.

Data in InPhonDB were inputted by the authors, who also established clear entry criteria, and made analyses scripts openly available. New records can be submitted through method-specific submission forms, which are checked by the site managers prior to acceptance. The final database contains 97 columns, including dimensions encoding the study and record identity, methods, and results.

Contributors, however, only need to fill in key information (which takes approximately 15 minutes), as other fields are completed by the managers or calculated via scripts. This facilitates the task for contributors and ensures standards for users. Users can also contribute recoding of extant data (e.g., an addition of a dimension that was initially overlooked), further analysis scripts, and questions through an associated blog. Thus, the time cost from all actors (site managers, contributors, users) is relatively small, and yet – as we will show – a great deal stands to be gained by allowing a broad and statistically informed view of a subfield.

We exemplify the power of such databases through three findings uniquely afforded to us by InPhonDB. First, we inspected the distribution of effect size as a function of precision. As precision increases, the effect size recovered is probably closer from the true effect size of the distribution. If all findings were reported, measurements with low precision should be as far from the true effect size in the positive direction as in the negative direction, while an asymmetrical distribution would indicate a bias in reported findings. In contrast, if experts are tempted to "clean up" the noisy infant data, asymmetries in this distribution could be observed.

Second, findings from different methods and paradigms are implicitly assumed to be comparable. One often reads statements like the following: "Eilers, Wilson, and Moore (1977) report that [...] 12–14-month-olds fail to discriminate [a given contrast], while 2-month-olds [...] succeeded [...] according to Levitt et al. (1988)" (Cristia, McGuire, Seidl, & Francis, 2011). And yet, these two datapoints come from studies using radically different techniques, one requiring the same infants to produce or withhold headturns contingent on a syllable change, and the other monitoring between-group differences in pacifier sucking rate depending on syllables presented. Are results from such conceptually diverse methods effectively comparable?

Finally, we illustrate how users could gain conceptual insights with minimal effort, by coding studies along a cognitively relevant dimension and analyzing effect sizes in the subset of papers where multiple conditions differ along that dimension. For this example, we investigated the impact of contrast size. If infants are sensitive to the extent to which a pair of vowels differ, there should be larger effect sizes

elicited by a large contrast (e.g., the vowels in *shop-sheep*) compared to a small contrast (e.g., those in *ship-sheep*). To quantify this "difference", however, the subfield needs to agree on a metric. We compare two of them by putting into subsets papers where records varied along either or both metrics. The *acoustic* distance relies on a description of stimuli's physical properties that does not necessitate much linguistic knowledge. Alternatively, infants may employ *linguistic* representations involving phonological features (cf. White & Morgan, 2008, on word recognition).

2 Methods

2.1 Search protocol

We compiled a list of journal articles, manuscripts, and theses (the conjunction of which we call "papers") through (1) a search on scholar.google.com with the keyword combination "{infant|infancy} & {vowel|speech sound|syllable} & discrimination" in English, French, German, Japanese, and Spanish; (2) an automatic alert was set up; (3) scouring those references; (4) communicating with researchers active in the subfield.

The search sample was narrowed down to 51 papers based on the following inclusion criteria: [1] the study contained data on infants aged 15 months or younger; [2] discrimination was the key component of the task; [3] discrimination depended solely on auditory vowel quality or quantity (if a visual stimulus was presented, it was only for the purpose of indirectly measuring infants' attention by looking time, or in order to distract infants with unsystematic stimuli). At present, only the 34 papers using behavioral methods have been exhaustively entered. The behavioral methods included central fixation (CF), Conditioned HeadTurn (CHT), Headturn Preference Paradigm (HPP), High-Amplitude Sucking (HAS), Anticipatory Eye-Movements (AEM), and Stimulus Alternation (SA). For introductions to these methods, readers are referred to Jusczyk (1997), McMurray and Aslin (2004), and Werker, Cohen, Lloyd, Casasola, and Stager (1998). One additional study using heart rate, another using amount of limb movement, and 15 studies using neurophysiological correlates were excluded from the analyses below to keep the final sample homogeneous. Most of the papers have been published in journals

(*N* = 28; between 1973 and the present); others are manuscripts in review or in preparation (*N = 4*), progress reports (*N = 1*), and theses (*N = 1*). They contain a total of 181 records that are candidates for effect size estimations. A *record* is a result from a group of infants being tested with a vowel contrast.

2.2 Coding and calculation of effect sizes

Appropriate formulas for effect size estimation were selected from Lipsey and Wilson (2001). Some of the effect sizes arose from methods yielding a single performance measure that is then compared to a chance level. This pertained percent correct (chance level 50%), d' (chance level 0) and A' (chance level 0.5), among CHT measures, the proportion of trials anticipated in AEM (chance level 0.5), and the attention differences in SA (chance level 0.5). Other effect sizes were based on between-participant-measures (e.g., most HAS studies have an experimental group exposed to a vowel change, and a control group that hears no change). Yet others arose from repeated measures within participants (e.g., habituated versus novel stimulus in CF and HPP). In the latter case, correlations between repeated measures were obtained by personal communication with the authors for 63 CF records, and their distribution was used to impute correlation values to the remaining 22 records. To this end, we used the *impute* function in the Hmisc package (Harrell, 2013) in R (R Core Team, 2012), which samples randomly from the available observations. Among HPP studies, 22 correlations were made available to us, which were used to impute the 15 remaining values.

There was sufficient information to calculate effect sizes for 138 records (76%). Three of these were eventually excluded because they were more than 3 standard deviations away from the mean effect size.

2.3 Moderators and statistical analyses

For the present analyses, only two potential moderators are relevant:

a) Method. The most prevalent method in our sample was CF, with 48 effect sizes. Other common methods were CHT (45 effect sizes), HPP

(22 effect sizes), and HAS (14 effect sizes). AEM contributed 4 effect sizes, and SA contributed 3 effect sizes.

   b) Contrast size. In order to assess the predictive value of acoustic distance, we used the first and second formant values (F1 and F2), which, as broadly agreed upon descriptors of the spectral characteristics of vowels, were reported in over 80% of the records. A single acoustic distance was calculated as the Euclidean distance between the two vowels in Bark-based F1 x F2 space. We avoided the "apples and oranges" problem by including only the 11 papers with multiple records varying along this acoustic distance (Albareda-Castellot, Pons, & Sebastián-Gallés, 2011; Bohn & Polka, 2001; Bosch & Sebastián-Gallés, 2003; Mazuka, Hasegawa, & Tsuji, 2013; Phan & Houston, 2008: Polka & Bohn, 1996, 2011; Sebastián-Gallés & Bosch, 2009; Swoboda, Morse, & Leavitt, 1976; Trehub, 1973), for a total of 91 records.

   Contrasts were linguistically coded using 3 levels of height (i.e., high, mid, low); 3 levels of backness for high and mid vowels (i.e., front, central, back) and 2 for low vowels; 2 levels of roundness (rounded, unrounded); 2 levels of nasality (nasal, oral); and 2 levels of tenseness (tense, lax). A single linguistic distance was then computed as the sum of feature changes (without assuming feature redundancy - i.e., [i-u] represents a 3 feature difference: 2 levels of backness plus roundness). As above, we only included in this analysis 62 records drawn from 8 papers reporting multiple linguistic distance conditions (Albareda-Castellot et al., 2011; Benders, submitted; Bohn & Polka, 2001; Mazuka et al., submitted; Phan & Houston, 2008: Sato, Sogabe, & Mazuka, 2010; Sebastián-Gallés & Bosch, 2009; Trehub, 1973). Notice that the two lists overlap although not perfectly, as some studies included contrasts of the same linguistic distance but different spectral distance (e.g., Polka & Bohn, 2011, reports multiple vowel contrasts, but they all span a linguistic distance of 1).

   Analyses were performed with the meta (Schwarzer, 2012) and metafor (Viechtbauer, 2010) packages in R (R Core Team, 2012).

3 Results and targeted discussions

   The mean weighted effect size under a random effect model without moderator variables (k = 135) was ES = 0.617 (*SE* = 0.055), with the lower

bound of the 95% confidence interval CI_L = 0.513, and the higher bound CI_H= 0.722. The estimated total amount of heterogeneity was τ^2 = 0.262 (estimated by restricted maximum likelihood, REML). The ratio of this amount to total variance was I^2 = 75.94%. A Cochran's Q-test rejected the null hypothesis of no sample heterogeneity [$Q(134)$ = 440.54, p < .0001], which is consistent with the idea that there are moderators leading to heterogeneous results.

3.1 Reporting bias

The funnel plot of standard error as a function of effect size including the 138 records (left panel of Figure 1; the 3 CHT outliers are noted with open symbols) shows an underrepresentation of data points in the lower left corner: there are fewer records with a high standard error and small or negative effect size, than records with an equally large standard error and a large positive effect size. Linear regressions following Egger, Smith, Schneider, and Minder (1997) confirm the presence of a significant asymmetry in the dataset as a whole [$t(133)$ = 9.312, p = 4 x 10^{-16}], as well as within the 4 methods with more data points [CF $t(46)$ = 2.68; CHT $t(42)$ = 3.5; HAS $t(12)$ = 2.52; HPP $t(20)$ = 2.96].

Funnel plot asymmetry could be due to human biases in data collection, selection, submission, and publication. That is authors may choose not to report a study with an "unexpected" effect size when the sample is small; or they may continue adding observations until either the effect size becomes more moderate, and/or the standard error of the effect size becomes smaller. Additionally, studies with small or negative effect sizes could meet with lower acceptance rates at the publication stage as reviewers and editors might more easily accept a large positive effect size than a small or negative one. These kinds of practices inflate Type I errors (Gupta & Stopfer, 2011), and fit the criticisms often leveled against psychological science.
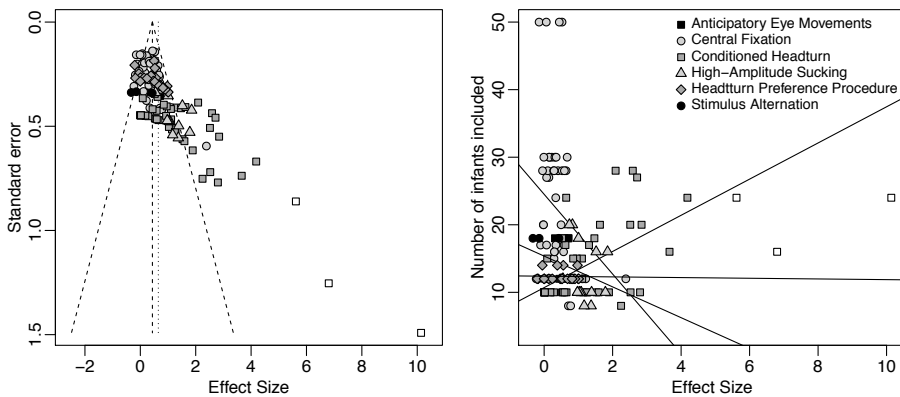
Figure 1: Funnel plot of precision (standard error of the effect size, on the left panel, and sample size in the right panel) as a function of effect size; different symbols represent different methods (see legend). On the left panel, the vertical line indicates the model estimate along with a +- 1.96 SE pseudo confidence interval. On the right panel, lines indicate simple regressions within method. A similar plot based on a subset of these data has been published in Tsuji and Cristia (2013; Chapter 4 of this dissertation).

But this is not the only interpretation (cf. Lau, Ioannidis, Terrin, Schmid & Olkin 2006). To begin with, since standard error of effect size for the measures under study actually depends both on sample size and effect size, studies with larger effect sizes will necessarily have larger standard errors. Nonetheless, lack of numeric independence from effect size is not the only factor, given that there are evident relationships between effect size and sample size in 3 of the 4 methods with a sizable N (see right panel in Figure 1). This still leaves open a host of possible sources for the non-independence between precision and effect size, such as heterogeneity across studies with different methods and sample sizes.

Particularly in view of the difficulty of measuring the presence and source of bias, a more realistic view of actual effect sizes could be gained by compelling authors to announce their results publicly *regardless of the effect size sign*. Repositories like InPhonDB could be a good venue for this, as they would only require an additional 15 minutes from a study's author for the key data to be publicly available.

3.2 Comparability across methods

A second analysis introduced method as a moderator variable. Heterogeneity between methods was significant when all methods were declared [k = 129, $QM(5) = 70.98$, $p < .0001$], and when the methods with only a few studies were removed [k = 124, $QM(3) = 64.13$, $p < .0001$].
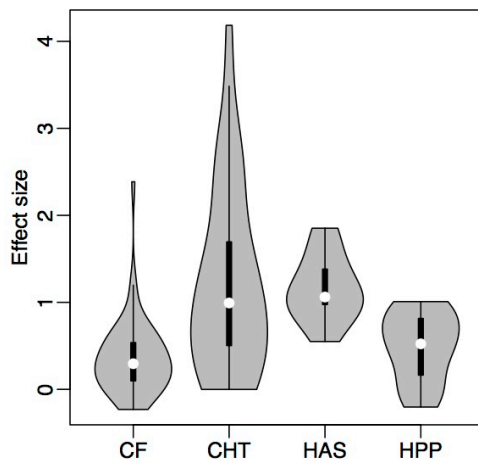
Figure 2: Violin plots of effect sizes as a function of method: CF stands for Central Fixation-Stimulus Alternation, CHT for Conditioned Headturn including also Anticipatory Eye Movements, HAS for High Amplitude Sucking, HPP for Headturn Preference Procedure. The white dot and black vertical line within each violin represent the median and the 25th/75th percentiles. The height represents the data range after outlier exclusion, and the width represents the probability density.

It could be argued that different methods employ different effect size formulas, and thus they cannot be directly compared (see e.g., Morris & DeShon, 2002). We therefore repeated this analysis restricting the sample to CF and HPP, which are both based on looking-time preference in a repeated measures, habituation-dishabituation design. The null hypothesis, that methods did not account for substantial heterogeneity in effect sizes, failed to be rejected by a narrow margin [k = 68, $QM(1) = 3.65$, $p = .056$]. Figure 2 shows violin plots for effect sizes as a function of method.

Although rarely explicitly addressed in the discussion of infant discrimination outcomes, our results suggest that method impacts effect sizes, cautioning us when comparing results gained by different methods.

To follow up on this central methodological finding, the actual impact of method should be experimentally investigated by testing infants on different methods while controlling for all other factors.

3.3 Sensitivity to contrast size

Given the results above, we controlled for method. Methods with few datapoints were merged with the conceptually most similar one: AEM with CHT (they are both forced choice, one reinforces headturn, the other gaze direction), SA with CF (which are only marginally different). A first mixed effects model for meta-analyses assessed the predictive value of spectral distance, controlling for contrast-coded method (CF-SA, AEM-CHT, HAS, HPP). Significant residual heterogeneity remained [$QE(86) = 170.903$, $p < 0.001$], and the Q-test of moderators was significant [$QM(4) = 27.116$, $p < 0.001$]. While estimates for methods were significant (with CHT and HAS departing from the baseline CF-SA), that for spectral distance was not ($ß = -0.021$, SE = 0.031; $z = -.671$, $p = 0.5$; notice the direction is opposite to predictions).

Using a distance measure based on phonological features yielded a different picture. As in the previous analysis, both moderator [$QM(4) = 66.14$, $p < 0.001$] and residual [$QE(57) = 88.251$, $p = 0.005$] heterogeneity were significant, and the same two methods had significant estimates. However, unlike spectral distance, linguistic distance did predict a significant proportion of the variance in the expected direction ($ß = 0.138$, SE = 0.053; $z = 2.596$, $p = 0.009$; see Figure 3). To directly compare the two distance measures, we ran a third model on the subset of 50 records that had been selected for both analyses, and declared the two distances (in addition to controlling for method). The linguistic distance predictor remained significant and virtually unchanged ($ß = 0.142$, SE = 0.063; $z = 2.256$, $p = 0.02$).
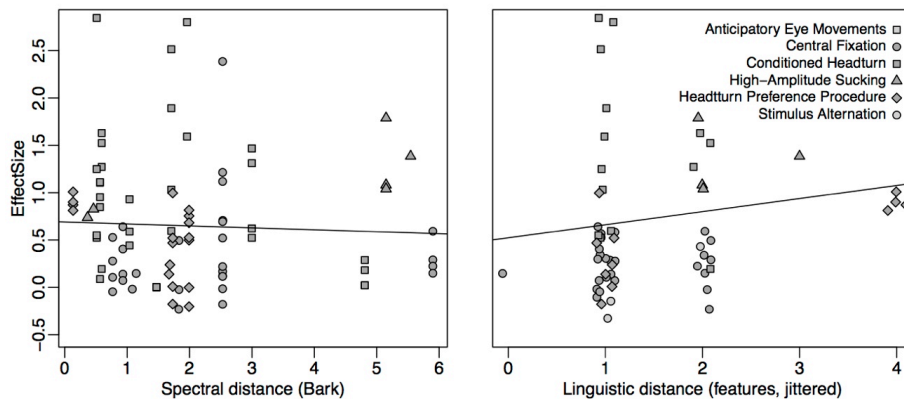
Figure 3: Effect size as a function of spectral (left panel) and linguistic (right panel) distance. Different symbols represent different method, as shown in the legend. Lines indicate meta-analytic regression of effect size by spectral or linguistic distance fitted to the relevant set of points. Notice that points have been jittered along the linguistic distance dimension solely for the purposes of visual inspection, but not for the actual regression.

Thus, these meta-analytic results on sound discrimination suggest that even in such simple tasks, infants' perception may be shaped by vowel differences that are best captured with a linguistic, rather than an acoustic description. Whether infants' perceptual sensitivity is better represented by *acoustic* or *linguistic* measures of distance has been a central question of the infant speech perception literature ever since the seminal study of Eimas, Siqueland, Jusczyk, and Vigorito (1971). In this study, infants' were shown to discriminate two consonants with a given acoustic distance better when they came from two different linguistic categories than when they came from the same category. The present fascinating finding on vowel discrimination should inspire a direct experimental comparison where acoustic and linguistic distances are independently manipulated.

4 General discussion

Through the example of infant vowel discrimination, we have illustrated the power of meta-analyzable databases that are openly available for use. Indeed, the current meta-analysis finds evidence for

non-independence of effect size from precision, as well as for an impact of method on effect sizes. We argue that InPhonDB can help improve the current methodological state of affairs in the field of infant speech perception by its dynamic and open-access nature. First, the subfield can combat under-reporting of null or negative results by facilitating the public availability of both published and unpublished results. Second, the publication of positive results with high standard errors can be constrained by establishing clear guidelines for appropriate sample sizes based on power calculations on suitable, method-specific subsets of InPhonDB. Third, explicit consideration and discussion of the impact of method will lead to a more accurate view of infant discrimination performance.

Our example analyses also aimed to demonstrate the potential such databases hold to provide conceptual insights. Our results on the predictive value of phonological features ideally motivate targeted and controlled experiments comparing linguistic versus spectral predictors of infant speech sound discrimination. Teasing apart the effects of spectral and linguistic predictors would have been hard to assess in a qualitative review, in which continuous measures are difficult to capture and compare. This exercise further illustrates the potential of databases like InPhonDB: Users can, with little effort, investigate the importance of a dimension by creating a subset from all entered studies that contains the most relevant studies to their specific question.

There are, undoubtedly, limitations to community-augmented, meta-analyzable databases. First and foremost, there is an initial cost in the construction. To attract novel contributors and analyzers, one or a few people may need to invest by inputting all previous work, to make the database attractive. At that point, there is also a cost/gain balance to be reflected upon in terms of the details requested from subsequent contributors. The more information one asks, the richer the database potentially is, but the more time volunteers will have to spend to submit new entries. There is also a small maintenance cost, in that new entries should be checked for consistency with previous ones, and/or novel contributors should be registered. These responsibilities could be even extended to identifying new relevant papers, contacting the authors, and ensuring that the entry is made (by the author, or else by the site

managers). At this point, we cannot prove statistically that these costs will outweigh the potential gains, a question that will have to be evaluated in the future.

Others may find that such repositories are insufficient, and more comprehensive ones are desirable (e.g., with outcomes encoded through the individual participants' results files). Nonetheless, a more comprehensive repository will require considerably more dedication in terms of data assembly, standardization and maintenance for both contributors and users, and it is an open question to what extent the additional benefits will balance out these larger costs. Preliminary answers for many methodological and conceptual questions can be found by using our compact database, as we demonstrated above. An additional advantage of such compact databases over comprehensive repositories is that they are backwards-compatible. For instance, infant discrimination studies have a relatively long history, with our first entry dating back to 1973, and it is virtually impossible to recover information much beyond effect sizes and method descriptors for "old" data points. Furthermore, even if community-augmented meta-analyses are eventually phased out, they could be an important stepping stone towards a more comprehensive solution.

Finally, there are certainly restrictions to the technique of meta-analysis itself: a meta-analysis is only as good as the data it contains together with the analysis applied. Over this backdrop, we contend that a community-augmented format can lead to a more comprehensive and unbiased accumulation of data, which moreover remains accessible for alternative and/or newer analyses (e.g., Pfeiffer, Bertram, & Ioannidis, 2011).

In conclusion, we propose that community-augmented and meta-analyzable databases are a low-effort and timely way for a field to keep up to date with its methodological and theoretical state, and a first step towards estimating, and controlling for, both weaknesses and limitations in experimental research.

Acknowledgements

References

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

Benders, T. (submitted). Learning phonemes from multiple auditory cues: Dutch infants' language input and perception.

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*, 217–243. doi: 10.1177/00238309030460020801

Bohn, O.-S., & Polka, L. (2001). Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *Journal of the Acoustical Society of America, 110*(1), 504-515. doi: 10.1121/1.1380415

Cristia, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics, 39*(3), 388-402. doi: 10.1016/j.wocn.2011.02.004

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634. doi:10.1136/bmj.315.7109.629

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental Changes in Speech Discrimination in Infants. *Journal of Speech and Hearing Research, 20*, 766-780.

Eimas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(3968), 303-306. doi: 10.1126/science.171.3968.303

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology, 61*, 191-218. doi: 10.1146/annurev.psych.093008.100408

Gibbons, A. (1992). Databasing the brain. *Science, 258*(5090), 1872-1873. doi: *10.1126/science.1470907*

Gupta, N., & Stopfer, M. (2011). Negative results need airing too. *Nature, 470,* 39. doi:10.1038/470039a

Harrell, F. E. (2013). Harrell Miscellaneous. R package http://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf

Hentschke, H., & Stüttgen, M. C. (2011). Computation of measures of effect size for neuroscience data sets. *European Journal of Neuroscience, 34*(12), 1887-1894.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Lau, J., Ioannidis, J. P., Terrin, N., Schmid, C. H., & Olkin, I. (2006). Evidence based medicine: the case of the misleading funnel plot. *BMJ: British Medical Journal*, *333*(7568), 597-600.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Maddieson, Ian. (1984). *Patterns of sounds*. Cambridge studies in speech science and communication. Cambridge: Cambridge University Press.

Mazuka, R., Hasegawa, M., & Tsuji, S. (2013). Development of non-native vowel discrimination: Improvement without exposure. *Developmental Psychobiology*. Advance online publication. doi: 10.1002/dev.21193

McMurray, B., & Aslin, R. N. (2004). Anticipatory Eye Movements Reveal Infants' Auditory and Visual Categories. *Infancy, 6*(2), 203-229. doi: 10.1207/s15327078in0602_4

Morris, S. B., & DeShon, R. P. (2002). Combining Effect Size Estimates in Meta-Analysis With Repeated Measures and Independent-Groups Designs. *Psychological Methods, 7*(1), 105-125. doi: 10.1037//1082-989X.7.1.105

Phan, J., & Houston, D. M. (2008). Infant Dialect Discrimination. *Research on spoken language processing Progress Report, 29.* Indiana University.

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

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

Pfeiffer T, Bertram L, Ioannidis JPA (2011) Quantifying Selective Reporting and the Proteus Phenomenon for Multiple Datasets with Similar Bias. *PLoS ONE 6*(3): e18362. doi:10.1371/journal.pone.0018362

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641. doi: 10.1037/0033-2909.86.3.638

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

Schwarzer, G. (2012). meta: Meta-Analysis with R. R package version 2.1-3. Retrieved from http://CRAN.R-project.org/package=meta

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science, 22*(11), 1359-1366. doi: 10.1177/0956797611417632

Simonsohn, U. (2013). Just Post It: The Lesson From Two Cases of Fabricated Data Detected by Statistics Alone. *Psychological Science, 24*(10), 1875-1888. doi:10.1177/0956797613480366

Spellman, B. (2012). Introduction to the Special Section: Data, Data, Everywhere . . . Especially in My File Drawer. *Perspectives on Psychological Science, 7* (1), 58-59. doi: 10.1177/1745691611432124

Swoboda, P.J., Morse, P.A., & Leavitt, L.A. (1976). Continuous Vowel Discrimination in Normal and At Risk Infants. *Child Development 47*(2), 459-465.

Trehub, S. E. (1973). Infants Sensitivity to Vowel and Tonal Contrasts. *Developmental Psychology, 9*(1), 91-96.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1-48. URL http://www.jstatsoft.org/v36/i03/

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. *Perspectives on Psychological Science, 4*(3), 274-290. doi: 10.1111/j.1745-6924.2009.01125.x

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology, 34*(6), 1289-1309. doi: 10.1037/0012-1649.34.6.1289

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59*, 114-132. doi: 10.1016/j.jml.2008.03.001

Fifty years of infant vowel discrimination research: what have we learned?

Chapter 3

Based on:

Tsuji, S., & Cristia, A. (2013). Fifty years of infant vowel discrimination research: what have we learned? *Journal of the Phonetic Society of Japan, 17*(3), 1-11.

Abstract

This article provides the first comprehensive qualitative review of infant vowel discrimination, including results from behavioral, psychophysiological, and neuroimaging methods in infants during their first year of life. Even prior to extensive language experience, discrimination of vowel contrasts is robust in the face of concurrent variation in other dimensions (e.g., pitch), although infants do not discriminate all vowel contrasts equally well. Moreover, while infants are sensitive to within-category variation, discrimination is not a linear function of physical distance between sounds. There is clear evidence for experience-dependent changes in behavioral and neural responses, with some variation regarding the precise age at which language exposure warps perception. Finally, differences and similarities in vowel perception in infants at risk for language impairments and bilingual infants, compared to typically-developing monolinguals, are reviewed.

1 Introduction

The study of infant vowel perception over the last decades has spawned 60 papers, among which a diverse range of insights on early perceptual abilities and their development become evident (see Tsuji & Cristia, 2014, for details on how these papers were selected). The aim of the present article is, therefore, to review these experimental results to provide a broad and current picture of infant vowel perception.

This picture is organized into four key categories, bearing respectively on methodological considerations; the format of representations prior to, or independent from, language exposure; changes (or lack thereof) with development; and a comparison of "standard" (monolingual, typically developing) infants with other populations. Within each category, we group studies as they relate to key theoretical and empirical questions. Each of these subsections begins with one paragraph stating the general question, and the overall answer that has been achieved. This article can be read from beginning to end, to gain a complete overview of the field, or it can be used as a reference piece, to consult the literature on one specific topic.

2 Methodological considerations

2.1 Characteristics of vowels

Vowels can differ phonemically in quality or quantity (length). Vowel quality differences are typically represented by their articulatory or acoustic characteristics. The articulatory features can be specified along the following dimensions: *Backness* represents the horizontal tongue position relative to the back of the mouth; *height* represents the vertical position of the tongue relative to the roof of the mouth; *roundedness* encodes lip rounding; and *nasalization* indicates whether air flows through the nose during vowel articulation. *Tenseness* has been difficult to define articulatorily, but it is generally associated with fronting and raising.

As to acoustic characteristics, vowel contrasts are captured through the first three formants (commonly referred to as F1, F2, F3), in combination with their duration (Hillenbrand, Getty, Clark, & Wheeler,

1995). Changes in formant values reflect different positions of lips, tongue and jaw and are thus linked to articulation. In particular, F1 is associated with height, F2 with backness, and F3 with roundedness.

2.2 Methods for assessing infant speech sound perception

For reasons of space, we cannot provide a full methodological review. Nonetheless, it suffices for our goals to state that a variety of behavioral methods have been used, including High-amplitude sucking (HAS; for details, see Jusczyk 1985); Central Fixation (CF; Werker, Cohen, Lloyd, Casasola, & Stager, 1998); Headturn Preference Paradigm (HPP; Kemler Nelson, Jusczyk, Mandel, Myers, Turk, & Gerken, 1995); Conditioned Head-Turn (CHT; Werker, Polka, & Pegg, 1997); Anticipatory Eye-Movement (AEM; Albareda-Castellot, Pons, & Sebastián-Gallés, 2010); and, more rarely, infant body movement (Weir & Lamb, 1990). Psychophysiological methods used include heart rate (Clarkson & Berg, 1983), electroencephalography (EEG; Cheour, Leppänen, & Kraus, 2000), and magnetic encephalography (MEG; Kujala et al., 2004). Neuroimaging methods are so far limited to Near-infrared spectroscopy (NIRS; Minagawa-Kawai, Mori, Hebden, & Dupoux, 2008). These methods have been combined with a variety of paradigms, including simply spontaneous responses to specific stimuli; responses to novelty (after familiarization for a fixed duration or habituation; or the detection of an infrequent sound over the background of more frequent ones); and trained change detection or classification.

The possibility that different methods and paradigms can lead to different results is commonly recognized, because some require an overt response and attention to the task but not others, and some rely on instinctive responses whereas others rely on trained ones. To provide an integrated picture across all findings, we do not group studies by method, but instead simply indicate key methodological choices. Readers can download a spreadsheet containing all methodological details (including infant age and language) from sites.google.com/site/InPhonDB.

2.3 How do experimental manipulations affect discrimination?

Not only do methods differ between studies, but even within one method, stimuli can be presented under different conditions, for instance with different inter-stimulus intervals (ISI). Several studies have explored the impact of such differences by combining the study of vowel characteristics with temporal aspects of stimulus presentation. Briefly, longer ISI and longer stimuli promote better discrimination.

Clarkson and Berg (1983) presented newborns with changes in vowel quality between [a] and [i] tokens that were either separated by 0 or 500 ms of silence, and measured their cardiac response. Newborns only reacted to changes in the latter case, suggesting that silences were necessary to avoid loading infants' short term memory. This result was replicated by Byrne, Miller, and Hondas (1994) with 3- and 6-month-old infants and using variants of the diphthong [ai]. The role of ISI in the more specific context of informational masking was further explored by Cowan, Suomi, and Morse (1982), who documented that infants reacted to a vowel change in a backward-masking condition (e.g., a change from a vowel pair like [a, a] to a different one like [ɛ, a]) if the ISI was 300 ms, but not if it was 150 ms. In consonance with adults' results, infants succeeded in a forward-masking condition (e.g., a change from [a, a] to [a, ɛ]) at both ISI.

ISI is also acutely relevant when measuring change detection using EEG, at least in newborns. In Leppänen, Pikho, Eklund, and Lyytinen (1999), larger responses to both a repeated [ka:] and an infrequent [ka] were registered with 855 ms compared to 425 ms ISI.

Finally, the comparison of two studies assessing 2-month-old infants suggests that sheer stimulus duration may affect discriminability (Swoboda ,Kass, Morse, & Leavitt, 1976; Swoboda, Morse, & Leavitt, 1978). Infants were able to notice within-category contrasts when vowels were 250 ms long, but not when they were 60 ms long.

2.4 What acoustical and multimodal cues affect infants' vowel representations?

Vowels in natural speech mostly occur within strings of speech sounds, and are connected with visual talker information. This section

shows that infants track and use multiple sources of information, both in the auditory signal and in other modalities.

Bohn and Polka (2001) investigated which aspects of the acoustic signal in CVC syllables infants relied upon in order to distinguish vowel categories. German 6- to 12-month-old infants were tested on their discrimination ability of native vowels in a [dVt] context which was either unmodified or had various kinds of information removed. Infants were equally able to discriminate the original syllables in conditions which preserved both onset and offset formant transitions, and those where only the stable vocalic portion was present. In contrast, onset and offset formant transitions on their own were insufficient for discrimination.

The possibility that infants represent vowels in a modality-independent way has captured researchers' attention. American English-learning infants as young as 2 months tend to look longer at faces with open mouths while hearing [a], and faces with closed, spread lips while hearing [i] (e.g., Kuhl & Meltzoff, 1984). In fact, in terms of mismatch responses, visual and auditory information cuing the large vowel quality contrast [a-i] may have a similar informational value (Bristow et al., 2008). Indeed, EEG revealed that brain responses to a vowel change occurring in a crossmodal condition (i.e., background trials cued through a silent video, test trials cued auditorily) did not differ from those occurring in a unimodal condition (both background and test trials were auditory). Finally, it has been reported that already by 3-5 months of age American English-learning infants tend to approach vowel qualities observed on an audiovideo in their own productions (Kuhl & Meltzoff, 1996), suggesting that audiovisual vowel information can bias infants' vocalizations.

3 The format of representation

3.1 Do vowel quality contrasts have a privileged status?

Vowels in infants' natural input show variability on multiple dimensions. A set of studies demonstrate that infants attend to variation in pitch, talker, and vowel quality. However, their discrimination of vowel quality changes elicits the largest responses and is particularly robust in the face of variation in other dimensions, with syllables as the basic encoding units.

Trehub (1973) reported that 1-4-month-old learning infants in an HAS study discriminated the contrasts in [pa-pi], [ta-ti], [a-i] and [u-i], but not tonal square wave or sine wave contrasts. Similarly, newborns presented with either a change of vowel quality ([æ-i]) or pitch (low-high) responded with increased motor movements to both, but responses to the former were stronger (Weir & Lamb, 1990).

Infants' sensitivity to vowel quality may organize their perception, and this sensitivity may therefore resist competing variation. Most 6-month-old infants tested with CHT generalized from a trained male-spoken [a-i] contrast to tokens with different voice (female, child) and pitch (rising, falling) characteristics with little training (Kuhl, 1979), and some could do so for the more difficult [a-ɔ] contrast (Kuhl, 1983). Even younger infants prioritize vowel quality over other contrasts, as shown with 1- to 4-month-old infants who could discriminate [a-i] both in the presence and in the absence of pitch variation (but had difficulties discriminating between pitch contrasts in the presence of vowel variation, Kuhl & Miller, 1982); and with 2-, 3-, and 6-month-old infants who responded more strongly to a stimulus where both voice and vowel category changed, than when only voice changed (Marean, Werner, & Kuhl, 1992). The primacy of vocalic over pitch changes is even reflected in the ease with which change detection can be localized within the infant brain. In a unique MEG study, Kujala et al. (2004) could localize the neural source for the detection of an infrequent [i:] over the background of repeated steady-pitched [a:] for all of the 10 newborns tested, whereas the detection of an infrequent rising-pitched [a:] could only be localized in 6 of them.

This is not to say that pitch variation is completely orthogonal to vowel discrimination. On the contrary, a CHT study varying pitch height and pitch contour parametrically suggests that discrimination improves when the pitch changes over the course of the vowel, but is not overly high (Trainor & Desjardins, 2002). This appears to ensue from tradeoffs in perceptibility of the formants (which are blurred by high pitch) and increased attention (captured by the changing pitch contour).

Other studies assessed to what extent infants pay attention to the vowel separately from the consonantal context. Miller and Eimas (1979) habituated 2- to 4-month-olds with a pair of syllables and tested them on

a change of vowel ([ba, da] versus [bæ, dæ]) or a recombination ([ba, dæ] versus [bæ, da]). Since the strengths of dishabituation did not differ significantly across conditions, the authors propose that infants may encode the stimuli holistically, a conclusion repeated in Jusczyk and Derrah (1987).

3.2 Can infants perceive within-category variation?

The format of representations has been a key interest in the field. This strand of research suggests that infants are sensitive to within-category variation in vowels, but their sensitivity might be organized in a non-linear way depending on prototypes and category boundaries, as follows.

Early reports had concluded that vowels were not perceived categorically, based on evidence that 2-month-old American English-learning infants in a HAS study were equally able to discriminate within- and between-category contrasts drawn from a continuum between [i] and [ɪ] (Swoboda et al., 1976). However, other work documented greater sensitivity for between- than within-category differences (although acoustic distance is not always matched across the two types of contrasts). In a follow-up with shortened vowels, infants discriminated [i] from [ɪ], but not from the intermediate token (Swoboda et al., 1978). Similarly, EEG reveals that, over the background of a repeated [y], the detection of a categorically different [i] is significant whereas that of an ambiguous [iy] is not, both for infants tested shortly after preterm birth (Cheour-Luhtanen et al., 1996), fullterm birth (Cheour-Luhtanen et al., 1995), and at about 3 months of age (Cheour et al., 1997). Using NIRS, Minagawa-Kawai, Mori, Naoi, and Kojima (2007a) documented that Japanese infants' brain responses to within- versus between-category vowel length contrasts indeed differed at certain points of development (cf. Section 4.2).

Another body of literature suggests that the sensitivity to within-category changes depends on the extent to which vowel exemplars are prototypical of a sound category, as shown by two groups of results. First, infants seem to respond more vigorously to sounds that are prototypical of a given vowel category in newborns, as documented for both native [i, u] and non-native [y, ɯ] through HAS (Aldrigde, Stillman, & Bower, 2001). Other work suggests that within-category sensitivity around prototypical

vowel exemplars is reduced for native language categories in 6-month-olds (as in the CHT studies Grieser & Kuhl, 1989; Kuhl, 1991; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992, cf. Section 4.1), or possibly by birth (in a HAS study by Moon, Lagercrantz, & Kuhl, 2013) using the native [ɪ] and the non-native [y].

Second, an EEG study with 6- to 12-month-old infants exposed to American English found greater and more synchronous brain responses to [i] exemplars with exaggerated formants compared to non-exaggerated exemplars (Zhang et al., 2011). Naturally, it is unclear to what extent these results are due to innate biases towards certain vowel regions, or to the effects of exposure.

3.3 Can infants initially discriminate all linguistic contrasts?

It is often assumed that infants, before language exposure shapes their perception, are able to hear all contrasts that are used in any language. Although more evidence is needed to systematically map out early discrimination abilities, it is clear that infants do not require experience to discriminate vowels, but they do not discriminate all contrasts and all contrast *directions* equally well.

Indeed, while infants sometimes succeed with even small non-native contrasts (e.g., [pa-pã] in 1- to 4-month-olds, Trehub, 1976; [bu:k-by:k] in 4.5-month-old Japanese infants using CF, Mazuka, Hasegawa, & Tsuji, 2013), failures have been recorded even for large contrasts (Nittrouer, 2001 found that 35% of 6- to 14-month-old American English-learning infants failed to discriminate native [a-u] in CHT; Lacerda, 1991 found no statistical evidence for discrimination in a group of Swedish 1- to 6-month-olds for [a-u], or the more difficult [ba-bæ], for which a marginal effect was found in Miller & Eimas, 1979, also HAS; see also Mazuka et al., 2013, for failures to discriminate [bu:k-bo:k] and [bi:k-be:k]).

The Natural Referent Vowel (NRV) framework (Polka & Bohn, 2011) has been proposed to capture a phenomenon called perceptual asymmetry: peripheral vowels (cf. Section 2.1) can act as anchors, rendering discrimination from these vowels towards central ones more difficult than the reverse. Numerous reports are consistent with this prediction, namely: [i-ɪ] in 2-month-olds (Swoboda et al., 1978); [dæt-dɛt], [dut-dʏt] and [dʊt-dʏt] in both Canadian English- and German-learning 6-

8- and 10-12-month-olds (CHT, Polka & Werker, 1994; Polka & Bohn, 1996); and [i-e] in Spanish- and Catalan-learning 4- and 6-month-olds (Pons, Albareda-Castellot, & Sebastián-Gallés, 2012). A more complex pattern of results is documented in Polka and Bohn (2011), where Danish-learning 6- to 12-month-olds were tested on the non-native [dɒt-dʌt], or the native [det-dɛt] or [det-døt]. For the non-native English contrast, both the younger and the older half of infants showed an asymmetry in the predicted direction. For the [det-dɛt] contrast, only the younger half of the infants showed this asymmetry. Finally, for the [det-døt] contrast, again only the younger age-group showed an asymmetry; however, the asymmetry was in the direction opposite from the prediction. Asymmetries may disappear with age and/or experience (e.g., Polka & Werker 1994; Pons et al., 2012).

4 Changes with age: A developmental perspective

4.1 How does vowel quality perception develop?

Infant vowel perception changes during the first year of life. Language exposure affects both the internal organization of sound categories (cf. Section 2.4 and 3.3), and the discrimination between categories. It should be noted that, although a recent study challenges the assumption that newborns' perception is not yet shaped by language (Moon et al., 2013; cf. Section 3.2), there is ample evidence that infants' perception is shaped by language exposure during the first year of life beyond potential *in utero* modulations. Moreover, a meta-analysis places the key age at about 6 months (Tsuji & Cristia, 2013). Additional findings suggest that maturation and experience could enhance certain sensitivities, and lead to differential neural processing.

The process by which exposure to a given language influences speech sound perception has been captured in several models of early speech perception, most prominently the Native Language Magnet model (NLM; Kuhl, 1994; Kuhl et al., 2008), the Perceptual Assimilation Model (PAM; Best, 1994), and the developmental framework for Processing Rich Information from Multidimensional Interactive Representations (PRIMIR; Werker & Curtin, 2005). NLM and PRIMIR in particular assume experience-based perceptual reorganization in the first year, which leads

to decreases in sensitivity to non-native contrasts, and increases in sensitivity to native contrasts. PAM, in contrast, does not bear on the process of attunement, but rather explains how non-native contrasts are processed by reference to native ones.

Early evidence for language-specific vowel perception relied on non-linearities in the detection of within-category changes, demonstrated by infants' better ability to discriminate vowels in the direction from a non-prototypical to a prototypical native exemplar of [i] than vice versa (Grieser & Kuhl, 1989; Kuhl, 1991; cf. Section 3.2). In a subsequent seminal CHT study, Kuhl et al. (1992) suggested that this perceptual pattern was tied to native language perception. American English-learning 6-month-olds failed to detect vowel changes around the prototypical [i] in their language but were sensitive to the same acoustic distances centered around non-native [y], while Swedish infants tested with the same stimuli readily heard such changes around the non-native [i] and missed them around native [y].

Other studies have focused on infants' between-category discrimination. Declines for non-native contrasts have been recorded repeatedly: Polka and Werker (1994) found it for Canadian English-learning infants' performance on non-native [u:-y:] and [ʊ-ʏ] using both CF (between 4 and 6 months) and CHT (between 6-8 and 10-12 months); Mazuka et al. (2013) for Japanese infants' discrimination of German [bu:k-by:k] between 4.5 and 10 months of age; and Jansson-Verkasalo et al. (2010) for Finnish 6- versus 12-month-olds responding to the non-native [ɤ-e], as detected with EEG. A comparable decline has been found for the discrimination of American English dialectal variants of [aɪ] between 7 and 11 months of age (using CF; Phan & Houston, 2008). Other work combines cross-linguistic and cross-sectional data. Indeed, while both Spanish- and Catalan-learning infants were able to discriminate the Catalan contrast [dɛði -deði] at 4 months of age, only the latter did so at 8 months (Bosch & Sebastián-Gallés, 2003, tested with HPP).

A great deal of work documents maintenance for native contrasts, namely EEG-recorded responses to the native [i-y] contrast in Finnish newborns and 3-month-olds (Cheour et al., 1998); [a-i] in 2-, 3-, and 6-month-old American English learners (Marean et al., 1992); [doði-duði] and [deði-duði] in both Catalan- and Spanish-learning infants at 4 and 8

months of age (Sebastián-Gallés & Bosch, 2009); [ø-e] in Finnish learners tested at 6 or 12 months (Jansson-Verkasalo et al., 2010); and [sɑk-sa:k] in Dutch 11- and 15-month-olds (Benders, submitted). Interestingly, a training study showed equally good discrimination of the vowel contrast embedded in [tɪb-teb] after exposure to a monomodal or a bimodal distribution in 8-month-old Canadian English learners, suggesting that when a contrast is discriminated well in the first place, vowel perception is resilient to short-term distributional learning (Pons, Sabourin, Cady, & Werker, 2006a).

Another group of studies has actually found sensitivity can increase with age. In a rare longitudinal CHT study on vowel discrimination, Cardillo (2010) documented an improvement between 7 and 11 months in American English learners' discrimination of Finnish [u-y]. Similarly, Mazuka et al. (2013) reported enhancement for the German-spoken [i:-e:] contrast between 4.5 and 10 months. These improvements could be due simply to maturation, or they could be due to infants learning about their native categories, and later mapping these non-native sounds into two separate native categories, as predicted by PAM.

Finally, a recent trend has been to assess to what extent vowel contrasts lead to left-dominant activations, in the context of the hypothesis that asymmetric processing is a sign of the emergence of brain networks that have specialized for the ambient language. Minagawa-Kawai, Cristia, and Dupoux (2011) provide a comprehensive review of results using NIRS. They conclude that vowel contrasts are processed by a largely bilateral network early on in development, which becomes increasingly left-lateralized with age. The precise age at which this occurs appears to vary across different contrasts. For example, stable left-dominance in temporal brain areas associated with auditory discrimination was evident as early as 7-8 months (but not yet at 3-4 months) in Japanese infants for native [i-ɯ], but not the non-native [u-ɯ] (Minagawa-Kawai, Naoi, Nishijima, Kojima, & Dupoux, 2007b). In contrast, left-dominance was only evident from 11-12 months onwards but not yet between 7-10 months in Japanese infants presented with native [itta-itte] (Sato, Mori, Furuya, Hayashi, Minagawa-Kawai, & Koizumi, 2003). Recent research by Arimitsu, Uchida-Ota, Yagihashi, Kojima, and Watanabe (2011) suggests that left-dominance can already be present in newborns,

albeit in more posterior brain regions and possibly reflecting auditory short-term memory activation.

To our knowledge, no systematic review has sifted through EEG evidence for lateralization. Certainly, spatial localization with EEG cannot be undertaken lightly, and even using source localization methods can be challenging for infant data because accurate localization requires good signal-to-noise ratios as well as precise knowledge of the physical properties of the system in which the signal is traveling, the infant head (e.g., Hämäläinen, Ortiz-Mantilla, & Benasich, 2011; Whittingstall, Stroink, Gates, Connolly, & Finley, 2003). Moreover, until the emergence of high-density EEGs, hemispheric asymmetry descriptions were done at the sensor level, where the source is particularly uncertain given that it depends on the precise characteristics of cortical folding, which changes greatly with age. Not surprisingly, reports of lateralization in such vowel discrimination work are sparse. Only Bristow et al. (2008) have reported localization results using dipole modeling, finding that vowel repetition resulted in greater reconstructed amplitudes in left temporal cortices, and smaller reconstructed amplitudes in right frontal cortices in a group of 2-month-old infants. Future work could exploit this strategy to shed further light on this question, and complement the poor temporal resolution of NIRS.

4.2 How does vowel quantity perception develop?

The perception of vowel quantity contrasts has been studied much less than that of quality, despite the fact that length plays a key phonemic or acoustical role in many languages. An overview of this literature reveals a somewhat mixed pattern of results in terms of infants' sensitivity to this kind of contrast and in terms of changes with age/language exposure.

Indeed, some behavioral studies reported that at 4-6-month-old English-learning infants discriminate vowels differing only in length (100, 200 or 300 ms vowels in items such as [mad-ma:d], Eilers, Bull, Oller, & Lewis, 1984; and English learners tested with 88 ms versus 180 ms vowels in items like [tɛki-tɛ:ki], even after hearing monomodal distributions of vowel length, Pons, Mugitani, Amano, & Werker, 2006b). In contrast, Sato, Sogabe, and Mazuka (2010) found little evidence of

discrimination in Japanese 4-month-olds tested with a vowel length contrast between 100 and 200 ms in [mana] and [ma:na]. By the end of the first year, discrimination is certainly in place (English 11-month-olds in Eilers et al. 1984; Japanese 10-month-olds in both Sato et al. 2010 and Mugitani, Pons, Fais, Dietrich, Werker, & Amano, 2009).

Interestingly, the picture is even more complex when ERP and NIRS evidence is taken into account. Friederici, Friedrich, and Weber (2002) reported that German 2-month-olds detected a change from a frequent short vowel (202ms) to an infrequent long one (341ms), whereas they did not in the opposite direction (notice that the timing, complexity, and distribution of responses varied depending on infants' awake status). The same asymmetry was reported in Friedrich, Weber, and Friederici (2004), who were able to find some index of change detection in the short to long condition provided that infants were fully awake during testing. Friederici et al. (2008) also found a larger response for a long deviant in a sequence of short vowels in German 1-month-old infants (note that no reliable response was observed for male infants with high testosterone levels). Whereas none of the behavioral studies noted above reported significant asymmetries, one has been documented in Japanese toddlers of about 18 months of age, but in the *opposite* direction: they detected a change from long to short, but not short to long (Mugitani et al., 2009). Since no such asymmetry was found in a group of English-learning 18-month-olds tested with the same stimuli, Mugitani et al. (2009) conclude that the Japanese 18-month-olds' behavior reflected language exposure.

The richest longitudinal sample to date comes from a cross-sectional NIRS study that suggests that the development of discrimination for vowel length is not linear. Minagawa-Kawai et al. (2007a) measured the brain responses of Japanese infants in several age groups between 3 and 28 months of age to vowel length contrasts drawn from the same category (both classified as e.g. short by adult Japanese listeners), or from different categories (one short and the other long). While brain responses during change trials did not differ significantly in the within-category versus between-category change blocks at 3-4 and 10-11 months, they did at 6-7, 13-14, and 25-27 months. Interestingly, left-lateralization was evident later than in vowel quality studies (see previous Section): only the last group exhibited significantly left-dominant responses.

Given the diversity in methods, future longitudinal work may be able to shed light on the contributions of age, experience, and experimental choices in evoking asymmetrical patterns of discrimination for vowel quantity. In addition, it is of interest to study how vowel quality and quantity changes interact, particularly in languages like German or Dutch, where vowels often contrast simultaneously on both dimensions. Bohn and Polka (2001) reported that German 6- to 12-month-old infants were negatively affected in their ability to discriminate vowel quality changes by the removal of vowel duration information (see also Section 2.4). Similarly, Dutch infants of 11 and 15 months of age discriminated between typical examples of the native vowels [ɑ] and [a:], which differ in both vowel quality and duration, but not to stimuli that differed in only one of these dimensions (Benders, submitted).

5 Special populations

5.1 Assessing language development in at-risk populations

Since vowel discrimination can be studied very early, it could potentially provide an index of language development of individual infants to inform clinical application, and it could be applied to group differences that highlight the effects of maturation and/or experience on vowel perception. A more extensive review of the predictive value of infant speech perception measures has been undertaken elsewhere (see Cristia, Seidl, Soderstrom, & Hagoort, in press). Briefly, in terms of vowel processing, group comparisons between at-risk and typically-developing infants appear more robust and interpretable than the prediction of individual variation among the latter population.

In an early study, Swoboda et al. (1976, 1978) investigated vowel discrimination in 2-month-olds with a history of perinatal complications (such as low Apgar scores, premature birth, or extremely low birth weight). Like control participants, these infants reacted to both within-category and between-category contrasts when instantiated in long (250ms) vowels; however, they differed from control infants in the weakness of the between-category response when short (60ms) vowels were employed. That early study also reported a host of other differences between groups that extended beyond the presence or absence of a

response to a change, and pertained learning of the contingent sound presentation and the effect of ISI in the strength of the discrimination response.

A sizable EEG literature has extended our knowledge well beyond these initial results by comparing brain responses in infants at familial risk of language impairment with control infants. Some work suggests the key difference is in the timing of responses. Specifically, brain responses to a change from a short (202ms) to a long (341ms) vowel differed between 2-month-olds at familial risk of language impairments and controls not in the amplitude of the positive deflection to the deviant, but in its longer latency (Friedrich et al., 2004). Making precise claims is, however, difficult, given that not all experimental manipulations allow such group differences to emerge (Leppänen et al. 1999; Pihko et al. 1999).

Another strand in this literature investigates prematurely born infants, who tend to experience language delays in childhood (see Bosch, 2011 for an insightful discussion). Comparing fullterms and maturation-matched preterms, both poorer discrimination early on and poorer language-specific tuning have been reported. Figueras-Montiu and Bosch (2010) document, with HPP, poorer discrimination of native [doði-duði] at 4 months (with no differences across groups at 8 months). As for tuning, Jansson-Verkasalo et al. (2010)'s EEG work revealed no difference across fullterms and preterms in detection of native vowel changes measured at 6 and 12 months, nor on detection of a non-native contrast measured at 6 months. However, responses were *greater* for the non-native contrast in preterms than fullterms at 12 months. This was interpreted as poor neural commitment to the native language patterns, a conclusion strengthened in this study through a longitudinal investigation, which revealed smaller vocabularies at 24 months the larger the brain response to the non-native contrast at 12 months.

Results are more variable for the prediction of individual variation among typically developing infants. Whereas Tsao, Liu, and Kuhl (2004) report that poorer performance in Finnish [y-u] discrimination predicts slower language development, this result did not replicate in the same lab (Cardillo, 2010).

5.2 Vowel perception in bilinguals

A growing literature investigates the development of speech perception skills in bilinguals, who may follow a different developmental path than monolinguals by virtue of their different life experience (for a recent review, see Curtin, Byers-Heinlein, & Werker, 2011). While Spanish-Catalan bilinguals at 4 months of age performed equally to monolingual peers, at 8 months they had difficulties in distinguishing the contrasts [dɛði-deði] and [doði-duði] (Bosch & Sebastián-Gallés, 2003; Sebastián-Gallés & Bosch, 2009). However, they succeeded if the contrast was acoustically more distinct ([deði-duði]). This finding suggests that an interaction of bilingual status and acoustic characteristics influences discrimination. Differential distribution of attention in monolinguals versus bilingual infants may also play some role in explaining the difference in performance. The studies above were carried out with CF, a paradigm which relies on infants' recovery of attention. A newer method, AEM, calls on a different suite of cognitive skills, namely on infants' anticipatory eye movements to a previously learned association between vowel and side of screen; thus, this method operates like a forced-choice categorization. When tested with the latter method, 8-month-old bilinguals' performance in the acoustically less distinct contrasts was not significantly worse compared to their monolingual peers (Albareda-Castellot et al. 2011).

On the basis of such statements, it would seem that electrophysiological methods that do not load on attention could shed clearer light on whether monolingual and bilingual infants differ in vowel discrimination. In the preliminary results from a large study, which unfortunately does not include any statistical analyses, Shafer, Yu, and Garrido-Nag (2011) recorded EEG in Spanish-English bilinguals and English monolinguals, in age groups ranging from 3 to 36 months; here, we concentrate on their results from the first year (3, 6, and 12 months). Infants were presented with repeated [e], which were interspersed with infrequent [ɪ] near the middle and the end of the sequence. Overall, infrequent vowels elicited brain responses with a positive mismatch detection response, whose latency decreased with age in the monolingual group, whereas it did not seem to do so in the bilingual group.

Additionally, the amplitude of this response did not differ considerably across the monolingual and bilingual group, except possibly in the 12-month-old group. Unlike other groups, female bilinguals tended to show responses with an inverse polarity, which is in fact the typical polarity in the adult literature.

In a follow-up with the same stimulation, Shafer et al. (2012) concentrated on a group of 6-month-old Spanish-English bilinguals and another of 6-month-old monolinguals to explore the hypothesis that this polarity difference indicated increased sensitivity to the vowel change in female bilinguals. Based on the idea that final syllables attract more attention than medial syllables, the authors hypothesized that separate analyses of medial and final deviants would shed light on the polarity of the mismatch response. In consonance with their predictions, there was a negative response to the deviant in final position, but a positive one in the medial position, with the strength of both responses being inversely correlated when individual data were inspected. It is unclear to what extent this explains the group and sex differences described in the 2011 article, since neither language group nor sex appeared to interact with the polarity shift dependent on position. Nonetheless, there was an interaction with gender within the final position, as females showed relatively more negative responses than males. No interaction with group was reported in this paper, suggesting that at 6 months, mismatch responses are not significantly different in monolingual and bilingual infants.

6 Conclusion

Pooling all available published studies, this review has revealed a rich and diverse literature on infant vowel discrimination. To date, a similar up-to-date review on the early perception of consonants is missing. Evidence from a variety of studies suggests that some aspects of vowels and consonants are not processed in the same way (e.g., Bonatti, Peña, Nespor, & Mehler, 2004). Although these differences are not restricted to discrimination, in order to gain a comprehensive picture of early speech sound perception it will be important to assess to what extent the

conclusions from the present review would hold for vowels specifically, or for speech sounds in general.

Working towards this end, we have created the open-access, updatable online resource InPhonDB (sites.google.com/site/InPhonDB). It currently contains the majority of studies reviewed in this article, and it may be extended to studies on consonants. Currently, the spreadsheet-format resource is a useful complement to this qualitative review in providing quantitative information on relevant independent and dependent variables in each experiment.

Acknowledgements

References

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

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

Arimitsu, T., Uchida-Ota, M., Yagihashi, T., Kojima, S., & Watanabe S. (2011). Functional hemispheric specialization in processing phonemic and prosodic auditory changes in neonates. *Frontiers in Psychology, 2*, 202. doi: 10.3389/fpsyg.2011.00202

Benders, T. (submitted). Learning phonemes from multiple auditory cues: Dutch infants' language input and perception.

Best, C. T. (1994). *The emergence of native-language phonological influences in infants: A perceptual assimilation model*. In Judith. C. Goodman & Howard. C. Nusbaum (Eds.) The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words, 167–224. Cambridge, MA: The MIT Press.

Bohn, O. S., & Polka, L. (2001). Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *The Journal of the Acoustical Society of America, 110*(1), 504–15. doi: 10.1121/1.1380415

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2004). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16*, 451-459. doi: 10.1111/j.0956-7976.2005.01556.x

Bosch, L. (2011). Precursors to language in preterm infants: speech perception abilities in the first year of life. In O. Braddick, J. Atkinson, and G. Innocenti (Eds.) *Progress in Brain Research Volume 189*, 239-257. Amsterdam: Elsevier.

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*, 217–243. doi: 10.1177/00238309030460020801

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J. F. (2008). Hearing faces: how the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience, 21*(5), 905-921. doi:10.1162/jocn.2009.21076

Byrne, J. M., Miller, C. L., & Hondas, B. (1994). Psychophysiologic and behavioral responsivity to temporal parameters of acoustic stimuli. *Infant Behavior and Development, 17*, 245-254. doi: 10.1016/0163-6383(94)90003-5

Cardillo, G. C. (2010). *Predicting the predictors: Individual differences in longitudinal relationships between infant phonetic perception, toddler vocabulary, and preschooler language and phonological awareness.* (Doctoral dissertation). University of Washington.

Cheour, M., Alho, K., Ceponiené, R., Reinikainen, K., Sainio, K., Pohjavuori, M., … Näätanen, R. (1998). Maturation of mismatch negativity in infants. *International journal of psychophysiology, 29*(2), 217–26. doi: 10.1016/S0167-8760(98)00017-8

Cheour, M., Alho, K., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., … Näätanen, R. (1997). The mismatch negativity to changes in speech sounds at the age of three months. *Developmental Neuropsychology, 13*(2), 167–174. doi: 10.1080/87565649709540676

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

Cheour-Luhtanen, M., Alho, K., Sainio, K., Rinne, T., Reinikainen, K., Pohjavuori, M., … Näätanen, R. (1996). The ontogenetically earliest discriminative response of the human brain. *Psychophysiology, 33*(4), 478-481. doi: 10.1111/j.1469-8986.1996.tb01074.x

Cheour, M., Leppänen, P. H. T., & Kraus, N. (2000). Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. *Clinical Neurophysiology, 111*(1), 4-16. doi: 10.1016/S1388-2457(99)00191-1

Clarkson, M.G., & Berg, W. K. (1983). Cardiac orienting and vowel discrimination in newborns: Crucial stimulus parameters. *Child Development*, *54*(1), 162-171. doi: 10.1111/j.1467-8624.1983.tb00345.x

Cowan, N., Suomi, K., & Morse, P. A. (1982). Echoic storage in infant perception. *Child Development, 53*(4), 984–90. doi: 10.2307/1129138

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (in press). Infant predictors of language. *Child Development.* [online resource: sites.google.com/site/InVarInf]

Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics, 39*(4), 492–504. doi: 10.1016/j.wocn.2010.12.002

Eilers, R. E., Bull, D. H., Oller, D. K., & Lewis, D. C. (1984). The discrimination of vowel duration by infants. *The Journal of the Acoustical Society of America, 75,* 1213-1218. doi: 10.1121/1.390773

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

Figueras Montiu, M., & Bosch Galceran, L. (2010). Capacidades de discriminación fonética de un contraste vocálico nativo en el prematuro. *Psicothema, 22*(4), 669-676. Retrieved from http://www.unioviedo.es/reunido/index.php/PST/article/view/8935

Friederici, A. D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *NeuroReport, 13*(10), 1251-1254.

Friederici, A. D., Pannekamp. A., Partsch, C.-J., Ulmen, U., Oehler, K., Schmutzler, R., & Hesse, V. (2008). Sex hormone testosterone affects language organization in the infant brain. *NeuroReport, 19*(3), 283-286. doi: 10.1097/WNR.0b013e3282f5105a

Friedrich, M., Weber, C., & Friederici, A. D. (2004). Electrophysiological evidence for delayed mismatch response in infants at-risk for specific language impairment. *Psychophysiology, 41*(5), 772-782. doi: 10.1111/j.1469-8986.2004.00202.x

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25*(4), 577–88. doi: 10.1037/0012-1649.25.4.577

Hämäläinen, J. A., Ortiz-Mantilla, S., & Benasich, A. A. (2011). Source localization of event-related potentials to pitch change mapped onto age-appropriate MRIs at 6months of age. *NeuroImage, 54*(3), 1910-1918. doi: 10.1016/j.neuroimage.2010.10.016

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America, 97*, 3099. doi: 10.1121/1.409456

Jansson-Verkasalo, E., Ruusuvirta, T., Huotilainen, M., Alku, P., Kushnerenko, E., Suominen, K., ...Hallman, M. (2010). Atypical perceptual narrowing in prematurely born infants is associated with compromised language

acquisition at 2 years of age. *BMC Neuroscience, 11*(1), 88. doi:10.1186/1471-2202-11-88

Jusczyk, P. W. (1985). The high-amplitude sucking technique as a methodological tool in speech perception research. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of Audition and Vision in the First Year of Postnatal Life: A Methodological Overview* (pp. 195–222). Westport, CT, US: Ablex Publishing.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology, 23*(5), 648–54. doi: 10.1037/0012-1649.23.5.648

Kemler Nelson D.G., Jusczyk P.W., Mandel D.R., Myers J., Turk A., & Gerken L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*(1), 111-116. doi 10.1016/0163-6383(95)90012-8

Kuhl, P. K. (1979). Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America, 66*(6), 1668–79. doi: 10.1121/1.383639

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6*(2-3), 263–285. doi:10.1016/S0163-6383(83)80036-8

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. doi: 10.3758/BF03212211

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology, 4*(6), 812–822. doi:10.1016/0959-4388(94)90128-7

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

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development, 7*(3), 361-381.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America, 100*(4), 2425-2438.

Kuhl, P. K., & Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception & Psychophysics, 31*(3), 279–292. doi: 10.3758/BF03202536

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 31*(255), 606–608. doi: 10.1126/science.1736364

Kujala, A., Huotilainen, M., Hotakainen, M., Lennes, M., Parkkonen, L., Fellman, V., & Näätänen, R. (2004). Speech-sound discrimination in neonates as measured with MEG. *NeuroReport, 15*(13), 2089–92. doi: 10.1097/00001756-200409150-00018

Lacerda, F. (1991). Perception of CV-utterances by young infants: Pilot study using the high-amplitude-sucking technique. In: *Perilus XII. Experiments in Speech Processes.* University of Stockholm, Institute for Linguistics.

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

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

Mazuka, R., Hasegawa, M., & Tsuji, S. (2013). Development of non-native vowel discrimination: Improvement without exposure. *Developmental Psychobiology.* Advance online publication. doi: 10.1002/dev.21193

McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy, 6*(2), 203–229. doi:10.1207/s15327078in0602_4

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

Miller, J. L., & Eimas, P. D. (1979). Organization in infant speech perception. *Canadian journal of psychology, 33*(4), 353–67. doi: 10.1037/h0081732

Minagawa-Kawai, Y., Cristia, A., & Dupoux, E. (2011). Cerebral lateralization and early speech acquisition: a developmental scenario. *Developmental Cognitive Neuroscience, 1*(3), 217-232.

Minagawa-Kawai, Y., Mori, K., Hebden, J. C., & Dupoux, E. (2008). Optical imaging of infants' neurocognitive development: Recent advances and perspectives. *Developmental Neurobiology*, 68(6), 712-728.

Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007a). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *The Journal of neuroscience, 27*(2), 315–21. doi:10.1523/JNEUROSCI.1984-06.2007

Minagawa-Kawai, Y., Naoi, N., Nishijima, N., Kojima, S., Dupoux, E. (2007b). Developmental Changes in Cerebral Responses to Native and Non-Native Vowels: A NIRS Study. *Proceedings of the International Conference of Phonetic Sciences XVI.*

Moon, C., Lagercrantz, H., & Kuhl, P. (2013). Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatrica, 102*, 156-160. doi: 10.1111/apa.12098

Mugitani, R., Pons, F., Fais, L., Dietrich, C., Werker, J. F., & Amano, S. (2009). Perception of vowel length by Japanese- and English-learning infants. *Developmental psychology, 45*(1), 236–47. doi: 10.1037/a0014043

Nittrouer, S. (2001). Challenging the notion of innate phonetic boundaries. The *Journal of the Acoustical Society of America, 110*, 1598-1605. doi: 10.1121/1.1379078

Phan, J., & Houston, D. M. (2008). Infant Dialect Discrimination. *Research on Spoken Language Processing, Progress Report No. 29.*

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

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,* 577-592. doi: 10.1121/1.415884

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

Polka, L., & Werker, J. (1994). Developmental changes in perception of nonnative vowel contrasts. Journal of experimental psychology. *Human perception and performance, 20*(2), 421–435. doi: 10.1037/0096-1523.20.2.421

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), 965–76. doi: 10.1111/j.1467-8624.2012.01740.x

Pons, F., Mugitani, R., Amano, S., & Werker, J. F. (2006b). Distributional learning in vowel length distinctions by 6-month-old English infants. Presented at *the International Conference on Infant Studies;* Kyoto, Japan (abstract).

Pons, F., Sabourin, L., Cady, J. C., & Werker, J. F. (2006a). Distributional learning in vowel distinctions by 8-month-old English infants. In *28th Annual conference of the cognitive science society*, Vancouver, BC, Canada (abstract).

Sato, Y., Mori, K., Furuya, I., Hayashi, R., Minagawa-Kawai, Y., & Koizumi, T. (2003). Developmental changes in cerebral lateralization during speech processing measured by near infrared spectroscopy. *Japanese Journal of Logopedic Phoniatrics, 44*, 165–171. doi: 10.5112/jjlp.44.165.

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

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), 874–87. doi: 10.1111/j.1467-7687.2009.00829.x

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.

Shafer, V. L., Yu, Y. H., & Garrido-Nag, K. (2012). Neural mismatch indices of vowel discrimination in monolingually and bilingually exposed infants: Does attention matter? *Neuroscience Letters, 526*, 10-14.

Swoboda, P. J., Kass, J., Morse, P. A., & Leavitt, L. A. (1978). Memory factors in vowel discrimination of normal and at-risk infants. *Child development, 49*(2), 332–9. doi: 10.2307/1128695

Swoboda, P. J., Morse, P. A., & Leavitt, L. A. (1976). Continuous vowel discrimination in normal and at risk infants. *Child development, 47*(2), 459–65. doi: 10.2307/1128802

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin and Review, 9*(2), 335–40. doi: 10.3758/BF03196290

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

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

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

Tsuji, S., & Cristia, A. (2013). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, *56*(2), 179-191. doi: 10.1002/dev.21179

Tsuji, S., & Cristia, A. (2014). *Community-augmented meta-analyses: The example of infant vowel discrimination*. Manuscript in preparation.

Weir, C. G., & Lamb, J. (2009). Differential sensitivity of newborns to synthetic vowel stimuli. *British Journal of Audiology, 24*(6), 419–430. doi: 10.3109/03005369009076582

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology, 34*(6), 1289–309. doi: 10.1037/0012-1649.34.6.1289

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

Werker, J. F., Polka, L., & Pegg, J. E. (1997). The conditioned head turn procedure as a method for testing infant speech perception. *Early Development and Parenting, 6*(3-4), 171–178. doi:10.1002/(SICI)1099-0917(199709/12)6:3/4<171::AID-EDP156>3.0.CO;2-H

Whittingstall, K., Stroink, G., Gates, L., Connolly, J. F., & Finley, A. (2003). Effects of dipole position, orientation and noise on the accuracy of EEG source localization. *Biomedical Engineering Online, 2*(1), 14.

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., ...Carney, E. (2011). Neural coding of formant-exaggerated speech in the infant brain. *Developmental Science, 14*(3), 566–581. doi: 10.1111/j.1467-7687.2010.01004.x

Perceptual attunement in vowels: A meta-analysis

Chapter 4

Abstract

Although the majority of evidence on perceptual narrowing in speech sounds is based on consonants, most models of infant speech perception generalize these findings to vowels, assuming that vowel perception improves for vowel sounds that are present in the infant's native language within the first year of life, and deteriorates for non-native vowel sounds over the same period of time. The present meta-analysis contributes to assessing to what extent these descriptions are accurate in the first comprehensive quantitative meta-analysis of perceptual narrowing in infant vowel discrimination, including results from behavioral, electrophysiological, and neuroimaging methods applied to infants 0-14 months of age. An analysis of effect sizes for native and non-native vowel discrimination over the first year of life revealed that they changed with age in opposite directions, being significant by about 6 months of age.

1 Introduction

Over the last 50 years, the experimental study of infant speech sound discrimination has provided us with important insights into early perceptual abilities and their change as a function of development and language exposure. Much attention has been paid to perceptual narrowing: Infants are thought to start out with language-universal perceptual abilities (i.e., patterns of perception that are independent of language exposure), and these abilities would become tuned to the infant's ambient language as a function of exposure, culminating in the end of the first year of life with qualitatively different patterns of perception by infants exposed to different languages.

Perceptual narrowing provides crucial insights into the psychobiological bases of language because it is the first sign that infants are acquiring their native language. Therefore, attunement can shed light on the complex interplay of biological and experiential factors involved in the unfolding of linguistic abilities. For instance, we have recently learned that infants exposed to serotonin reuptake inhibitors prenatally show perceptual attunement *earlier* than control infants (Weikum, Oberlander, Hensch, & Werker, 2012). Additionally, individual variation in attunement predicts later language development (a recent review in Cristia et al., in press). Compared to consonants, vowels are more clearly heard in the womb (a recent summary in Granier-Deferre, Ribeiro, Jacquet, & Bassereau, 2011). Therefore, attunement for vowels results from speech exposure starting even before birth, and it has been thought to be evident earlier than consonants (a question we revisit below). Thus, vowel discrimination scores could be particularly useful to make decisions regarding both the at-risk status of specific infants and their priority for treatment, and the short-term effects of early treatments, at a very young age.

An additional reason for studying perceptual narrowing in vowels is internal to the field of infant speech perception. In fact, the majority of evidence for perceptual narrowing in speech perception comes from consonants. Nevertheless, prominent models of early speech perception by and large consider perceptual narrowing to apply to all speech sounds rather than to consonants in particular. Therefore, it is crucial to assess

how far such generalization is suitable, as some evidence suggests that vowels and consonants are not completely comparable. To begin with, a host of infant, child, and adult psycholinguistic evidence suggests that they are not processed in precisely the same way (e.g., Bonatti, Peña, Nespor, & Mehler, 2004; Caramazza, Chialant, Capasso, & Miceli, 2000 and references therein). Moreover, while infants' perception can change with brief lab-based exposures to consonants (e.g., Cristia, McGuire, Seidl, & Francis, 2011 and references therein) and lexical tones (Liu & Kager, 2011), such perceptual warping has failed to occur for vowels (Pons, Sabourin, Cady, & Werker, 2006; Pons, Mugitani, Amano, & Werker, 2006). Based on these substantial differences in findings on vowels and consonants, it is of particular interest to revisit the question of perceptual narrowing for vowels specifically.

Before turning to the quantitative study, we will provide a brief overview of a few prominent models of perceptual narrowing in infant speech perception. The Native Language Magnet model (NLM; Kuhl, 1994; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola, & Nelson, 2008) was originally based on evidence from vowel discrimination (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), and it is better specified than the other models in terms of when and how vowel perception becomes attuned to the native language (e.g., Kuhl et al., 2008). For this reason, we expand on this particular model and the evidence supporting it first.

The perceptual magnet effect refers to the phenomenon that vowel tokens are treated differently depending on how prototypical they are of a vowel category. Vowel prototypes in the context of NLM have been described as the representations most often activated (Kuhl et al., 2008), or as the centers of a vowel category (cf. Feldman, Griffiths, & Morgan, 2009). With exposure to the native language, prototypical vowels start acting like magnets, warping perceptual space such that it shrinks around prototypical vowels and creates non-linearities in perception. Thus, discrimination of tokens close to a prototype becomes worse than discrimination of tokens towards the category boundary. Since warping depends on exposure to sounds mapping on native vowels, no such magnet effect occurs for non-native vowels.

Early evidence for language-specific vowel perception relied on non-linearities in the detection of within-category changes. A first

indication for native vowel prototypes was given in two studies on 6-month-old English-learning infants, who were better able to discriminate vowels in the direction from a non-prototypical to a prototypical native exemplar of [i] (the vowel in the word 'sheep') than vice versa (Grieser & Kuhl, 1989; Kuhl, 1991). The seminal Kuhl et al. (1992) study subsequently documented that American English 6-month-olds failed to detect many vowel changes around the prototypical [i] in their language but were sensitive to the same acoustic distances centered around [y], while Swedish infants tested with the same stimuli readily heard such changes around the non-native [i] and missed them around native [y]. Based on this evidence, Kuhl and colleagues proposed that narrowing occurs earlier in vowels (by around 6 months) than in consonants (closer to 8-10 or as late as 10-12 months; Werker & Tees, 1984). The NLM model in its current form is not restricted to within-category changes, and has been invoked in several studies that document developmental changes (Polka & Werker, 1994), cross-linguistic differences (e.g., Bosch & Sebastián-Gallés, 2003), or cross-contrast differences (better discrimination for a native than a non-native contrast, e.g., Cheour et al., 1998; but see Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995).

NLM is not the only model that has been put forward to account for infant speech processing, and other existing models could also capture the aforementioned changes in vowel discrimination. The Perceptual Assimilation Model (PAM; Best, 1994) is also well known. However, it provides an account primarily in terms of how non-native sounds are processed once native perceptual categories have already been formed, rather than explaining the process by which native and non-native categories come to be treated differently, and thus it is not a model of perceptual attunement. We note here that PAM will become relevant once more in the final discussion below.

The developmental framework for Processing Rich Information from Multi-dimensional Interactive Representations (PRIMIR; Werker & Curtin, 2005) is another mainstream model of infant speech perception. In this model, perception always must be conceived as operating at multiple levels or planes at the same time. One of these is the General Perceptual plane, which encodes discrimination abilities that are initially independent of language exposure, and thus very similar in infants

exposed to different languages. As a function of language experience, including not only listening but also visual and articulatory experience, this plane is somewhat reorganized reflecting the native language categories, such that some innate boundaries are erased, enhanced, or shifted. This model also states that this representation, albeit language-specific, is not very robust or abstract. True phonological categories will only emerge as the child begins to learn words and store them in the Word Form plane, at which point a third plane (Phoneme plane) will begin to be developed (compare this with the Word Recognition and Phonetic Structure Acquisition, WRAPSA model, e.g. Jusczyk, 1993). Thus, PRIMIR differs from NLM in several aspects with regards to perceptual attunement. First, it more openly incorporates visual and articulatory experience in the process of attunement. Second, it predicts that reorganization may also be brought about by word learning.

Aside from these differences, both PRIMIR and NLM hold that infant vowel perception changes over the first year, with native discrimination improving and non-native discrimination deteriorating. As mentioned above, there is some evidence in favor of this view. However, other studies fail to find developmental changes (which are assumed to be due to experience) or cross-linguistic differences within the first year of life (e.g., Polka & Bohn, 1996; Sebastián-Gallés & Bosch, 2009). Moreover, where developmental changes are indeed reported, the time point of their occurrence is debated. While some studies find a modulation by 6-8 months of age (e.g., Bosch & Sebastián-Gallés, 2003; Kuhl et al., 1992; Polka & Werker, 1994), others only find modulations from 10 months of age onwards (e.g., Polka & Bohn, 2011; Pons et al., 2012). Therefore, based on these studies it is far from clear that the reorganization for vowels is truly robust; and that it happens earlier than 6 months.

Given the considerable diversity in outcomes, it was relevant to assess the evidence for perceptual narrowing in vowels critically. To this end, we carried out a comprehensive review of the vowel discrimination literature, and identified studies where two or more age groups of infants had been tested on the same vowel contrast. We then retrieved or calculated the effect size indicative of discrimination in each case, and combined effect sizes using meta-analytic methods, as explained in detail in the next section. We sought to answer the following questions. First, do

effect sizes change differently with infant age depending on whether the contrast is native or non-native? A change in opposite directions for native and non-native contrasts and with a more positive slope for native contrasts is indicative of perceptual narrowing. Subsequent questions investigated specific features of this process: Second, does native contrast discrimination improve with age? Third, does non-native discrimination deteriorate with age? Finally, do these changes occur by about 6 months?

2 Methods

2.1 Search protocol

A full search on scholar.google.com was conducted in September 2012 with the keyword combination "{infant|infancy} & {vowel|speech sound|syllable} & discrimination". Additionally, the search terms were translated into French, German, Japanese, and Spanish for additional searches. We also asked experts in the field to inform us of any published or unpublished studies we had missed. Experts were defined as scientists having participated in at least 2 studies identified in our intermediate search sample or who were part of a lab where such research had taken place, and who were still active in the field or could be otherwise contacted. Further, articles were added based on a screening of articles cited and articles citing the articles in the remaining search sample. The complete sample is available as a public resource (Tsuji & Cristia, 2014, https://sites.google.com/site/inphondb/).

The search sample was narrowed down to the final search sample of 19 articles based on the following inclusion criteria: (1) The study focused on normally developing infants, with at least one age group involved being 12 months of age or less. (2) At least two age groups were assessed on the same vowel contrast. (3) Discrimination was the key component of the task. (4) The two stimuli being discriminated were described as differing only in vowel quality or quantity. (5) The two stimuli being discriminated were auditory only. If a visual stimulus was presented, it was only for the purpose of indirectly measuring infants' attention by looking time, or in order to distract infants with unsystematic stimuli. (6) The article was published in any source, including peer-reviewed journals (N = 16, in addition, 1 article is under review: Benders, 2013, and 2 articles are in

preparation: Liu & Kager, in preparation a, and Liu and Kager, in preparation b), conference proceedings ($N = 1$), and theses ($N = 1$). Given that the key question pertained to the first year, we excluded records focusing on infants older than 15 months of age.

The 19 articles of the final search sample contained 116 eligible records. We define a record as an experimental unit for which a separate result was reported. In most cases, this was one experiment on one group of infants, but sometimes it was the case that, for instance, values for different orders of presentations were reported separately. In such cases, we counted each reported unit as one record.

2.2 Experimental methods for assessing infant speech sound discrimination

Before turning to the quantitative analysis, we will give a short overview of the methods used to assess speech sound discrimination in infants. Along with the methods themselves, we will outline the respective dependent variables on which later effect size calculations were based. Although the methods combined in this meta-analysis are varied, they all focus on the same construct, namely infants' response to a sound change. As such, they are suitable for combination into one meta-analysis.

Central Fixation (CF), also sometimes referred to as Visual Habituation, is a paradigm where a central audiovisual stimulation is presented contingent on the infants' attention (for details, see Werker, Cohen, Lloyd, Casasola, & Stager, 1998). Therefore, it can be used in combination with habituation-dishabituation designs, where the same stimuli are presented repeatedly until attention wanes. It can also be used in familiarization-preference designs, where the initial exposure is fixed in duration (rather than dependent on a decline of attention). In both cases, the habituation or familiarization phase is followed by a test phase, in which the infant is presented with one or multiple trials of the same stimulus, as well as one or multiple trials of a novel stimulus. The looking times to the same and novel trials are the dependent variables, and the difference in looking times is measured within-participants. All but one of the studies using CF in the current sample followed the above design. One study (Benders, 2013) employed the stimulus alternation design, a variant

of CF in which infants are presented non-alternating trials with repetitions of the same stimulus as well as alternating trials in which the same stimulus alternates with a novel stimulus, without a prior habituation or familiarization phase. The study with this design assessed differences in looking times by calculating the ratio of look duration during alternating trials divided by the look duration during the surrounding non-alternating trials.

In the Headturn Preference Paradigm (HPP), audiovisual stimulation is presented on the right and left sides of the infants contingent on their head-turns to the respective sides (for details, see Kemler Nelson et al., 1995). Like CF, HPP can be used in familiarization-preference designs such that the infant is initially exposed to repetitions of the same stimulus until a fixed looking time has accumulated. In the subsequent test phase, the infant is presented with multiple trials of the same or a novel stimulus, which are presented on either the left or the right side paired with a flashing light in pseudo-random order. The difference in infants' orientation times to trials with the same or novel stimulus is measured within-participants.

The Conditioned Head-Turn (CHT) paradigm also makes use of infants' headturns towards a visual reinforcement. Infants are trained to respond to sound changes by turning their head towards a visual reinforcement each time there is a sound change. At a subsequent stage, the visual reinforcement becomes conditional on correct headturns (details in e.g. Werker, Polka, & Pegg, 1997). After training infants on this contingency, they are tested on the sound contrast of interest (sometimes on several contrasts over subsequent days). A single measure per participant, such as the percent of correct headturns to a sound change is reported as the dependent measure. While some studies also report the sensitivity measures d-prime or a-prime, we base our effect size calculations of percent correct in the current sample because this was the measure consistently reported in all studies.

In electroencephalography (EEG), the electrical activity of the brain is measured with electrodes placed on the scalp. Infant speech sound discrimination has often been measured through the mismatch response (MMR), an event-related potential (ERP) response that appears when a rare (deviant) stimulus is presented in a row of repeated (standard) (for

details, refer to Cheour, Leppänen, & Kraus, 2000). As the method does not require attention to stimulation, infants are often silently entertained with toys or a silent movie during the experiment. The MMR is defined as the difference wave between the response to standard and deviant stimuli. Both the latency and amplitude of the MMR constitute important measures. For the purpose of the current study, we chose to base effect size calculations on the amplitudes. The auditory MMR in adults occurs as a fronto-central negative potential at around 150-250 ms after onset of stimulation, while in infants both positive and negative polarities in a broader time-range are observed. In one of the two EEG studies included in the final analysis, the MMR was defined as the most negative peak in a time window of 200-500 ms, and amplitude was calculated from a 50 ms time-window centered around the peak at right frontal electrode F4. In the other study, the MMR was defined as the most negative peak in a time-window from 150-300 ms, and amplitude was calculated as the average over fronto-central bilateral electrodes F3, C3, P3, F4, C4, P4 in a 100 ms time-window centered around the peak.

Near-infrared spectroscopy (NIRS) measures changes in hemoglobin oxygenation in specific brain regions. Speech sound discrimination in infants is measured by presenting blocks in which a single (type of) stimulus is repeated, as well as "alternating" blocks, in which that stimulus is interspersed with a novel one. As in EEG, infants do not need to attend to stimulation and are often entertained with unrelated visual stimuli during the experiment. Two types of dependent variables have been typically used for measuring speech sound discrimination in infants: changes in oxygenated or deoxygenated hemoglobin concentration between the two types of blocks mostly in probes over the superior temporal gyrus (STG) in the left hemisphere, or a laterality index calculated from probes over STG in both hemispheres, indicating how selective the activation is. As the former is regarded as a measure of pure discrimination, while the latter is regarded to reflect more linguistic processing, we aimed to include the former in the analysis. However, for the three studies included in the final analysis, we succeeded in retrieving the former in two, and the latter in all three studies. We therefore decided to calculate the effect sizes based on the laterality index for all three studies.

We decided on the effect size measure by experimental method as outlined below. We then divided the articles randomly and coded them independently. After the coding process, records were cross-checked for inconsistencies several times.

2.3 Selection of samples and coding of effect size

Of the 116 records, we succeeded in calculating effect sizes for 100 records (86%) out of 18 studies (cf. Table 1 for an overview of studies for which effect sizes could be calculated). The articles for which we were able to calculate effect sizes were published between 1992 and 2012 (2 were under review and 2 in preparation) by 15 different first authors. Following standard meta-analytic practice, we removed outliers above or below 3 SD from the sample mean (Lipsey & Wilson, 2001). Three records were removed by this criterion (cf. Fig.1). Thus, the final dataset included 97 records, 75 for native and 22 for non-native. The records were based on a total of 1613 unique infants, some of them measured repeatedly for a total of 1882 unique measurements.

Effect sizes were calculated based on Lipsey and Wilson (2001). As outlined in 2.2, depending on the method, the outcome was either reported as a comparison between two conditions within one group of infants (CF[1], HPP), or a single score that could be a ratio (one CF study), a difference score (ERP, NIRS), or a percentage (CHT). Cohen's d, an effect size measure that involves dividing the differences in means by their standard deviation, was calculated in all cases. As the majority of records had a sample size < 20, Hedges' correction for small samples was applied to all effect sizes.

In CF and HPP studies (57 records), the difference between same and novel trials in the test phase was a within-subject measure. For these two methods, the standardized mean gain effect size for within-subject comparisons (Lipsey & Wilson, 2001) was calculated, in which the mean difference score between same and novel trials is divided by their pooled standard deviation. In calculating the standard error of the standardized mean gain effect size, the correlation between the means of the same and

---

[1] Excluding one study using the stimulus alternation paradigm and calculating a ratio as the outcome variable.

Table 1. Studies contributing data-points to the meta-analysis

No	Authors	Year	Age (days)	Native language	Contrasts	Nativeness	Method	Dependent Variable
1	Benders	2013	359, 461	Dutch	sɑk-sɑːk, sɑk-sak/sɑːk-sɑːk, sɑk-sɑːk/sɑk-sɑːk	Native, non-native	CF	Ratio of alternating to non alternating looks
2	Bosch & Sebastián-Gallés	2003	139, 246, 371	Catalan, Spanish, bilingual	dɛði-deði	Native, non-native	HPP	Looking time to old and novel trials
3	Cardillo	2010	213, 324	English	u-y	Non-native	CHT	Mismatch response
4	Cheour, Alho, Ceponiene, Reinikainen, Sainio, et al.	1998	3, 97	Finnish	y-i	Native	EEG	Mismatch response
5	Figueras Montiu & Bosch Galceran	2010	140, 264	Catalan, Spanish	doði-duði	Native	HPP	Looking time to old and novel trials
6	Jansson-Verkalo, Ruusuvirta, Huotilainen, Alku, Kushnerenko, et al.	2010	183, 365	Finnish	e-ö, e-ō	Native, non-native	EEG	Mismatch response
7	Liu & Kager a	prep.	177, 270, 348, 443	Dutch	I-i	Native	CF	Looking time to old and novel trials
8	Liu & Kager b	prep.	174, 266, 356, 447	bilingual	I-i	Native	CF	Looking time to old and novel trials
9	Marean, Werner, & Kuhl	1992	60, 91, 183	English	a-i	Native	CHT	% correct headturns
10	Mazuka, Hasegawa, & Tsuji	2013	141, 309	Japanese	biːk-beːk, buːk-boːk, buːk-byːk	Non-native	CF	Looking time to old and novel trials
11	Minagawa-Kawai, Mori, Naoi, & Koijima	2007	106, 198, 319	Japanese	mama-mamaː	Native, non-native	NIRS	Laterality Index
12	Minagawa-Kawai, Naoi, Nishijima, Kojima, & Dupoux	2007	119, 229	Japanese	i-uu, u-uu	Native, non-native	NIRS	Laterality Index
13	Mugitani, Pons, Fais, Dietrich, Werker, & Amano	2009	302	Japanese	taku-taːku	Native	CF	Looking time to old and novel trials
14	Polka & Bohn	2011	103, 144	Danish	dɒt-dʌt, det-dɛt, døt-det	Native, non-native	CHT	% correct headturns
15	Pons, Albareda-Castellot, & Sebastián-Gallés	2012	138, 200, 375	Catalan, Spanish	bel-bil	Native	CF	Looking time to old and novel trials
16	Sato, Mori, Furuya, Hayashi, Minagawa-Kawai, & Koizumi	2003	100, 127, 154, 181	Japanese	itta-itte	Native	NIRS	Laterality Index
17	Sato, Sogabe, & Mazuka	2010	121, 228, 288	Japanese	mana-maːna, mana-mina	Native	CF	Looking time to old and novel trials
18	Sebastián-Gallés & Bosch	2009	132, 252, 380	Catalan, Spanish, bilingual	doði-duði, do-du, deði-duði	Native	HPP	Looking time to old and novel trials

novel trials is taken into account. The inclusion of a correlation term leads to a smaller standard error the larger the correlation, thus taking into account the increased precision of within-subject measures. This correlation was not reported by any of the studies included, but we were able to obtain the original correlations from the first authors of six studies (personal communication), which covered 42 experiments. For the remaining 15 experiments, we chose the median correlation of these 42 data points, which was $r = 0.505$ ($SD = 0.255$).

All other studies reported one value per record. This value could either be a ratio (one CF study, 3 records), a difference score (ERP and NIRS, 23 records), or a percentage (CHT, 14 records). For these cases, we calculated the standardized mean difference score (Lipsey & Wilson, 2001) for between-subject comparisons. This effect size is equivalent to the standardized mean gain score when sample sizes of control group and experimental group are the same. In order to calculate the effect size, we assumed a control group performing at the respective chance level (1 for the CF study, 0 for ERP and NIRS, 50% for CHT). The standard error of the effect size for uncorrelated samples was calculated. The weight of all effect sizes was obtained as the inverse of the squared standard error.

2.5 Coding of moderator variables

The only relevant participant characteristic for the present analyses was infant age. We entered mean or median age in days into the analysis. If a range was reported instead of a mean or median, we chose the midpoint of the range as an estimator of age. If only age in months was reported, we estimated the age in days by multiplying the number of months by 30.42. We were able to estimate age for all experiments based on these procedures.

The only relevant stimulus characteristic included in the current analyses was the phonemic status of the stimulus in the infants' native language[2]. Stimuli were coded as native if the vowels were reported to be

---

[2] Additionally, we coded measures of spectral and temporal distance between stimuli. Spectral distance refers to differences in vowel formant frequencies, and temporal distance refers to differences in vowel length. For the present sample, a spectral distance could be estimated for only 60% of records, and a temporal distance for 36% of records. Including these measures in the key regression for this study was not possible, as it would have imposed a serious curfew on our statistical power.

present in the vowel inventory of the language by the authors. All other stimuli were coded as non-native. Non-native stimuli could thus either be non-native vowels, or speech sounds that were modified such that they were not contrastive in the infants' native language. The latter was the case for two studies using a vowel length distinction outside of the contrastive range for the native language (e.g., Minagawa, Mori, Naoi, & Kojima, 2007), and one study where one of a pair of identifying features was neutralized (either quality or length, Benders, 2013).

3 Results

3.1 Preliminary Analyses

A set of preliminary analyses was conducted to assess overall sample characteristics. We specifically aimed at investigating (1) possible asymmetries in the funnel plot as a potential indicator of publication bias, (2) if there was sufficient heterogeneity in the sample to justify further analysis, and (3) if effect sizes from different methods could be combined into a single analysis, to boost power. Analyses were performed with the meta (Schwarzer, 2012) and metafor (Viechtbauer, 2010) packages for R (R Core Team, 2012).

We analyzed funnel plot asymmetry as a potential indicator of publication bias (Egger, Smith, Schneider, & Minder, 1997). In a funnel plot effect sizes are plotted against some measure of study size, and in a symmetric plot large studies are expected to cluster in the middle, while smaller studies spread to both sides. Figure 1 shows an underrepresentation of studies in the lower left corner, that is, studies with a high standard error and small effect size. This could occur for a variety of reasons, including that such studies may be set aside before or after the submission stage on the grounds that the sample size is too small. Please note that the rightmost three datapoints are outliers over 3 SD from the sample mean and were excluded from subsequent analyses. A linear regression on funnel plot asymmetry reaches significance [$t(95) = 4.93$, $p < .001$], suggesting bias (publication or otherwise) in our sample.
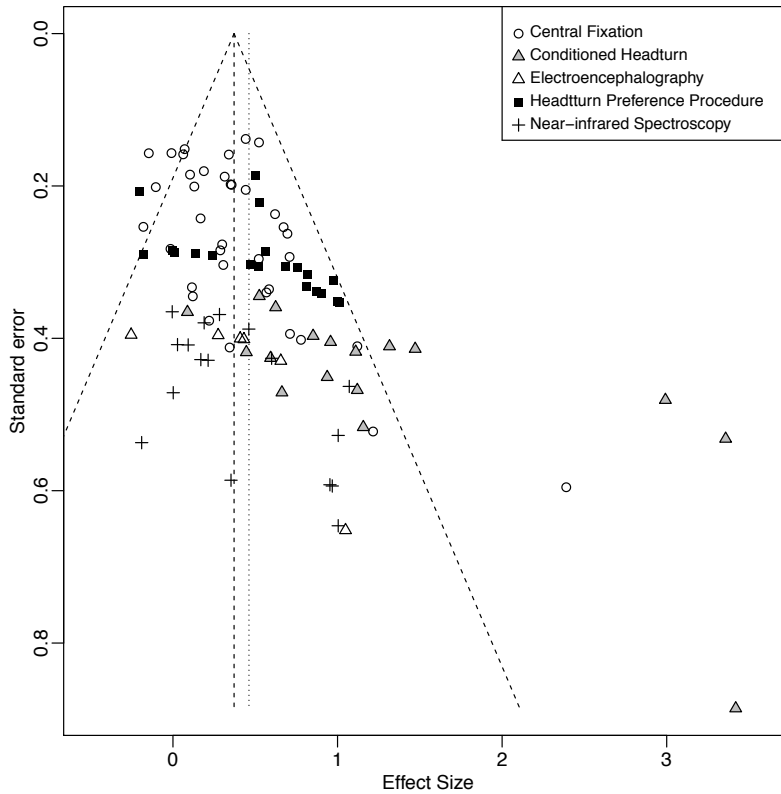
Figure 1: Funnel plot of effect sizes by method. Different methods are represented with different symbols, as shown in the legend.

To investigate whether the asymmetry we found reflected different effect size distributions across methods rather than an overall bias, analyses of funnel plot asymmetry were also conducted separately by method. We found significant asymmetry for all methods, with the sample of EEG studies being too small to assess asymmetry. These results are not reported here but available on request.

Figure 1 furthermore gives an indication that experiments cluster by method. We followed up on this observation by measuring the sample characteristics, first overall and then by method. As a first step, we estimated the overall effect size. We chose a random effects model for the analysis, which allows heterogeneity between studies due to differences in, for instance, sample characteristics or method chosen. The mean weighted effect size under a random effects model was *estimate* = 0.398 (*SE* = 0.039), with the lower bound of the 95% confidence interval CI_L = 0.322, and the higher bound CI_H= 0.475. This effect size was significantly

different from zero (z = 10.19, p < .001). As a second step, we assessed heterogeneity of the sample. Next to estimating the mean true effect, the amount of heterogeneity among the true effects needs to be estimated in a random-effects model. τ^2 measures between-study variance as an estimate of the difference between total observed variance and within-study variance. The total amount of between-study variance was τ^2 = 0.050 (estimated by restricted maximum likelihood, REML). Cochran's Q-test for homogeneity indicated significant sample heterogeneity [Q(96) = 158.069, p < .001]. Expressed in percentages, the variability explained by heterogeneity rather than sampling error was I^2 = 38.31% [CI_L = 20.61%, CI_H = 59.83%]. This result indicates that the sample variance is larger than would be expected from sample error, which justifies the introduction of moderator variables into the analysis.

In order to estimate the variance explained by the experimental method, we conducted a second analysis on overall sample characteristics, introducing experimental method as a moderator variable. The CF method was used as the reference level for this factor, because it has the largest amount of observations (40) and the lowest mean effect size. The Q-test showed significant heterogeneity between methods [Q(4) = 17.727, p = .001], and the effect of CHT (*estimate* = 0.524, z = 4.03, p < .001) and HPP (*estimate* = 0.178, z = 1.97, p = .049) were significant, with a significantly higher mean effect size than CF. Residual heterogeneity remained significant [τ^2 = 0.033, Q(92) = 133.282, p = 0.003], indicating that method did not account for all the variance.

The above analyses show considerable heterogeneity between methods, cautioning us to be careful in combining effect sizes from different experimental methods into one analysis. Moreover, residual heterogeneity also remains considerable, suggesting that the sample contains variability beyond the portion accounted for by method. We therefore included method as a moderator variable. It should also be noted that data on native contrasts (k = 75) outnumber data on non-native ones (k = 22), as evident in Figure 2.

3.2 Does effect size vary developmentally as a function of whether the contrast is present in the infants' native language?

We entered vowel nativeness (native, non-native), age (in days), and their interaction into the analysis. Given the heterogeneity of effect sizes across methods, method was entered as an additional factor. There is no reason to predict that the relationship between age and nativeness will interact with method; moreover, there are too few points to reliably estimate the slope of the change in native and non-native discrimination as a function of age separately for each method. Therefore, no interactions with method were declared. The categorical factors nativeness and method were contrast-coded. Thus, the intercept estimates the weighted mean effect size at age = 0. The comparison level for method was again CF.

The Q test for moderators was significant [$Q(7) = 29.932$, $p < .001$], showing that the regressors that we included accounted for a substantial proportion of variance. The Q test on residual heterogeneity was also significant [$Q(89) = 117.978$, $p = 0.022$], which indicates that further factors may be needed to account for the remaining variance. The model intercept was significant (*estimate* = .466, *SE* = .112, $z = 4.129$, $p < .001$), suggesting that baseline discrimination levels were significantly different from zero. Additionally, there was a significant interaction between nativeness and age (*estimate* = -.0021, *SE* = 0.0009, $z = -2.316$, $p = .021$), which is consistent with the hypothesis that developmental trends for native and non-native contrasts diverge. The CHT method (*estimate* = .584, *SE* = 0.137, $z = 4.242$, $p < 0.001$) and the HPP method (*estimate* = .178, *SE* = 0.089, $z = 1.998$, $p < 0.046$) showed a significant effect. We carried out a number of follow-up analyses to make sure that these results were robust. For the sake of simplicity, we do not report them in detail here. In one set of follow-ups, we assessed the possibility that method accounted for the results found above. To this end, we separated CHT, HPP and other methods, as well as removed the NIRS results; the same pattern of results found in the general analyses obtained in all three regressions. Additionally, we conducted two analyses declaring either study or sound contrast instead of method as a structuring variable. These also replicated the previous results, as the interaction between nativeness and age remained significant in both of them.
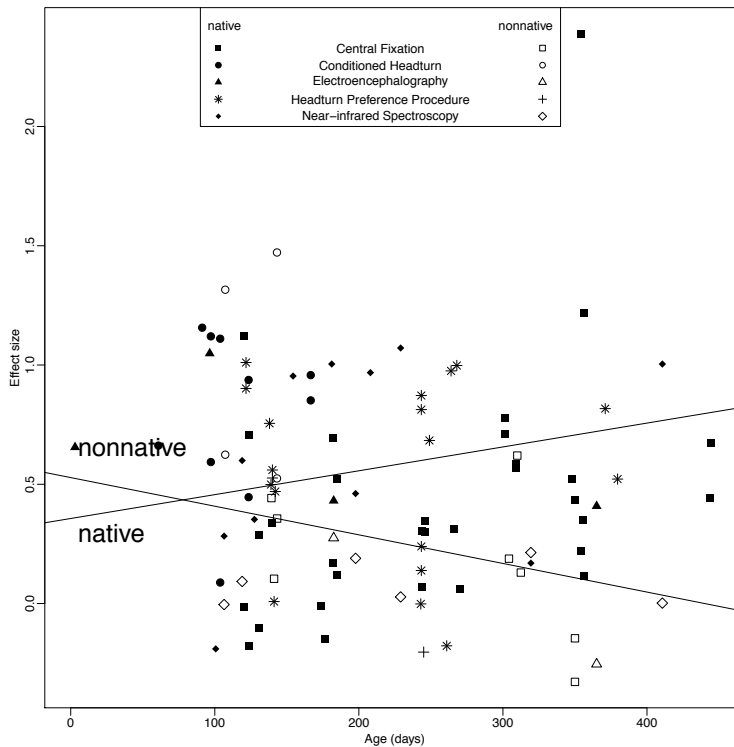
Figure 2: Effect size as a function of age, nativeness, and method. Different methods as well as nativeness are represented with different symbols, as shown in the legend. Lines indicate meta-analytic regression of effect size by age fitted to the relevant set of points. These lines do not take method into account.

3.3 How does discrimination of native contrasts change with age?

We followed up on the divergence in developmental trends by fitting separate models for native and non-native contrasts. For the native contrasts (k = 75), the Q test for moderators reached significance [$Q(5)$ = 18.279, p = .003], suggesting that our regressors were capturing meaningful variation. Additionally, the Q test for residual heterogeneity was also significant [$Q(69)$ = 90.892, p = .040], indicating that a substantial proportion of variance remained to be explained. In this statistical analysis, the baseline discrimination level again differed from zero, because the intercept reached significance (*estimate* = .357, *SE* = .105). The linear slope for age also reached significance (*estimate* = .001, *SE* = .0004, z = 2.247, p = 0.025). Additionally, the methods CHT (*estimate* = .581, *SE* = .163, z = 3.559, p < 0.001), HPP (*estimate* = .250, *SE* = .096, z = 2.618, p = 0.009), and NIRS (*estimate* = .303, *SE* = .163, z = 1.858, p = 0.063)

showed significant effects. We conducted additional analyses to assess if age was better captured with quadratic or cubic trends, but neither of these predictors (derived from a centered version of age) had a significant slope in subsequent polynomial regressions.

3.4 How does discrimination of non-native contrasts change with age?

For the non-native contrasts (k = 22), the test for moderators was significant [Q(5) = 15.397, p = .009], whereas the test for residual heterogeneity was not [Q(16) = 18.047, p = .321], suggesting that our regressors succeeded in structuring the variance in the dependent variable. The baseline level of discrimination for non-native contrasts was above zero, as the intercept was significant (*estimate* = .528, *SE* = .194; z = 2.720, p = .007). The slope for CHT was also a significant predictor (*estimate* = .596, *SE* = .239, z = 2.376, p = 0.018), again indicating that effect sizes with this method are substantially higher. The slope for age did not achieve significance, although the estimate was in the predicted negative direction (*estimate* = -.0012, *SE* = .0008, z = -1.452, p = .146). Quadratic and polynomial regressors based on age did not have a significant estimate in this analysis either.

3.5 At what age does vowel perception become language-specific?

Given the interest that there has been for the age of the emergence for language-specific perception, we sought to provide some rough estimation that could be further investigated in future research. There are several possible ways of approaching the question of the age at which attunement occurs. One is to identify the crossover, given that a linear fit was accurate for at least native perception. The crossover of weighted linear regression lines for native and nonnative effect sizes was at 78 days (2.6 months; cf. Fig. 2). Another possibility is to group effect sizes as a function of the age at which the data had been collected. We divided age-groups into 4 quartiles and carried out weighted regressions exactly as those above (declaring nativeness and method) in each of those quartiles, to assess at which age group native and non-native effect sizes diverged. Nativeness did not have a significant estimate in the first two quartiles (3 to 131 days, and 132 to 185 days) but it was a significant predictor of

effect size in the later two quartiles, namely between 6 months and 10 months (*estimate* = -.661, *SE* = .235; *z* = -2.817, *p* = 0.005), and 10 and 14 months (*estimate* = -.346, *SE* = .120; *z* = -2.885, *p* = 0.004).

4 Discussion

In standard theoretical views (including NLM and PRIMIR), discrimination improves for native vowels within the first year of life, whereas it declines for non-native vowels during that time. We carried out a meta-analysis of developmental infant vowel discrimination literature to assess these predictions. Detailed statistical analyses provided evidence for perceptual narrowing in vowels, in the form of an interaction between vowel nativeness and age. This interaction was due to significantly different slopes for native and non-native sounds. Moreover, effect sizes for native vowel discrimination increased significantly with age. Statistically significant evidence for non-native vowel discrimination was not found, a point to which we return below. As for the age at which attunement occurs, significant differences between effect sizes elicited using native and non-native contrasts were apparent in data collected after, but not much before, 6 months of age.

The first conclusion to be drawn from these data is that there is clear statistical support in current developmental vowel discrimination data, from a variety of paradigms, that perception of native and non-native vowels comes to diverge over the first year of life. This conclusion is not trivial in view of the fact that several null results have been reported for changes in perception with age (and thus language exposure and/or across two language backgrounds; e.g., Polka & Bohn, 1996; Sebastián-Gallés & Bosch, 2009). We believe that our results put both positive and negative previous results in a new, holistic perspective of infant perception, as follows.

To begin with, the presence of an interaction between age and nativeness together with an effect of nativeness in datapoints gathered after 6 months confirm the predictions from perceptual attunement in general, and the description made from the NLM and PRIMIR models in particular. Indeed, enhancement in discrimination of native contrasts had mainly been documented in consonants (Kuhl, Stevens, Hayashi, Deguchi,

Kiritani, & Iverson, 2006; Narayan, Werker, & Beddor, 2009; see also Pons et al., 2012), and thus it is compelling that the present meta-analysis, profiting from the power of studies testing over a thousand infants, was able to confirm that the extrapolation of this process to vowels was justified. At the same time, the lack of a significant slope for non-native datapoints when taken separately cautions us both about the strength of the effect and the design that should be adopted in the future.

This is especially true because the decline in discrimination of non-native vowels has, in a way, been a stronger tenet in the literature on perceptual narrowing in speech sound contrasts. Early findings of a decline in non-native speech perception (Werker & Tees, 1984) led researchers to assume a universal listener who is able to discriminate all speech sound contrasts in the world, and whose ability to do so declines with language exposure. Only recently have reports of improvement began to appear (Kuhl et al., 2006), resulting in the presently predominant view of both decline and enhancement based on language exposure. Our results suggest that the changes in non-native discrimination are more variable and they cannot be distinguished from the null hypothesis independently.

One possibility we considered related to PAM (Best, 1995), a model discussed briefly in the introduction. In it, non-native contrasts are not all difficult to discriminate. On the contrary, those non-native contrasts that can be mapped onto native ones may remain quite discriminable. For instance, both English and German contrast the vowels [i-ɪ], as in the English words 'sheep' and 'ship'. Although these vowels are not exactly the same across the two languages, the German contrast is quite easy to discriminate by native American English listeners because the German [i] maps onto their native English [i], and the German [ɪ] maps onto the English [ɪ]. Thus, one may wonder if some of the non-native results might have been of this 'easy' type. Deciding on this would require a relatively extensive study of the infants' native language and the stimuli used, which could be explored in future research. Nonetheless, we are not confident that this analysis is promising, given that the statistic for remaining variance to be explained was not significant. Instead, we suggest that the current null result for the change with age among non-native effect sizes could be due to insufficient power, because we benefited from only 22

non-native compared to 75 native effect sizes. Therefore, future work including non-native contrasts would be desirable to make the native and non-native samples more comparable.

We propose to take these results as indication that a stronger measure of language attunement would be obtained as the difference between *two* discrimination indices from the same children, one for a native contrast and the other for a non-native one. Such a design has already been successfully employed in the study of consonant attunement (Conboy, Sommerville, & Kuhl, 2008), where investigators cleverly selected a single standard sound as background (voiceless unaspirated /t/) and measured reactiveness to two oddballs. One of the oddballs was contrastive in the infants' native language (either voiced /d/ for Spanish learners, or aspirated /tʰ/, for English learners). Such an oddball paradigm is compatible with both CHT and ERPs. This design would also keep a better handle on random acoustic differences across the contrasts tested; that is, to some extent, one could have feared that nativeness effects might have been obscured if all the native sounds employed happened to be more acoustically dissimilar than non-native contrasts. By testing three sounds in a single continuum or matching the two pairs in acoustic distance, future research would be better able to measure language-specific effects.

Another interesting finding obtained in the present meta-analysis relates to the discussion of whether vowel perception attunes earlier than consonants (e.g., Pons et al., 2012). Our analyses show that perception indeed differs as a function of nativeness as early as 6 to 9 months of age, but not much before this point. We would like to, however, withhold judgment as to whether this age range is earlier for vowels than consonants until the appropriate meta-analysis has been done with consonantal data.

It should be noted that, albeit significant, the effects observed for age are rather small. An analysis on consonantal data would shed light on whether these small attunement effects reflect a minor role of language exposure in shaping perception or rather are peculiar to vowels. As mentioned in the introduction, infants' vowel perception is less pliable in laboratory learning experiments than similar approaches in consonants.

Before concluding, it is relevant to discuss the limitations of the

current study. The first three are inherent to meta-analyses, which are only as good as the data they are based on. Thus, one important limitation relates to sample size for analyzing the effect of potential modulating factors. Indeed, we could not conduct separate analyses within methods, or even include further moderator variables like acoustic distance between stimuli, acoustic distance of non-native stimuli from native categories, as well as further experimental and stimulus characteristics in a quantitative way.

The second, which must also temper our enthusiasm for the attunement effects described above, relates to the possibility that our data reflects a publication bias which is, itself, shaped by theoretical expectations. Notice in particular that the great majority of results came from published studies, with only 3 being manuscripts at this point. In our searches, we have not come across theses or reports in conferences, which are more likely to contain null results that are usually not accepted in peer-reviewed journals. As with any other meta-analysis, this one is only as truthful as the data it includes. In fact, we found statistical evidence for a bias in our data suggesting that small effect sizes were being under-reported. It should be clarified, however, that this is not akin to a publication bias regarding age and nativeness interactions. That is, our sample is biased towards reporting positive *discrimination* results beyond age and the native/non-native status. Nonetheless, bias remains an important consideration that should be kept in mind, particularly given that only developmental studies (i.e., reporting more than one age group) were included.

A third limitation of the present work relates to the 'apples and oranges' problem constitutive of meta-analysis. This type of research necessarily builds on diverse studies, and ours is no exception. We included here a host of different studies, with variable designs, and which load to a variable extent on discrimination skills per se. For example, CHT studies require of the infant not only that she hears the difference between two tokens, but also that she refrains from making a response when no change has occurred, which undoubtedly involves executive abilities beyond linguistic discrimination. Infants tested in CHT also go through a long period of shaping and are highly trained in the task, whereas infants in, for example, NIRS studies will typically simply be

presented with either one or two vowels, with no specific training to perform a discrimination task. This difference could possibly lead to a higher likelihood of finding mixed results, and might be one reason why effect sizes derived from CHT were significantly higher than those derived from other methods.

A related limitation goes beyond the meta-analytic nature of the present research, and relates to the underlying phenomenon under study. Discrimination has been used as an early index of language acquisition, but the precise mechanisms by which this occur remain poorly understood, as evidenced by the differences across the NLM and PRIMIR models of attunement. Primarily due to limitations in the available data, the current meta-analysis has not taken into account factors such as acoustic distance between vowels or acoustic variability induced by number of tokens or talkers, which are certainly relevant for a more differentiated picture of perceptual narrowing. More in general, we cannot speak to the fundamental question of at what level reorganization occurs. There is considerable evidence from adult studies that we retain sensitivity to non-native contrasts (particularly vocalic ones, e.g., Beddor & Strange, 1982). Such findings have led to the hypothesis that language acquisition operates in a 'structure-building' process, and that cross-linguistic differences in perception are driven by top-down influences, for example through biases induced by certain types of tasks (Schouten, Gerrits, & van Hessen, 2003), whereas lower levels of perception remain completely faithful to the signal (but see Chandrasekaran, Krishnan, & Gandour, 2007 for evidence that language experience can shape even the brainstem's response to non-linguistic sounds). Furthermore, attunement in discrimination is clearly only the first of many steps on the road to the native language. Put into a lexical context, infants do not simply discriminate phonemes along the relevant dimensions to make lexical distinctions, but also attend to indexical information like talker identity (e.g., Houston & Jusczyk, 2003; Rost & McMurray, 2010). Even within speech perception alone, infants must also gain a host of other abilities and considerable knowledge at many other levels of representation (e.g., Werker, Fennell, Corcoran, & Stager, 2002, Fernald, Perfors, & Marchman, 2006). These interesting questions go well beyond the present meta-analysis, although they may be amiable to future ones in which more

automatic (i.e., EEG, NIRS) and more "decision-based" (i.e., CHT) discrimination responses can be directly compared.

To conclude, we sought experimental evidence concerning the emergence of native language perception patterns for vowels in infancy. A meta-analysis supported the contention that native and non-native discrimination develop in opposite directions over the first year of life. Moreover, a distinction is evident already by about 6 months of age. In addition to substantiating claims made from mainstream models (NLM and PRIMIR), the present results suggested that a fruitful future avenue of research could employ multiple measures for better capturing infants' budding linguistic knowledge.

Acknowledgements

References

(References marked with * were included in the meta-analysis)

Beddor, P. S. & Strange, W. (1982). Cross-language study of perception of the oral-nasal distinction. *Journal of the Acoustic Society of America, 71*, 1551-1561.

Benders, T. (2013). Learning phonemes from multiple auditory cues: Dutch infants' language input and perception. Manuscript submitted for publication *

Best C.T., McRoberts G.W., LaFleur R., & Silver-Isenstadt J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development, 18*(3), 339-350. doi: 10.1016/0163-6383(95)90022-5

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 167–224). Cambridge, MA: The MIT Press.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler. J. (2004). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16,* 451-459.

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*, 217–243. doi: 10.1177/00238309030460020801 *

Caramazza, A., Chialant, D., Capasso, D., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature, 403*, 428-430.

Cardillo, G. C. (2010). *Predicting the predictors: Individual differences in longitudinal relationships between infant phonetic perception, toddler vocabulary, and preschooler language and phonological awareness*. (Doctoral dissertation). University of Washington.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport, 18*, 1963-1967.

Cheour, M., Alho, K., Ceponiené, R., Reinikainen, K., Sainio, K., Pohjavuori, M.,... Näätanen, R. (1998). Maturation of mismatch negativity in infants. *International Journal of Psychophysiology, 29*(2), 217–26. *

Cheour, M, Leppänen, P.H. T., & Kraus, N. (2000). Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in

infants and children. *Clinical Neurophysiology, 111*(1), 4-16. doi: 10.1016/S1388-2457(99)00191-1

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

Cristia, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics, 39*(3), 388-402. doi: 10.1016/j.wocn.2011.02.004

Cristia, A., Seidl, A, Junge, C., Soderstrom, M., & Hagoort, P. (in press). Predicting individual variation in language from infant speech perception measures. *Child Development.*

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634. doi:10.1136/bmj.315.7109.629

Feldman, N.H., Griffiths, T.L., & Morgan, J.L. (2009). The influence of categories on perception: Explaining the Perceptual Magnet Effect as optimal statistical inference. *Psychological Review, 116*(4), 752-782. doi: 10.1037/a0017196

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding. *Developmental Psychology, 42*(1), 98-116.

Figueras Montiu, M., & Bosch Galceran, L. (2010). Capacidades de discriminación fonética de un contraste vocálico nativo en el prematuro. *Psicothema, 22*(4), 669-676. *

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

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25*(4), 577–88. doi: 10.1037/0012-1649.25.4.577

Houston, D. M., & Jusczyk, P. W. (2003). Infants long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance, 29*(6), 1143-1154. doi: 10.1037/0096-1523.29.6.1143

Jansson-Verkasalo, E., Ruusuvirta, T., Huotilainen, M., Alku, P., Kushnerenko, E., et al. (2010). Atypical perceptual narrowing in prematurely born infants is associated with compromised language acquisition at 2 years of age. *BMC Neuroscience, 11*(1), 88. doi:10.1186/1471-2202-11-88 *

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics, 21*(1-2), 3–28.

Kemler Nelson D. G., Jusczyk P. W., Mandel D. R., Myers J., Turk A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*(1), 111-116. doi 10.1016/0163-6383(95)90012-8

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. doi: 10.3758/BF03212211

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology, 4*(6), 812–822. doi:10.1016/0959-4388(94)90128-7

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). P*hilosophical transactions of the Royal Society of London. Series B, Biological sciences, 363*(1493), 979–1000. doi:10.1098/rstb.2007.2154

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13-F21. doi: 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, 31*(255), 606–608. doi: 10.1126/science.1736364

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE.

Liu, L., & Kager, R. (2011). How do statistical learning and perceptual reorganization alter Dutch infant's perception to lexical tones? *ICPhS XVII,* 1270-1273. Retrieved from http://www.icphs2011.hk/resources/OnlineProceedings/RegularSession/Liu,%20Liquan/Liu,%20Liquan.pdf

Liu & Kager (in preparation a). Infants' perceptual development towards a native vowel contrast. *

Liu & Kager (in preparation b). Bilingual infants' perceptual development towards a native vowel contrast. *

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

Mazuka, Hasegawa, & Tsuji (2013). Development of non-native vowel discrimination: Improvement without exposure. *Developmental Psychobiology.* Advance online publication. doi: 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. *The Journal of Neuroscience, 27*(2), 315–21. doi: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. *

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

Narayan, C., Werker, J. F., & Beddor, P. S. (2009). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Developmental Science, 13*(3), 407-420. doi: 10.1111/j.1467-7687.2009.00898.x

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

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

Polka, L., & Werker, J. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 421–435. doi: 10.1037/0096-1523.20.2.421

Pons, F., Mugitani, R., Amano, S., & Werker, J. F. (2006). Distributional learning in vowel length distinctions by 6-month-old English infants. Presented at the *International Conference on Infant Studies*; Kyoto, Japan (abstract).

Pons, F., Sabourin, L., Cady, J. C., & Werker, J. F. (2006). Distributional learning in vowel distinctions by 8-month-old English infant. Presented at the *28th Annual Conference of the Cognitive Science Society*; Vancouver, BC, Canada (abstract).

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), 965–76. doi: 10.1111/j.1467-8624.2012.01740.x *

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of non-contrastive phonetic variability in early word learning. *Infancy, 15*(6), 608-635. doi: 10.1111/j.1532-7078.2010.00033.x

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

Sato, Y., Mori, K., Furuya, I., Hayashi, R., Minagawa-Kawai, Y., & Koizumi, T. (2003). Developmental changes in cerebral lateralization during speech processing measured by near infrared spectroscopy. *Japanese Journal of Logopedic Phoniatrics, 44*, 165–171. doi: 10.5112/jjlp.44.165. *

Schouten, B., Gerrits, E., & van Hessen, Arjan (2003). The end of categorical perception as we know it. *Speech Communication, 41*, 71-80.

Schwarzer, G. (2012). meta: Meta-Analysis with R. R package version 2.1-3. Retrieved from http://CRAN.R-project.org/package=meta

Tsuji, S., & Cristia, A. (2014). *InPhonDB: A developing meta-analysis of infant vowel perception*. Manuscript in preparation.

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), 874–87. doi: 10.1111/j.1467-7687.2009.00829.x *

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48. URL http://www.jstatsoft.org/v36/i03/

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

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

Werker, J. F., & 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. doi: 10.1016/S0163-6383(84)80022-3

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology, 34*(6), 1289–309.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words. *Infancy, 3*(1), 1-30.

Werker, J. F., Polka, L., & Pegg, J. E. (1997). The conditioned head turn procedure as a method for testing infant speech perception. *Early Development and Parenting, 6*(3-4), 171–178. doi:10.1002/(SICI)1099-0917(199709/12)6:3/4<171::AID-EDP156>3.0.CO;2-H

The more, the better? Behavioral and neural correlates of frequent and infrequent vowel exposure

Chapter 5

Based on:

Tsuji, S., Fikkert, P., Minagawa-Kawai, Y., & Cristia, A. (2014). *The more, the better? Behavioral and neural correlates of frequent and infrequent vowel exposure*. Manuscript in preparation.

Abstract

A central assumption in the perceptual attunement literature holds that exposure to a speech sound contrast leads to improvement in native speech sound processing. Yet, whether the amount of exposure matters for this process has not been put to a direct test. To elucidate behavioral and neural indicators of frequency-dependent perceptual attunement, we compared 5-8-month-old infants' processing of tokens containing a highly frequent [hɪt-he:t] and a highly infrequent [hʏt-hø:t] Dutch vowel contrast in a behavioral visual habituation paradigm and using near-infrared spectroscopy. Behavioral and cerebral hemodynamic responses provided converging evidence that infants discriminated both contrasts, but that discrimination ability was not modulated by frequency of exposure. Infants' neural responses also showed a tendency to be processed in a left-dominant network, possibly indicating increasingly more linguistic processing of these native contrasts. The degree of left-lateralization in response to the frequent and the infrequent contrast was affected differently by infant age, however, the directionality of the effect could not be clarified in post-hoc tests. These results illustrate the importance of taking parameters like the nature of the contrast and the amount of exposure into account when assessing perceptual attunement.

1 Introduction

During their first year of life, infants' ability to discriminate non-native contrasts declines, while it is maintained or improved for native contrasts. In this developmental process, often referred to as *perceptual attunement*, frequency of exposure is assumed to play a critical role: the more tokens of a given speech sound category infants hear, the more evidence they can accumulate for that particular category in their native language. This assumption is put to a direct test in the present study.

Mainstream language acquisition models propose that infants' perception becomes specialized to the categories present in the native input. A variety of mechanisms have been proposed to account for how exactly this is implemented. One possibility is that the perceptual space gets 'warped' through the formation of prototypes that act as magnets; as a consequence, discrimination around frequently heard tokens (prototypes) declines because perceptual space is shrunk there compared to the regions of perceptual space where few tokens are heard (WRAPSA, Jusczyk, 1993; NLM-e, Kuhl et al., 2008). Another possibility is that infants use frequency *distributions* rather than prototypes. The distributional learning account proposes that infants track the frequency of occurrence of acoustic correlates. When there is a single category, these distributions will tend to be unimodal, whereas multiple categories could lead to multimodal distributions. Infant discrimination would then improve for speech sounds that belong to two different modes compared to those that lie in the same mode, indicating enhanced between- and reduced within-category discrimination (Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002). Regardless of the specific mechanism, all current proposals assume that infants profit from accumulating evidence for a given category. Accordingly, frequency of occurrence, and not merely presence versus absence of a contrast, plays a key role.

And yet most experimental studies on natural speech have captured developing speech sound perception in a rather categorical way, namely by comparing discrimination of (non-native) contrasts with zero exposure versus (native) contrasts with above-zero exposure. The seminal study by Werker and Tees (1984) was the first to demonstrate how language exposure alters speech sound discrimination during the first year of life,

showing that English-learning infants' ability to discriminate two non-native consonant contrasts (a Hindi dental-retroflex contrast [t̪- ʈ], and a Nthlakampx glottalized velar versus uvular contrast [k'-q']) declined between a group aged 6-8 months and a group aged 10-12-months. By contrast, Hindi- or Nthlakampx-learning 10-12-month-old infants continued to discriminate their respective native contrast. Evidence for perceptual attunement was subsequently also reported with regard to vowel perception. Kuhl, Williams, Lacerda, Stevens, and Lindblom (1992) found that 6-month-old English-learning infants failed to discriminate between prototypical and less prototypical tokens of the native vowel [i], whereas they succeeded in making an equivalent discrimination between non-native tokens of the Swedish vowel [y]. The reverse pattern of discrimination was found in Swedish-learning infants, providing evidence for language-dependent differences in within-category structure. Finally, evidence for enhancement in the discrimination of native contrasts was reported by demonstrating that English-learning infants improved in their discrimination of the native [l-r] contrast at the same time as Japanese infants' ability to discriminate this non-native contrast declined (Kuhl et al., 2006). These basic patterns of decline, changes in within-category structure, and enhancement have been shown to generalize to a large variety of contrasts and languages. Further, a number of studies using the event-related potential (ERP) technique have reported that neural markers of early speech sound discrimination conform to the pattern reported in behavioral studies, such that mismatch responses become weaker for non-native, and are maintained or become stronger for native contrasts (cf. Tsuji & Cristia, 2013a, for review).

Neuroimaging research adds one more piece to the puzzle. Near-infrared spectroscopy (NIRS) measures changes in blood oxygen level in cortical regions as an index of neural activity. Two measures of relevance are the change in blood oxygenation (measured in the superior temporal gyrus, STG, bilaterally), which corresponds to a change detection response, and a laterality index (L-R)/(L+R)), which measures the relative left hemisphere advantage and is assumed to reflect increasingly linguistic processing. As in the behavioral and ERP research, the focus is on native versus non-native contrasts. Broadly speaking, this work shows heightened change detection as well as emergent left dominance for the

former and not the latter (see Minagawa-Kawai, Cristià, & Dupoux, 2011, for a theoretical review and Tsuji & Cristia, 2013a, for an empirical review). For example, Minagawa-Kawai, Naoi, Nishijima, Kojima, and Dupoux (2007a) found increased left-hemisphere activation as well as left-dominance at 7-8 months (but not yet at 3-4 months) for the native Japanese contrast between [i] and [o], while no such development was attested for the non-native contrast between [o] and [u].

The above studies provide important evidence that the presence versus absence of exposure has an impact on developing speech sound perception. None of them, however, has directly tested the key assumption that the *amount* of exposure matters for attunement. Only two studies have investigated the influence of input frequency on perception by comparing discrimination of native contrasts with little exposure (low frequency in the input) to discrimination of native contrasts with much exposure (high frequency in the input). A study of English infants' discrimination of the same two non-native contrasts that have been used in Werker and Tees (1984), the coronal Hindi dental-retroflex and the dorsal Nthlakampx glottalized velar-uvular contrast, exploited the fact that coronal [t] is more frequent than dorsal [k] in English (Anderson, Morgan, & White, 2003). While 6.5-month-olds discriminated both non-native contrasts, 8.5-month-olds only discriminated the dorsal one. The authors suggest that frequent native categories get robust earlier, acting as attractors for close non-native contrasts, for which discrimination in turn declines. Pons, Albareda-Castellot, and Sebastián-Gallés (2012) in turn focused on frequency-dependent changes in native discrimination, assessing discrimination of a contrast consisting of one frequent and one infrequent vowel. They showed that both Catalan-and Spanish-learning 12-month-olds discriminated [i] and [e] only if the change went from the less frequent to the more frequent speech sound in their respective native language. These perceptual asymmetries indicate that frequent speech sounds can also act as attractors to less frequent native speech sounds, actually reducing contrast discrimination in one direction.

The two studies mentioned above imply that the frequency of native speech sounds influences the decline in non-native discrimination, and that asymmetries in native speech sound frequency can lead to difficulties

in native discrimination. What has not been focused on in either of these studies, however, is the impact of input frequency on *improvement* in native discrimination: the ability to discriminate more frequent native speech sound contrasts (where both speech sounds are frequent) should improve earlier than the ability to discriminate less frequent native speech sound contrasts (where both speech sounds are infrequent). This is the central prediction tested in the current study.

1.1 The current study

The central aim of the current study was to compare infants' discrimination of a frequent and an infrequent native speech sound contrast, critically assessing whether input frequency had an influence on discrimination performance. In addition to making this central comparison, the current study extended previous work in three aspects. To control for the fact that basic discrimination ability might differ between contrasts, we chose two contrasts that were matched on their respective acoustic distance; to account for the fact that age, usually tested categorically, should have a continuous influence on discrimination ability, we recruited infants in a wide enough age range to allow us to include infant age as a continuous predictor; and to seek converging evidence from multiple methods, we compared indices of discrimination and more linguistic processing from behavioral and neuroimaging methods.

As explained in more detail further below, we matched acoustic distances by selecting speech sound contrasts that differed in the same phonological features, and by assessing their discriminability scores in a multi-class classifier model. For selecting an appropriate age range, we referred to a recent meta-analysis on published studies, which confirmed six months as the critical age for perceptual attunement in vowels (Tsuji & Cristia, 2013b). We therefore decided to test infants in a narrow age range spread around this critical age.

Finally, we decided to include three dependent measures in our study, derived from a behavioral study (Experiment 1) and a NIRS study (Experiment 2). In both experiments, we used a variant of a stimulus alternation paradigm, in which infants' responses to trials with repetitions of the same speech sound (non-alternating trials) are

compared to their responses to trials in which the same speech sound is alternated with a novel speech sound (alternating trials) (cf. Procedure sections for details). The frequent and infrequent contrasts were assessed in a between-participants design in Experiment 1, because infants are often tired even after a single habituation phase, and thus it is not feasible to test infants with two pairs of sounds behaviorally. An advantage of NIRS, however, is that no habituation phase is required, and infants are only required to passively listen to the presented speech sounds while being silently entertained with toys. Therefore, it is possible to measure infants' perception of both contrasts in a within-participants design.

The dependent variable in Experiment 1 was looking time differences between non-alternating and alternating trials as an index of discrimination. There were two dependent variables in Experiment 2: bilateral differences in blood oxygenation as an index of discrimination, and lateralization as an index of more linguistic processing. We predicted a main effect of frequency and of age for all three measures: Infants would be better at processing the frequent compared to the infrequent contrast, and infants would be better at processing with age.

2 Experiment 1

2.1. Participants

Forty-one monolingual Dutch full-term infants were included in the final analysis. Twenty-one of these infants were assigned to the frequent condition (8 females, mean age = 6.10 months, range = 4.80-6.80 months), and another twenty were assigned to the infrequent condition (11 females, mean age = 6.57 months, range = 5.00-7.66 months). Eleven more infants were excluded from data analysis due to fussiness (frequent: 5, infrequent: 2), failure to habituate (frequent: 2), experimenter error (infrequent: 1), dialectal language background (infrequent: 1).
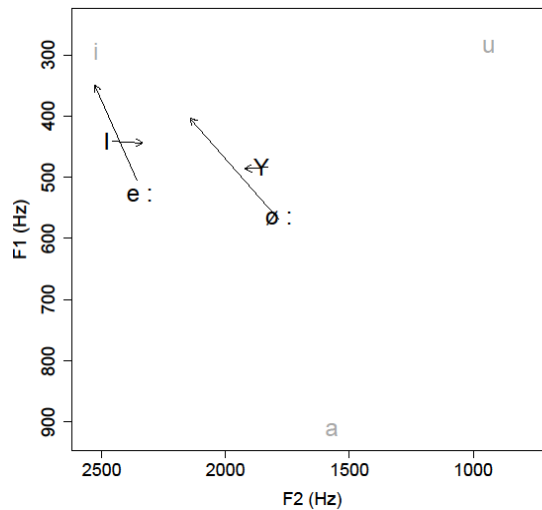
Figure 1. Formant values of experimental tokens. Endpoints of arrows represent mean formant values in the first and fourth quantiles of the respective vowel (also cf. Table 2). Average values for the point vowels [a,i,u] pronounced by 20 female native speakers of Standard Dutch (Adank, Hout, & Smits, 2004) are added in gray for reference.

2.2 Stimuli

Two pairs of Dutch vowels, [ɪ - e:] and [ʏ - ø:], were selected. They were selected as to differ maximally in their token frequencies, but minimally in their acoustic/perceptual characteristics. As illustrated in Table 1, [ɪ] and [e:] are several times more frequent than [ʏ] and [ø:] (counts based on two corpora of spoken Dutch: CGN, Oostdijk, 2000; IFA corpus, van Son, Binnenpoorte, van den Heuvel, & Pols, 2001). While the pairs thus differ markedly in frequencies, they have similar acoustic characteristics in that they both consist of a short close vowel and a long diphtongized close-mid vowel. As shown in Figure 1, these tokens are relatively close in F1/F2 space.

In order to measure the similarity of their acoustic/perceptual characteristics, a multi-class classifier model based on mel frequency cepstral coefficients (MFCCs, e.g., Kirchhoff & Schimmel, 2005) derived from the original tokens in the corpora assessed the discriminability of the four vowels, showing that the discriminabilities of the short vowels,

and the discriminability scores of the diphtongized vowels are comparable (cf. Table 1).

Table 1. Frequency and discriminability of chosen vowels. Frequency counts are token frequencies derived from CGN (Oostdijk, 2000) and IFA corpus (van Son, Binnenpoorte, van den Heuvel, & Pols, 2001), and discriminability scores represent F1 scores derived from a multi-class classifier model.

Contrast	Vowel	Frequency	Discriminability
Frequent	ɪ	10489	0.514
	e:	10087	0.652
Infrequent	Y	629	0.413
	ø:	2533	0.661

Table 2. Acoustic properties of experimental tokens. Pitch and formant values were measured over the vowel part.

Contrast	Stimulus	Length (ms)	Mean pitch (Hz)	F1 1st quantile (Hz)	F1 4th quantile (Hz)	F2 1st quantile (Hz)	F2 4th quantile (Hz)
Frequent	hɪt	411	248	441	444	2456	2335
	he:t	520	255	505	350	2356	2527
Infrequent	hYt	401	262	484	486	1830	1920
	hø:t	521	266	568	404	1786	2142

Experimental tokens were recorded in an infant-directed register by a female native speaker of Dutch. The vowels were embedded in a [hVt] context. One token of each vowel was chosen based on similarity in length and pitch characteristics (cf. Table 2 & Figure 1). For each of the two conditions (frequent, infrequent), two lists of stimuli, a non-alternating and an alternating list, were created. Non-alternating lists contained 17 repetitions of the stimulus including the short vowel ([hɪt] in the frequent condition, [hYt] in the infrequent condition). Alternating lists contained 16 stimuli, and repeatedly alternated between the stimulus

including the diphtongized vowel ([he:t] in the frequent condition, [hø:t] in the infrequent condition) and the respective stimulus containing the short vowel. These lists always started with the stimulus containing the short vowel. The inter-stimulus interval (ISI) was 750 ms for all four lists. The alternating lists contained 16 rather than 17 stimuli in order to keep total list length as constant as possible, because the diphtongized vowels were longer than the short vowels (cf. Table 2). The total list length for the non-alternating list in the frequent condition was 19.79 s; for the non-alternating list in the infrequent condition was 19.42; for the alternating list in the frequent condition was 19.49 s; and for the alternating list in the infrequent condition was 19.61 s.

2.3 Procedure

To assess infants' discrimination abilities, the hybrid visual habituation method (Houston, Horn, Qi, Ting, & Gao, 2007) was implemented with the LOOK software (Meints & Woodford, 2008). Infants were seated in a car seat on their caregiver's lap facing a TV screen. Caregivers were asked not to interact with their infant during the experiment, and both caregiver and experimenter wore headphones with masking music during the course of the experiment. Half of the infants was assigned to the frequent condition, and the other half of the infants was assigned to the infrequent condition. The experiment consisted of a habituation and a test phase. Each trial started with a silent attention getter (a video of a laughing infant). Once the infant looked at the screen, a picture of a colorful bull's eye appeared on screen. During habituation, the respective non-alternating list was repeatedly presented to infants until the habituation criterion (50% decrease in looking times compared to the first trials over a sliding window of three trials) was reached, or infants had reached a maximum number of 24 trials. During the test phase, infants were presented with 10 non-alternating and 4 alternating trials in pseudo-random order. Four different trials were created, in each of which the first trial was always non-alternating and the second trial alternating, and no two alternating trials ever followed each other. A trial was terminated when the infant looked away for more than 2 seconds. All trials were visually accompanied by the bull's eye picture. Between trials, the movie of a laughing infant appeared to capture infants' attention. The

next trial was started once the infant looked at the screen. Looking times were coded online by a trained experimenter.

2.4 Results
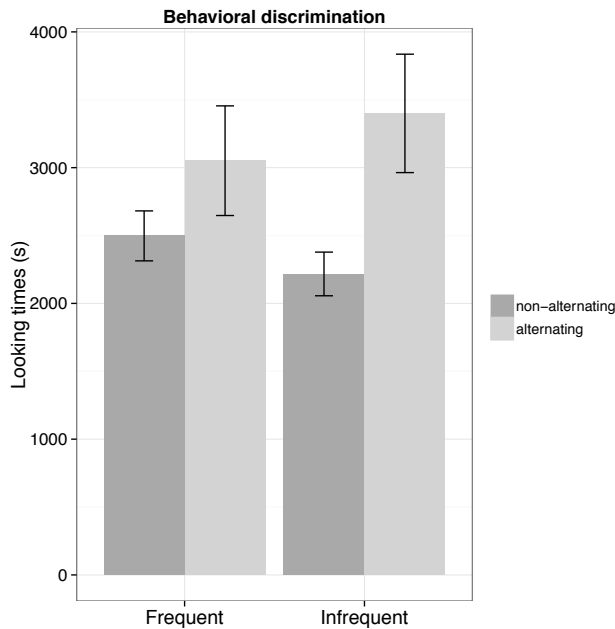
Figure 2. Looking times by Trial Type and Condition. Error bars represent 95% confidence intervals.

A linear mixed effect model (lme in nlme; Pinheiro, Bates, DebRoy, Sarkar, & Team, 2012) with the within-subject predictors Condition (frequent, infrequent), Trial Type (non-alternating, alternating), the between-subject predictor Age, and the random effects of infant was fitted (full model: looking time ~ Condition * Trial Type * Age , random = (~1|Infant/Condition/TrialType), number of observations: 82). The model showed a significant main effect of trial type (F = 31.68, $p < .001$), with a higher amount of looks to alternating (*mean* = 3.64 s) than to non-alternating (*mean* = 2.51 s) trials (Figure 1). None of the other effects reached significance (Condition: $F = 0.14$, $p = .712$; Age: $F = -0.30$, $p = .584$; Condition x Trial Type: $F = 2.67$, $p = .111$; Trial Type x Age: $F = 0.50$, $p = $

.484; Condition x Age: F = 0.06, p = .809; Condition x Trial Type x Age: F = 0.02, p = .900).

2.5 Discussion

Based on the prediction that the *amount* of exposure matters for perceptual attunement, we hypothesized that the frequent native contrast would be discriminated better earlier compared to the infrequent contrast. However, no evidence for frequency-dependent or age-dependent differences in the discrimination of the two native speech sound contrasts was found in Experiment 1. This lack of any effect was rather unexpected, because our predictions had been derived from central assumptions in the speech sound acquisition literature, and we had chosen a strong frequency manipulation to put them to test (cf. Table 1). We therefore sought to evaluate complementary measures for infants' processing of the two vowel contrasts. NIRS would possibly provide a more sensitive measure of discrimination, as well as of left-lateralization as a measure for increasingly linguistic processing.

As mentioned in the Introduction, bilateral differences in blood oxygenation as a measure of discrimination, and lateralization as a measure of more linguistic processing, have been used in previous NIRS studies on speech sound acquisition (e.g., Minagawa-Kawai, Mori, Naoi, & Kojima, 2007b). NIRS might provide a more sensitive measure of discrimination, because within-subject designs can increase statistical power, and because cerebral responses do not rely on the recovery of attention. That infants' overall attentiveness in the test phase of Experiment 1 was rather moderate can be illustrated by the fact that they looked on average less than 4 s (of a possible 20 s) even to alternating trials. A measure circumventing to rely on infants' attention might therefore capture more sensitive responses.

In addition to this alternative measure of discrimination, NIRS would enable us to measure differences in infants' developing left-lateralization between the frequent and the infrequent contrasts. Differences in developing lateralization can be observed even when discrimination is stable over development, as has for instance been reported in a study on the processing of Japanese lexical pitch accent (Sato, Sogabe, & Mazuka, 2010a). Infants' behavioral discrimination of

pitch contrasts followed a pattern of maintenance, with equally successful discrimination at 4 and 10 months of age. However, their hemodynamic responses revealed a developmental difference in lateralization such that only 10-month-olds processed these native contrasts in a left-dominant network. Based on these considerations, we assessed another group of infants on their processing of the same frequent and infrequent vowel contrasts using NIRS in Experiment 2.

3 Experiment 2

3.1 Participants

Thirty-four infants were included in the final analysis (21 females, mean age: 6.71 months or 204 days, range: 5.45 - 8.48 months or 166 - 258 days). These infants were monolingual Dutch, full-term, in good health and without developmental, language or hearing problems according to parental report. Another 22 infants were excluded from the analysis for the following reasons: data loss resulting in less than 4 usable trials in each condition: 17; equipment error: 5. Caregivers signed a consent form approved by the local ethical committee (Commissie Mensgebonden Onderzoek Arnhem-Nijmegen, The Netherlands).

3.2 Stimuli and Paradigm

Our experimental design closely followed Minagawa-Kawai et al. (2007b). The exact same stimuli used in the behavioral study were employed here, with only two changes. First, the ISI was set to 1.25 s to match previous NIRS studies (Minagawa-Kawai et al., 2007b). Second, there were two versions of the non-alternating lists, one with 11 tokens (duration 18.79 s), the other with 12 tokens (duration 20.5 s), this variation in duration serving to jitter the alternating blocks. In alternating blocks, the two frequent, or the two infrequent, 11 tokens were presented in pseudo-random order with equal probabilities every 1.25 s (duration 19.35 or 19.47 s), with a block always starting with a change token.

Unlike Experiment 1, each infant in the present experiment was presented with both conditions, frequent and infrequent (order counterbalanced across participants). Infants were presented alternately

with non-alternating and alternating blocks for a total of 8 pairs per condition.
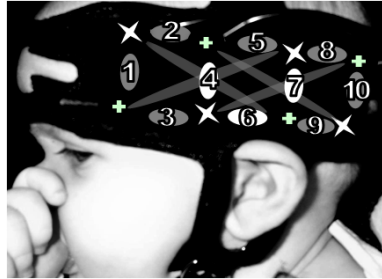

Figure 3. Probe array design showing distribution of sources and detectors. Crosses indicate detectors and stars sources. White channels indicate bilateral region of interest.

3.3 Equipment and data acquisition

Infants were seated on their caregiver's lap in a sound-proof booth and passively listened to the auditory stimuli. An experimenter silently entertained infants with toys during the course of the experiment. Both caregiver and experimenter were wearing headphones with masking music. Stimuli were presented with Psyscope B55 (Bonatti, 2009). The UCL-NTS fNIRS system (Department of Medical Physics and Bioengineering, UCL, London, UK) was used, which continuously emits near-infrared light of two wavelengths, 670 and 850 nm (for further technical details, see Everdell, Coulthard, Crosier, & Keir, 2005; Minagawa-Kawai, Cristià, Vendelin, Cabrol, & Dupoux, 2011). Probes and detectors were positioned on a 2 by 4 grid on each of the left and right pads, thus defining a total of 10 channels between optodes separated by 25 mm (see Figure 3), and 4 more between non-adjacent optodes. In the analyses, we focused on a region of interest (ROI) defined prior to the study and on the basis of previous research and anatomical considerations: channels 4, 6, 7. We used anatomical landmarks to align the bottom of the pad with the T3-T5 line of the 10/20 system, and used the ear as a midpoint reference (see Figure 3).

3.4 Data preprocessing and analysis

Light intensity signals were converted into oxy- and deoxy-Hb concentration with the modified Beer Lambert law. Data were analyzed by applying a general linear model (GLM) to the non-artifacted data of each channel, including nuisance regressors for long-distance, slow trends, and baseline changes as follows. Slow trends were captured through sine and cosine regressors (for each time-stretch of non-artifacted data longer than 20 s up to the whole duration of the experiment). The data were band-pass filtered between 0.02 and 0.7 Hz only for the following unsupervised artifact detection procedure. Artifacted data was identified as the time stretches in which concentration levels changed by more than 0.15 millimolars per millimeter (mM.mm) within 100 ms (time between two successive samples) in the total-Hb averaged over all channels associated with a given probe (Gervain, Macagno, Cogoi, Peña, & Mehler, 2008; Kotilahti et al., 2010). Artifacted stretches were silenced by giving them a weight of zero in the subsequent regression. A boxcar regressor for each new stretch of non-artifacted data was introduced. If there was less than 20 s of unartifacted data between two artifacted regions, this was also silenced, as it is difficult to estimate the hemodynamic response independently from any baseline level changes accompanying an artifact in such short stretches. The data of a channel was altogether excluded from analysis if there was unartifacted data for fewer than 4 trials in a given condition and infant. This did not affect the number of trials per condition (frequent: *mean* =7.01, infrequent: *mean* = 7.02), or per presentation order (frequent first: *mean* = 7.15 trials, infrequent first: *mean* = 6.85 trials).

In addition to these nuisance parameters, we declared a regressor based on the standard finite impulse response function (FIR) to estimate specifically concentration changes associated with change detection. The first dependent variable, aimed to capture differences in infants' discrimination response, consisted of the beta values obtained from a GLM where the FIR had been convolved with the duration of stimulation for each condition separately (similar to Cristia et al., in press). In accordance with previous infant studies, this analysis was based on oxy-Hb (see Lloyd-Fox, Blasi, & Elwell, 2010 for a discussion). The betas derived from the overall GLM fit were analyzed with a linear mixed effects

model, using the same method as in Experiment 1. The two categorical predictors Condition (frequent, infrequent) and Hemisphere (left, right) and the continuous predictor Age were included as fixed effects together with their interactions, and infants and channels were included as random effects (beta ~ Condition x Hemisphere x Age + (1|Infant) + (1|Channel), number of observations: 349). P-values were obtained by likelihood ratio tests on a forward-fitted model (cf. Field et al., 2012).

In order to assess differences in lateralization, the laterality index served as the dependent variable in a second analysis. It was calculated following Minagawa-Kawai et al. (2007b) using the formula (L-R)/(L+R). L and R represented the maximal changes in total Hb (oxy + deoxy) among the ROI channels in the left and right hemisphere, respectively. The laterality index was again analyzed with a linear mixed effect model and categorical predictor Condition (frequent, infrequent), the continuous predictor Age, and their interaction. Since the laterality index had been calculated based on the maximum value across channels, only infants were included as random effects (laterality index ~ Condition x Age + (1|Infant)).

3.5 Results

In the first analysis on differences in overall fit, we measured the overall discrimination effect by inspecting the intercept of the full model. The intercept was significant (F = 13.04, p <.001), indicating that there was an overall difference in processing of non-alternating and alternating trials (cf. Figure 4, left panel). However, no significant effects of any of the predictor variables or their interactions were obtained (Condition: F = 0.71, p = .404; Hemisphere: F = 0.20, p = .657; Age: F = 0.33, p = .578; Condition x Hemisphere: F = 0.96, p = .331; Condition x Age: F = 0.66, p = .421; Hemisphere x Age: F = 0.02, p = .884; Condition x Hemisphere x Age: F = 1.63, p = .207). The average activation change in the left hemisphere was β = 0.013 in the frequent condition, and β = 0.014 in the infrequent condition. In the right hemisphere, the average activation change was β = 0.015 in the frequent condition, and β = 0.006 in the infrequent condition.
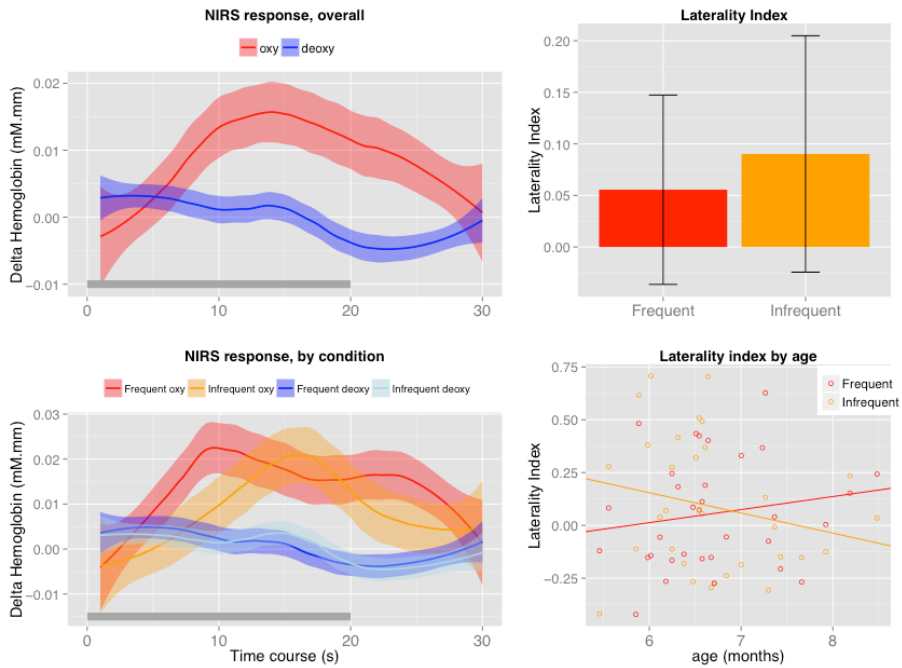
Figure 4. Left panels show the time course of hemodynamic responses, separated by oxy and deoxy Hb (upper panel), and by condition (lower panel). Lines are smoothed with a Gaussian filter for visualization, and shaded areas represent the 95% confidence interval. Grey bar at the bottom indicates 20 s time-window of stimulation. Right panels display the laterality index by condition (upper panel) and by age and condition (lower panel). Error bars represent 95% confidence intervals, and regression lines are based on intercepts and slopes of reported post-hoc regression analyses (cf. text for details).

The second analysis on the laterality index revealed no significant model intercept ($F = 2.85$, $p = .101$), indicating that there was no overall tendency for left-dominant activation. There were no significant main effects (Condition: $F = 0.72$, $p = .404$; Age: $F = 0.04$, $p = .836$), but a significant interaction effect between Condition and Age ($F = 5.19$, $p = .030$), such that left-lateralization increased with age for the frequent contrast, but decreased with age for the infrequent contrast (cf. Figure 4, right panel). To follow up on this effect, two models for the effect of age on laterality index were fit, one of which included the data from the frequent Condition, and the other included data from the infrequent Condition. The effect of Age reached significance in neither of these post-hoc analyses

(frequent: $b = 0.055$, $t = 1.21$, $p = .234$, infrequent: $b = -0.068$, $t = -1.22$, $p = .231$).

3.6 Discussion

Experiment 2 focused on 5-8 month-old infants' hemodynamic responses to vowel contrasts that are highly frequent or infrequent in their language, measuring frequency-dependent differences in discrimination and lateralization.

The discrimination measure (differences in bilateral blood oxygenation) showed the same pattern of results as the behavioral measure in Experiment 1: an overall discrimination effect, but no effect of frequency or age. By contrast, lateralization was affected by these two predictors, indicated by an interaction between frequency condition and infant age. This effect is, however, difficult to interpret, because follow-up tests did not indicate which factors contributed to this interaction. Descriptively, left-lateralization increased with age for the frequent contrast, which is consistent with our predictions: The more exposure, the earlier is a contrast processed linguistically. At the same time, left-lateralization decreased with age for the infrequent contrast, a finding not compatible with our predictions.

4 General Discussion

The present study assessed the influence of frequency of exposure on perceptual attunement, hypothesizing that frequently heard speech sound contrasts would lead to earlier perceptual attunement than infrequently heard contrasts. Given that models of language acquisition ascribe the frequency of exposure a central role in perceptual attunement, documenting such an influence would provide key evidence for the assumed mechanisms.

The overall evidence turned out to be rather weak, however. Both behavioral and neural measures of discrimination indicated that infants discriminated both contrasts equally well, regardless of frequency or infant age. In addition, the degree to which lateralization developed with age was the only instance in which responses to the infrequent and the frequent contrast significantly diverged. Analyses separating the frequent

and infrequent contrast found that left-dominance increased with age for the frequent contrast, but decreased with age for the infrequent contrast. Both of these tendencies did not approach significance, however, hampering the interpretation of this interaction.

How can the lack of any frequency-dependent differences in discrimination be interpreted in light of the perceptual attunement literature? One possible explanation for infants' ability to discriminate both contrasts well is that our choice of age-group was inappropriate, such that infants' discrimination ability for both contrasts had already improved prior to testing. However, we carefully selected the age-range based on a meta-analysis on extant studies of perceptual attunement in vowels (Tsuji & Cristia, 2013b), in which statistical evidence for a divergence between discrimination responses for native and non-native vowels was found from around 6 months of age onwards. We could therefore assume that the 5-8-month-old infants assessed in the current study were in the middle of perceptual attunement, and it was unlikely that they had already fully attuned especially to the infrequent contrast.

Instead, we propose that the chosen contrasts were inherently easy to discriminate for young infants, who needed to maintain rather than improve discrimination of these contrasts over the course of perceptual attunement. Although the contrast pairs had a relatively small spectral distance, differences in their duration and dynamics (short versus long and diphtongized, cf. Figure 1 and Table 2) might have rendered them salient. Assuming that discrimination of these contrasts was already fully in place prior to perceptual attunement, it makes sense that differences in exposure did not lead to additional improvement. Similar findings have been reported in two recent distributional learning studies, which found that infants were equally able to discriminate a vowel quality contrast ([i-e], Pons, Sabourin, Cady, & Werker, 2006a) or a vowel length contrast ([ɛ-ɛ:], Pons, Mugitani, Amano, & Werker, 2006b) after exposure to a unimodal or bimodal distribution. The authors suggest that vowels might not be affected by distributional learning to the same degree as consonants, for which successful distributional learning has been documented (Maye et al., 2002; 2008; cf. Introduction).

Assuming that the contrasts tested in the present study were unsuitable for measuring improvement in discrimination (regardless of

whether this would hold for vowels in general or for specific, salient contrasts), we consequently hypothesize that frequency does influence developing discrimination abilities of less salient contrasts. Testing this hypothesis is not a trivial task, however, because this would require selecting non-salient vowel contrasts differing markedly in frequency, but not in acoustic distance, a task we did not succeed in even in Dutch, a language with a relatively large number of 16 vowels (whereas an average vowel inventory in the world's languages contains between 5 and 6 vowels; Maddieson, 2013).

While the contrasts chosen for the current study might not have been suitable to assess changes in discrimination, our measure of lateralization was intended to tap more linguistic processing. It has been documented that infants show developmental changes in left-dominance for native contrasts even where discrimination is in place early on (cf. Sato et al., 2010a). Given that infants presumably had had significantly more exposure to the frequent compared to the infrequent contrast, it was conceivable that these developmental changes could be observed earlier in response to the frequent contrast. While there was no overall evidence of left-dominant processing, lateralization was directly affected by frequency. Lateralization in the frequent and the infrequent condition was, however, differently affected by age such that left-dominance increased with age for the frequent condition, while it decreased with age for the infrequent condition.

The lack of an overall tendency for left-dominant processing in our sample can be reconciled with previous NIRS studies, which have reported left-dominance in slightly older age-groups, namely from 7-8 months (Minagawa-Kawai et al., 2007a) or from 11-12 months (Sato et al., 2003) onwards. Such an account would also be compatible with the descriptive tendency of increasing left-lateralization with age for the frequent contrast, leading to the assumption that this trend would further increase in older infants. Testing more infants and widening the age-range to both sides might therefore provide stronger statistical evidence for this predicted tendency.

However, provided we were indeed capturing infants in the course of developing left-lateralization for the frequent contrast, responses to the infrequent contrast might not be observable yet, but should principally go

into the same direction. Contrary to this assumption, infants' responses to the infrequent contrast showed the opposite descriptive tendency, namely decreasing left-lateralization, which is equivalent to increasing right-lateralization. Finding a tendency of a right-lateralized response to vowels early in acquisition would not be entirely incompatible with previous findings, since slow acoustic transitions (such as pitch and prosody) have been found to elicit right-dominant responses, while fast changing sounds (such as consonants) have been found to elicit left-dominant responses, with vowels somewhat in between (cf. Minagawa-Kawai et al., 2011, for an overview). However, since the contrasts were well-controlled for acoustic differences, and the discrimination results did not indicate any differences in strength of discrimination, it is unlikely (although not entirely impossible) that differences in the spectral distance between the pairs, or differences in their prosodic realization, lead to differences in baseline lateralization. Still, such acoustic differences might explain initial stronger or weaker responses in one hemisphere. The observed divergence in developmental patterns does not follow from such initial differences, though. Since a review of available NIRS studies on phoneme discrimination suggests that native phonemes, even when processed in a right-dominant network first, move towards a left-dominant network with exposure (cf. Minagawa-Kawai et al., 2011), both contrasts would be predicted to get more left-lateralized with age. The only possible explanation for the opposite trend in response for the infrequent contrasts is that it was so low in frequency that it was treated as non-native, and its initially slightly left-lateralized response was proceeding towards a more bilateral response. Further study is needed to assess whether this unpredicted tendency would reach significance with a large dataset or rather represented measurement noise.

In conclusion, the current study found rather weak evidence for exposure-dependent differences in vowel processing. Despite testing infants just around the critical age for perceptual narrowing with a strong frequency manipulation, they showed no difference in their ability to discriminate a frequent and an infrequent contrast, or in the degree to which their responses were left-lateralized. Frequency seemed to affect lateralization differently with age, but the directionality of this effect needs further exploration.

Our study illustrates the need to critically assess the predictions made in the perceptual attunement literature by taking into account central parameters like the nature of the contrast, and the amount of exposure. Regarding the former parameter, it has been discussed in several contexts that the saliency of contrasts influences the degree to which perceptual attunement is observable (e.g., Best, 1994; Cristià et al., 2011; Werker & Curtin, 2005), but the questions of what determines the saliency of a contrast and how such differences would impact on later linguistic processing still calls for further study. Regarding the latter parameter, while our study was designed to investigate the effect of *relative* differences in frequency, an equally or even more critical question might be the *absolute* amount of input necessary to form a speech sound category. Indeed, our results might reflect that frequency of exposure does not affect the processing of speech sounds once a 'critical' number of instances has been encountered, and that the exposure even to very infrequent speech sound contrasts reaches this number relatively early in an infant's life.

The effect of frequency might also differ depending on the level of processing that is tapped, either the discrimination of 'phonetic' categories or the more linguistic processing of 'phonological' categories. With regard to the latter, the results on lateralization might suggest that the number of instances the speech sounds had been heard had been too low to count as linguistically relevant.

A related question has been a central topic in the word learning literature, where alternative hypotheses about the critical amount of exposure to acquire a word also inform theoretical debate on the mechanisms involved in word learning (cf. Frank & Goodman, 2012). This illustrates the broader relevance of investigating how the amount of exposure affects developing language learners.

Acknowledgements

References

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

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. doi:10.1177/00238309030460020601

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: The MIT Press. Retrieved from http://www.haskins.yale.edu/SR/SR107/SR107_01.pdf

Bonatti, L. (2009). Psyscope X Build 55 [Computer Program].

Cristià, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics*, *39*(3), 388–402. doi:10.1016/j.wocn.2011.02.004

Cristia, A., Minagawa-Kawai, Y., Egorova, N., Gervain, J., Filippin, L., Cabrol, D., & Dupoux, E. (in press). Neural correlates of infant accent discrimination: An fNIRS study. *Developmental Science*.

Everdell, N. L., Coulthard, M. G., Crosier, J., & & Keir, M. J. (2005). A machine for haemodialysing very small infants. *Pediatric Nephrology*, *20*, 636–643. doi:10.1007/s00467-004-1785-5

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London: SAGE Publications.

Frank, M., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998-998. doi: 10.1126/science.1218633

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(37), 14222–14227. doi:10.1073/pnas.0806530105

Houston, D., Horn, D., Qi, R., Ting, J., & Gao, S. (2007). Assessing Speech Discrimination in Individual Infants. *Infancy*, *12*(2), 119–145. doi:10.1111/j.1532-7078.2007.tb00237.x

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, *21*(1-2), 3–28.

Kirchhoff, K., & Schimmel, S. (2005). Statistical properties of infant-directed versus adult-directed speech: Insights from speech recognition. *The Journal of the Acoustical Society of America*, *117*(4), 2238–2246. doi:10.1121/1.1869172

Kotilahti, K., Nissilä, I., Näsi, T., Lipiäinen, L., Noponen, T., Meriläinen, P., … Fellman, V. (2010). Hemodynamic responses to speech and music in newborn infants. *Human Brain Mapping*, *31*(4), 595–603. doi:10.1002/hbm.20890

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

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. doi: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. doi:10.1126/science.1736364

Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neuroscience and Biobehavioral Reviews*, *34*(3), 269–84. doi:10.1016/j.neubiorev.2009.07.008

Maddieson, I. (2013). Vowel Quality Inventories. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology (Available online at http://wals.info/chapter/2, Accessed on 2014-02-01). Retrieved from http://wals.info/feature/2A#2/19.3/152.8

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, *11*(1), 122–34. doi:10.1111/j.1467-7687.2007.00653.x

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–11. doi:10.1016/S0010-0277(01)00157-3

Meints, K., & Woodford, A. (2008). Lincoln Infant Lab Package 1.0: A new programme package for IPL, Preferential Listening, Habituation and Eyetracking. [WWW document: Computer software & manual].

Minagawa-Kawai, Y., Cristià, A., & Dupoux, E. (2011). Cerebral lateralization and early speech acquisition: a developmental scenario. *Developmental Cognitive Neuroscience*, *1*(3), 217–232. doi:10.1016/j.dcn.2011.03.005

Minagawa-Kawai, Y., Cristià, A., Vendelin, I., Cabrol, D., & Dupoux, E. (2011). Assessing signal-driven mechanisms in neonates: brain responses to temporally and spectrally different sounds. *Frontiers in psychology*, *2*, 135. doi: 10.3389/fpsyg.2011.00135

Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *The Journal of Neuroscience*, *27*(2), 315–21. doi: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. In *Proceedings of the International Conference of Phonetic Sciences XVI* (pp. 1877–1880). Saarbrucken. Retrieved from http://www.icphs2007.de/conference/Papers/1487/1487.pdf

Oostdijk, N. (2000). Het Corpus Gesproken Nederlands. *Nederlands Taalkunde*, *Jaargang 5*, 280–284.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, the R. D. C. (2012). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-104.

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), 965–76. doi:10.1111/j.1467-8624.2012.01740.x

Pons, F., Mugitani, R., Amano, S., & Werker, J. F. (2006b). Distributional learning in vowel length distinctions by 6-month-old English infants (abstract). Presented at the *International Conference on Infant Studies*, Kyoto, Japan

Pons, F., Sabourin, L., Cady, J. C., & Werker, J. F. (2006a). Distributional learning in vowel distinctions by 8-month-old English infants (abstract). Presented at the 28th *Annual conference of the cognitive science society,* Vancouver, BC, Canada.

Sato, Y., Mori, K., Furuya, I., Hayashi, R., Minagawa-Kawai, Y., & Koizumi, T. (2003). Developmental changes in cerebral lateralization during speech processing measured by near infrared spectroscopy. *The Japan Journal of Logopedics and Phoniatrics*, *44*(3), 165–171.

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Development of hemispheric specialization for lexical pitch–accent in Japanese infants. *Journal of Cognitive Neuroscience*, *22*(11), 2503–2513. doi:10.1162/jocn.2009.21377

Tsuji, S., & Cristia, A. (2013a). Fifty years of infant vowel discrimination research: what have we learned? *Journal of the Phonetic Society of Japan, 17*(3), 1-11.

Tsuji, S., & Cristia, A. (2013b). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, *56*(2), 179-191. doi:10.1002/dev.21179

Van Son, R. J. J. H., Binnenpoorte, D., van den Heuvel, H., & Pols, L. C. W. (2001). The IFA Corpus: a Phonemically Segmented Dutch "Open Source" Speech Database. Poster presented at *Eurospeech* (pp. 2051–2054). Retrieved from http://www.fon.hum.uva.nl/Service/IFAcorpus/SLcorpus/AdditionalDocuments/IFAcorpusEurospeech2001.html

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

Werker, J. F., & 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. doi:10.1016/S0163-6383(84)80022-3

Phoneme frequencies and consonant-vowel association patterns in Japanese infant- and adult-directed speech

Chapter 6

Abstract

Japanese infant-directed speech (IDS) and adult-directed speech (ADS) were compared on their phoneme frequencies and consonant-vowel association patterns. Consistent with findings in other languages, a higher ratio of phonemes that are generally produced early was found in IDS compared to ADS: more labial consonants and low-central vowels, but fewer fricatives. Consonant-vowel associations in IDS also contained a high amount of the early-produced labial-central, coronal-front, coronal-central, and dorsal-back patterns. On the other hand, language-specific patterns included a higher frequency of dorsals, affricates, geminates and moraic nasals in IDS. These phonemes are frequent in adult Japanese, but not in the early productions or the IDS of other studied languages. In combination with previous results, the current study suggests that both fine-tuning (an increased use of early-produced phonemes) and highlighting (an increased use of language-specifically relevant phonemes) might modify IDS on phoneme level.

1 Introduction

Understanding the nature of infants' input is indispensable for research on language acquisition. Infants show an impressive ability to extract information from various sources of their ambient language such as the distribution of phonetic units, sequential probabilities between phonemes, and word stress. While dictionary counts or adult-directed speech (ADS) had been assumed a sufficient approximation of infants' input, evidence on differences between ADS and infant-directed speech (IDS), a speech-style used by caregivers when addressing their infants, is accumulating. Speech modifications in IDS have been reported at the phonological, prosodic, syntactic, and lexical levels (for an overview, see Soderstrom, 2007), documenting both commonalities and differences between languages. At the phonetic level, differences in vowel and consonant quality between IDS and ADS have been investigated (for an overview, see Cristia, 2013). Interestingly, however, only a few studies have addressed the characteristics of phoneme frequencies in IDS.

This lack of studies contrasts with the extensive literature on the development of early phoneme productions (e.g., Boysson-Bardies & Vihman, 1991; Jacobson, 1941/1968) and the grouping of phonemes into the most basic of association patterns, the consonant-vowel (CV) sequence (e.g., MacNeilage & Davis, 2000; Vihman, 1992). Only recently, studies comparing IDS and ADS in Korean (Lee, Davis, & MacNeilage, 2008) and English (Lee & Davis, 2010) have reported that phoneme distribution patterns of Korean and English IDS show both commonalities and cross-linguistic differences. However, research from a wider range of languages is needed to determine the role of input on infants' acquisition of phonological categories.

In order to gain further insight into possible modifications in IDS on the level of phoneme frequencies, the current study compares the frequency of occurrence of phonemes (consonants and vowels) and CV combinations in Japanese IDS with that of Japanese ADS. The findings are subsequently qualitatively compared with those of other languages, in particular with English and Korean. English has been studied extensively in terms of phoneme input and production, and is typologically and historically close to many other well-studied European languages. Korean

is, like Japanese, typologically and historically quite distinct from English, and as such instrumental to broadening our database. It is typologically and historically close to Japanese, but the phonologies of the two languages differ substantially, with Korean having a larger vowel and consonant inventory in addition to more complex syllable structure.

The following literature review describes how the production of phonemes and CV combinations develops in languages other than Japanese, highlighting cross-linguistic similarities and differences. Frequencies of phonemes and CV combinations in languages other than Japanese are then described and compared with those of Japanese. Finally, the structure of Japanese is outlined.

1.1 Development of early phoneme production

Ambient language input, in interaction with early production constraints, is considered a crucial source for learning to produce the native language phoneme inventory. Early claims of a rigid universal order of phoneme acquisition (Jacobson, 1941/1968) were not supported by subsequent studies that demonstrated a substantial variability in early production both within (Vihman, 1993) and across (Ingram, 1999) languages. Nonetheless, there is little doubt that motor constraints lead to a tendency of some phonemes to emerge earlier than others in babbling and early word production.

Based on an overview of several early production studies, Bernhardt and Stemberger (1998) reported that stops, nasals and glides are the manners of articulation that are produced earliest across languages, while fricatives, affricates and liquids occur comparatively late. This finding is consistent with more recent overviews of both American English (Smit, 2007) and British English (Howard, 2007), and members of other language families such as Cantonese (So, 2007), Finnish (Kunnari & Savinainen-Makkonen, 2007), Greek (Mennen & Okalidou, 2007), Spanish (Goldstein, 2007), and Thai (Lorwatanapongsa & Maroonroge, 2007). The fricative /h/ has also been reported to occur early across the languages Dutch (Fikkert, 1994), English, Swedish, and French (Vihman, 1992).

Regarding place of articulation, labials and coronals tend to be produced early compared to dorsals across languages (cf. Bernhardt & Stemberger, 1998). Languages do differ in the acquisition order of labials

117

and coronals, but there is no overall tendency for one to be predominantly produced earlier than the other. A consistent finding in studies and overviews of American English (Boysson-Bardies &Vihman, 1991; Smit, 2007), British English (Howard, 2007), Dutch (Fikkert & Levelt, 2008), French (Boysson-Bardies &Vihman, 1991; Rose and Wauquier-Gravelines; 2007), Spanish (Goldstein, 2007), German (Fox, 2007), Jordanian Arabic (Dyson & Amayreh, 2007), Cantonese (So, 2007), and Greek (Mennen & Okalidou, 2007) is an earlier onset of labial and coronal place of articulation compared to dorsals.

For vowels, front/central mid/low vowels (i.e., vowels located in the *lower left quadrant* of the F1/F2 vowel space) have been reported to be most frequent in early productions (Davis & MacNeilage, 1990). Similarly, for American English (Smit, 2007) it was reported that back vowels and the front-high vowel /i/ are rare in early productions, and that front-high /i/ and front-mid /ɛ/ remain erroneous for children between 1-3 years of age.

These common tendencies have mainly been explained with reference to articulatory restrictions. Stops and nasals, which are produced by a complete closure of the vocal tract, are considered relatively easy to produce compared to fricatives and affricates, which require a more complex coordination of articulatory position and airflow (cf. Kent, 1992; MacNeilage, Davis, Kinney, & Matyear, 2000). Vihman and colleagues suggested that labial and coronal stops are easy to articulate as they require simple mandibular oscillations (e.g., Vihman, 1993) and because the accompanying lip closure is considered an especially salient visual cue (Boysson-Bardies & Vihman, 1991). Despite these common tendencies, cross-linguistic variation exists. This variation is generally attributed to the nature of the input, which will be reviewed in the following section.

1.2 Phoneme characteristics of the input

One of the first studies on phoneme frequencies in IDS was a qualitative description of baby talk, words modified for infants, across 15 languages by Ferguson (1977). A later study (Vihman, Kay, Boysson-Bardies, Durand, & Sundberg, 1994) quantitatively compared the

characteristics of IDS in running speech, content words, word-initial phonemes of content words and adult target models of children's attempted words across American English, French and Swedish. In order to more specifically address the differences between IDS and ADS, two recent studies directly and quantitatively compared phoneme frequencies of Korean (Lee et al., 2008) and English (Lee & Davis, 2010) IDS and ADS. In both samples, the speech of ten mothers to their one-year-old infants was compared to a sample of ten women speaking to an adult experimenter.

The most consistent pattern with regard to place of articulation was a high frequency of labial place, both in terms of a high frequency compared to other places of articulation within IDS, and in terms of a high frequency in IDS compared to ADS. Ferguson (1977) reported a high frequency of labial and coronal stops. Vihman et al. (1994) also found support for a higher frequency of coronals and labials compared to dorsals in running speech across languages, with coronals being most frequent. Finally, comparisons between IDS and ADS showed that Korean IDS contained a significantly higher frequency of labial place, and a significantly lower frequency of coronal and glottal place compared to ADS (Lee et al., 2008). No differences with regard to place were found in the English sample (Lee and Davis, 2010).

Regarding manner, stops were reported to be frequent in two studies. Ferguson's (1977) sample contained a high frequency of nasals and a low frequency of liquids. In the sample of Vihman et al. (1994), stops were most frequent, followed by fricatives/affricates, nasals and glides. While Lee and Davis (2010) found a significantly higher frequency of stops and glides, and a lower frequency of fricatives, affricates, nasals and liquids for English IDS compared to ADS, no differences were found in the Korean sample (Lee et al., 2008).

Lee et al.'s study on Korean also reported the Korean fortis and geminate consonants to be more frequent in Korean IDS than ADS. In the same study, mid-central and low-central vowels were significantly more frequent and high-central and mid-front vowels were significantly less frequent in IDS compared to ADS.

We can find some consistencies between these IDS phoneme frequencies and the early productive tendencies reviewed in the previous

section: Fricatives both emerged later in early productions and were less frequent in IDS than ADS across the languages studied. The early-produced labial consonants and lower left quadrant vowels were relatively frequent in Korean IDS, and stops and glides were relatively frequent in English IDS. The relatively late-produced affricate and liquid consonants were less frequent in English IDS compared to ADS.

In summary, if anything, phoneme frequencies in IDS show a better fit with early production patterns than with ADS patterns. This tendency has been suggested to reflect that early produced phonemes are favored, and late produced phonemes are avoided or substituted in IDS (cf. Ferguson, 1977, Lee et al., 2008, Lee and Davis, 2010). On the other hand, the usage of late produced phonemes in IDS has also been suggested to reflect an increased use of language-specific, perceptually salient phonemes (Lee et al., 2008). Considering the rather small number of studies available to date, more data from other languages is necessary to evaluate these interpretations.

1.3 Phonemes in Japanese

Studies on Japanese children's development of phoneme productions diverge in part from the studies on other languages reviewed above. In a study comparing English-, French-, Japanese- and Swedish-learning children's babbling and early speech (Boysson-Bardies & Vihman 1991), Japanese children produced a relatively low number of labials and a high number of dorsals. Unlike English, French and Swedish children, whose use of fricatives and affricates decreased in first words compared to babbling, Japanese children showed no such decrease.

Edwards and Beckman (2008) reported that substitution patterns of Japanese- and English-acquiring children's early pronunciation errors reflected differences in phoneme distributions of the input: While English-learning children tended to substitute coronal [t] for dorsal /k/, Japanese children rather substituted [k] for /t/. A longitudinal study following phoneme mastery of ten Japanese children from 1;0-4;0 years (Uno, 2007) revealed that they first mastered the labial stop /b/ and nasal /m/ (by 1;3 years), immediately followed by the stops and nasals /p, t, d, k, n, g/. The alveopalatal affricate /tɕ/ was acquired by 1;6 years, relatively early in comparison to other languages (e.g., Bernhardt & Stemberger, 1998; Kent,

1992). Thus, Japanese children's early productions included a high amount of labial, stop and nasal consonants, consistent with universal tendencies. However, they also included a comparatively high amount of affricates and dorsal consonants.

Ferguson (1977) reported that baby talk words in Japanese often included geminates and affricates. More recently, it was reported that Japanese IDS contained a higher frequency of dorsal stops (/k, g/) than coronal stops (/t, d/) (Beckman, Yoneyama, & Edwards, 2003), in contrast to studies in other languages that have reported the opposite pattern. This recent study, though, was focused specifically on place of articulation in stop consonants. In order to study language-specific and language-general patterns beyond this sub-group, an overall quantitative analysis of phoneme distributions of Japanese IDS and ADS is mandatory.

1.4 Early CV association patterns

Phonemes, especially consonants, rarely occur in isolation. The grouping of phonemes into CV sequences is an important milestone towards the acquisition of speech, first occurring between the ages 0;6-0;8 in the stage of canonical babbling and necessarily preceding speech (Vihman, 1992). Similar to research on early phoneme production, research on early CV association patterns has considered constraints and regularities. MacNeilage et al. (2000) proposed that basic biomechanical constraints lead to three preferred association patterns in early production. In their Frame/Content theory, the association of labial consonants with central vowels reflects a *pure frame* resulting from simple mandibular oscillations. By adding a tongue movement to this basic oscillation, two additional associations, coronal-front and dorsal-back, are formed. These three association patterns were observed to be more frequent than expected by chance in the babbling and early speech of 15 English-learning infants, as well as overall in dictionary counts of the nine languages French, Swahili, Estonian, Hebrew, German, Spanish, English, Maori and Quechua.

Another investigation of early CV association patterns (Vihman, 1992) with samples of American, French and Swedish children found support for the labial-central association but not for the other two

association patterns. Instead of a coronal-front association, the sample showed a positive association between coronal consonants and central vowels. Regarding associations of dorsal consonants with vowels, it was difficult to find any pattern due to the low frequency of dorsal phonemes.

The relationship between IDS and early production of CV sequences has been investigated in two recent studies. A study on the relationship between IDS and 0;7-1;6 aged infants' output in Mandarin Chinese (Chen & Kent, 2005) found strong correlations between a subset of infants' predominant production patterns and caregivers' speech. Infant output provided support for the coronal-front and dorsal-back frame, as well as a labial-back association pattern. IDS correlated with the labial-back and dorsal-back, but not the coronal-front association pattern.

A study of Korean compared CV association patterns of infants' babbling and first words with IDS and ADS (Lee, Davis, & MacNeilage, 2007). They found support for the association patterns proposed by MacNeilage et al. (2000) in babbling, which were suggested to reflect early and possibly intrinsic constraints. In first words on the other hand, the coronal-front and dorsal-back, but not the labial-central associations were found. Both babbling and early words showed an imperfect but large overlap with IDS in the predominant association patterns, while there was no overlap with ADS. In IDS, a predominant coronal-front association pattern was observed, while ADS did not show any of the previously suggested basic patterns. These two studies show that there is a stronger relationship of early CV association patterns with IDS than with ADS, suggesting that the former resembles early production patterns more closely.

In sum, support for the labial-central association pattern comes from Swedish, English and French early productions as well as from Korean babbling, for the coronal-front pattern from English, Chinese, and Korean early productions and for the dorsal-back pattern from Chinese and Korean early productions. Further frequent patterns are coronal-central and labial-back, and early produced CV association patterns overlap with patterns in IDS, but not ADS.

CV association patterns in Japanese seem to diverge from the above-mentioned languages. The aforementioned dictionary study (MacNeilage et al., 2000) showed Japanese to be the only language in which the

average observed-to-expected frequency ratios for the three suggested patterns did not exceed chance level. Labial-central and dorsal-back associations showed a tendency in the expected direction, but the coronal-central association led to a higher observed-to-expected ratio than the coronal-front one. Similarly, Vihman (1992) found labial-central, dorsal-back and coronal-central associations for Japanese children. Notably, in her sample of four languages (Swedish, English, French and Japanese), only Japanese children produced a substantial frequency of dorsal consonants and back vowels, and consequently they alone contributed a high quantity of dorsal-back associations. Japanese children's early productions resemble the findings in other languages in that they contain labial-central associations. They do contain a high amount of coronal-central associations as proposed bu Vihman, but nor of coronal-front associations as proposed by MacNeilage et al. Finally, their productions include a high frequency of dorsal-back associations, consistent with the proposal made by MacNeilage et al. Japanese, along with Mandarin (Chen & Kent, 2005) and Korean (Lee, Davis, & MacNeilage, 2007), is a language in which children seem to produce many dorsal and back phonemes.

1.5 IDS patterns

The above review suggests that IDS is distinct from ADS at the phoneme level. Overall, IDS phoneme distributions parallel infants' early productions better than ADS. This is generally interpreted as a fine-tuning of caregivers' articulations to infants' capacities, favoring phonemes that are generally produced early while avoiding phonemes that are generally produced late. We will call this the *fine-tuning* pattern, following Cross (1977). This pattern predicts a universal tendency for a higher frequency of phonemes that are generally produced early and a lower frequency of phonemes that are generally produced late. However, other patterns of modification are also conceivable. Caregivers could produce language-specific phonemes more frequently, thus highlighting patterns that are important for the native language but are not necessarily acquired early in general. We will call this the *highlighting* pattern.

To distinguish fine-tuning and highlighting, it is necessary to examine a language in which aspects of phoneme acquisition divergence from common patterns, as otherwise the most frequent phonemes will

match what is easy for infants to produce. The above literature review shows that Japanese is such a language, with part of the early productions strongly reflecting language-specific characteristics. Before summarizing the aims of the current study, some relevant characteristics of Japanese phonology are described.

1.6 Japanese linguistic structure

Japanese is a mora-timed language, where one mora is a sub-syllabic unit that can consist of a single vowel (V), a CV sequence, the moraic nasal /N/ or the first half of a geminate consonant /Q/. Japanese light syllables consist of either V or CV, and heavy syllables are formed by vowel lengthening or by adding /N/ or /Q/ to a CV sequence. Consequently, Japanese syllables are mostly V or CV and the occurrence of consonant clusters is rare.

The Japanese vowel inventory consists of the five mono-moraic short vowels /a, i, u, e, o/, and their long bi-moraic counterparts /a:, i:, u:, e:, o:/. There are no quality differences between short and long vowels (Saito, 1997). As we are going to qualitatively compare our findings to previous findings in Korean and English later on, we are referring to characteristics of these languages where adequate. In terms of monophthong vowels, the Japanese inventory of five is smaller than the inventories of both Korean and English. Korean consists of eight, and English of twelve monophthong and an additional three diphthong vowels (Lee et al., 2008; Lee & Davis, 2010). Taking into account that Japanese distinguishes long and short vowels, the total number of vowel categories is ten, putting it in between Korean and English.

Japanese has 23 consonants, and all consonants except the moraic /N/ necessarily precede a vowel in a CV sequence. Additionally, Japanese has a geminate phoneme /Q/, which forms a geminate or long consonant combined with a singleton plosive or fricative consonant combined, and is in phonemic contrast to singleton consonants. As the consonantal status of the geminate phoneme is controversial (cf. Vance, 1987), we did not include it in our analysis of place and manner, but its frequency of occurrence in IDS and ADS separately. Based on the higher frequencies of geminate consonants in both Japanese baby talk words (Ferguson, 1977) and Korean IDS (Lee et al., 2008), a difference might be expected. The

number of phonemic consonants was reported to be 24 in English (Lee and Davis, 2010), and 19 in Korean (Lee et al., 2008). Like English, Japanese distinguishes voiced and unvoiced stops, while Korean makes a three-way distinction between lenis, fortis and aspirated. English contains nine fricatives while Japanese has eight (two of which are extremely rare) and Korean only three.

Japanese IDS further incorporates some salient characteristics, which are important to consider in addition to the Japanese phoneme inventory in order to make predictions on IDS-specific phoneme frequencies. It contains a high amount of specialized vocabulary that is often phonologically unrelated to the adult form. This is known as infant-directed vocabulary (IDV). A survey of mothers of 0;8-1;0 year-old infants reported 237 distinct infant-directed word-types (Mazuka, Kondo & Hayashi, 2008). For example, instead of *kuruma* (car) in ADS, *buHbu* is used in IDS, and instead of *gohan* (meal), *maNma* is used. Many of the expressions in IDV have their roots in onomatopoeia, and occur most frequently in heavy-light or heavy-heavy disyllabic forms (79% of word forms reported in the survey). Since heavy syllables necessarily contain either a geminate, a moraic nasal or a long vowel, it is of interest whether the frequency of occurrence of these three phoneme types differs between IDS and ADS. The moraic nasal is pooled with non-moraic nasals in consonant analysis, but in order to capture its exceptional status a separate analysis compares the frequency of occurrence of moraic and non-moraic nasals in IDS and ADS.

IDV is also related to the high frequency of *youon* consonants. Since the definition of youon depends on orthographic characteristics, it cannot directly be related to phonological categories. Orthographically, youon are formed by adding a small kana symbol that represents glides to a normal-sized one, for example in ギャ [gʲa] or チャ [tɕa]. One group of youon consists of consonants that are palatalized before a vowel, for example in [gʲa] or [kʷo]. The other group of youon includes fricative or affricate consonants that precede the vowels /a, u, e, o/, for example in [tɕa] or [ɕo]. Note that consonants preceding /i/ are palatalized in most cases but are not classed as youon as they do not have the orthographic distinction described above. Youon are often associated with a familiar/casual style of speech and with speech directed to young children and infants.

Examples include [tɕittɕai] instead of [tɕi: sai] 'small' and diminutive suffixes for people's names or kinship terms such as [onii-tɕaN] instead of [onii-saN] 'older brother'. They are also used frequently in onomatopoeic expressions (e.g. [tɕokitɕoki] to describe the action of cutting something with scissors), which are in turn used often in IDV, as discussed above. Therefore, a difference in the frequency of youon between IDS and ADS might be expected.

Word boundaries in Japanese can be determined by either referring to *short-unit* or *long-unit* words. Short-unit words roughly correspond to dictionary entries and are monomorphemic or at most bimorphemic. Long-unit words are combinations of words that may correspond to compound words. For instance, *baikiNmaN* (a Japanese cartoon character, 'germ-man') may be analyzed as two short-word units, *baikiN* 'germ' and *maN* 'man', or as one long unit. To our knowledge, there exists neither a strict agreement concerning when to use short and long units nor any reference discussing this topic. As long units are often perceived as the more natural boundaries, we chose those for analysis.

1.7 Aims of the current study

The above review shows that, if anything, IDS patterns fit early productions better than ADS patterns. These differences could be due to caregivers' fine-tuning, accounting for an increased use of phonemes that are generally produced early, and/or due to highlighting of language-specific patterns. The available data are too sparse and varied to establish the above tendencies, and data from additional languages are necessary to evaluate systematic differences between IDS and ADS at the phoneme level. Analyzing phoneme frequencies in a relatively large corpus of Japanese, a language in which early production patterns show both language-general and language-specific patterns, will provide a further step towards answering this question.

The current study will thus evaluate differences and similarities between Japanese IDS and ADS in light of the fine-tuning and highlighting accounts. If caregivers are fine-tuning their speech to infants' production capacities, we expect IDS to contain more frequently groups of phonemes that are generally produced early. This would be labial and coronal place of articulation; stop, nasal and glide manner; lower left quadrant vowels;

and labial-central, coronal-central, coronal-front and dorsal-back consonant-vowel associations. If on the other hand, caregivers are highlighting language-specific patterns, we expect that IDS contains more frequently groups of phonemes that are both frequent in Japanese and acquired rather late in general. Among these patterns are dorsals, affricates and dorsal-back consonant-vowel associations. Additionally, phoneme types occurring in Japanese infant-directed vocabulary (geminates, moraic nasals, youon) are expected to occur frequently in IDS.

We will compare phoneme frequencies of IDS and ADS for place of articulation, manner of articulation and vowels in that order, as well as evaluate the occurrence of consonant-vowel sequences in IDS and ADS.

The methods of phoneme comparison will closely follow those in Lee et al. (2008) and Lee and Davis (2010) in order to facilitate comparison (a more detailed explanation is provided in the Methods section). However, running speech, which was used in these studies, might not be the most representative measure of what matters to infants. There is evidence that children orient to initial consonants in word selection (Boysson-Bardies & Vihman, 1991) and that content-words are especially salient to infants (Shi & Werker, 2001). Moreover, phoneme frequency of content-word-initial phonemes has been found to better reflect children's early productions than running speech (Vihman et al., 1994). Therefore, both running speech and word-initial content-words were examined.

2 Methods

2.1 Corpus

The corpus used in this study contains the speech of 22 Japanese mothers from the Tokyo area and their 1;6-2;0 old children (Mazuka, Igarashi, & Nishikawa, 2006). Children of this age are at the early stage of their production and comprehend some of what their mothers say to them. Recordings of each mother-child dyad took place in a sound-attenuated room. The mother's utterances were recorded by a head-mounted dynamic microphone, and a condensor microphone placed on a table recorded the child's utterances. Additionally, dyads were video-recorded by means of a ceiling camera and microphone. Audio recordings

were made with DAT tapes, and video recordings with mini DV tapes. For IDS samples, two separate recordings were made. During the first 15 minutes, the mother was asked to play with the child using a number of picture books. Mothers could choose from seven books depicting a variety of animals, toys and actions and contained very little text. For the remaining 15 minutes the books were replaced by a set of silent toys such as animals, soft blocks and finger puppets. Mothers were free to use any of the materials but were not specifically instructed to do so. Some mothers in fact played with their child without using any of the materials that were provided. For ADS samples, a female experimenter subsequently entered the room and talked with the mother for ten minutes, in the child's presence, about topics related to child-raising. A total of approximately 45 minutes of recording per dyad was obtained.

2.2 Data coding

The IDS recordings totaled about 11 hours of speech and 50,000 words, the ADS recordings 3 hours and 30,000 words. Annotations were based on the schemes developed for the Corpus of Spontaneous Japanese (Maekawa, 2003). The phonetic transcriptions were performed by three highly trained phoneticians. In cases of disagreement or uncertainty, they examined the original sound files together in an effort to resolve the issue. When no agreement could be reached regarding some section, it was marked and excluded from the analysis. The entire corpus was double-checked for its accuracy by a single phonetician.

Table 1. Japanese consonant inventory

	labial	**coronal**	**dorsal**	**unclassified**
Stop	p b	t d	k g	
Fricative	ɸ v	s z ɕ ʑ ç	h	
Affricate		ts tɕ dʑ		
Nasal	m	n ɲ	ŋ ɴ	
Glide		j		w
Liquid		ɾ		

Table 2. Japanese vowel inventory

	front	central	back
High	i		u
Mid	e		o
Low		a	

Phonemes were transcribed according to the Japanese consonant and vowel inventory (cf. Tables 1 and 2). Additionally, the geminate phoneme /Q/ was coded. Transcribed consonants were classified for place of articulation as labials [p, b, ɸ, v, m], coronals [t, d, s, z, ɕ, z, ç, ts, tɕ, dʑ, n, ɲ, j, ɾ], and dorsals [k, g, h, ŋ]. The place of articulation of the moraic nasal /N/ depends on the place of articulation of the following consonant such that it is realized as [m] preceding labial consonants, [n] preceding coronal consonants and [ŋ] preceding dorsal consonants. We classified it post-hoc according to these rules. However, if /N/ was followed by a vowel or a pause, it could not be classified and was thus excluded from the analysis of place of articulation. About 34% of moraic nasals were excluded for that reason. Additionally, the glide [w] and the labiovelarized stops [kʷ] and [gʷ] were not classifiable for place of articulation post-hoc and therefore excluded from this analysis. For manner analysis, [p, b, t, d, k, g] were classified as plosives, [ɸ, v, s, z, ɕ, z, ç, h] as fricatives, [ts, ɕ, dʑ] as affricates, [m, n, ɲ, ŋ, ɴ] as nasals, [j, w] as glides, and [ɾ] as liquid. The phonemes that could not be classified for place of articulation were included in the manner analysis. Note that the corpus also phonetically coded phonotactic patterns that only occur in loanwords such as [tii] (tea). Adjectives, adjectival nouns, adnominals, adverbs, nouns, and verbs were considered content words.

2.3 Data analysis

The two types of IDS samples (book reading and toy playing) were collapsed after an initial comparison of the two data types did not show systematic differences. As total sample sizes of IDS and ADS differed, phoneme frequency ratios rather than absolute frequencies were used for

analysis. Ratios were calculated separately for IDS and ADS, for vowels and consonants, and for the analysis considering running speech and the analysis considering only content-word-initial phonemes. For example, the ratio of the labial stop [b] in infant-directed running speech was calculated by dividing the total number of [b] occurrences by the number of all consonants in infant-directed running speech. The obtained ratios were then subjected to an arcsine transformation, a common transformation recommended for stabilizing variances in proportional variables (Cohen, Cohen, West, & Aiken, 2003).

Repeated-measures analyses of variance (ANOVA) were conducted on place, manner and vowel contrasts, followed by Bonferroni-adjusted pairwise comparisons where appropriate. Greenhouse-Geisser corrected values were reported where the sphericity assumption was violated. Separate paired t-tests compared the frequencies of youon and of geminates, and a separate ANOVA compared the frequency of moraic and non-moraic nasals in IDS and ADS. Recently, the use of parametric statistical tests like ANOVA for analyzing phoneme frequencies in corpus data has been criticized due to their distributional properties and a non-parametric alternative was proposed (Daland, 2012). Results from an analysis following this method were comparable to those reported below, and are omitted due to space constraints.

For analysis of CV sequences, we calculated observed-to-expected ratios for each of the nine possible consonant-vowel association patterns for labial, coronal and dorsal consonants with front, central and back vowels, adopting the procedure introduced in Lee et al. (2007). Expected frequencies were obtained by multiplying the number of consonants in the respective place of articulation with the number of vowels in the respective position, and dividing this number by the total number of CV association patterns. For instance, the expected frequency of labial-front associations was obtained by multiplying the number of labial consonants with the number of front vowels, and dividing the result by the total number of CV associations. The observed-to-expected ratio was then calculated by dividing the observed frequency of each CV association pattern by its expected frequency. Chi-square tests were conducted to indicate if observed frequencies overall differed significantly from expected frequencies. If so, to determine which of the CV association

patterns contributed to this result, the standardized residuals for every association pattern were obtained, where a category with a standardized residual value above 2 is considered to be a major contributor to significance. As analyses of early CV association patterns in production mainly concentrate on the early-acquired groups of stops and nasals (cf. MacNeilage & Davis, 2000), we report results on this subgroup of phonemes in addition to results including all phonemes.

3 Results and Discussion

Overall, there were a total of 75,199 consonants in IDS and of 34,973 consonants in ADS. Vowel numbers totaled 78,583 in IDS and 37,154 in ADS. Assuming that the number of vowels roughly corresponds to the number of syllables in a corpus, our data is approximately 11 times the size of previous Korean and English studies.

3.1 Consonant place
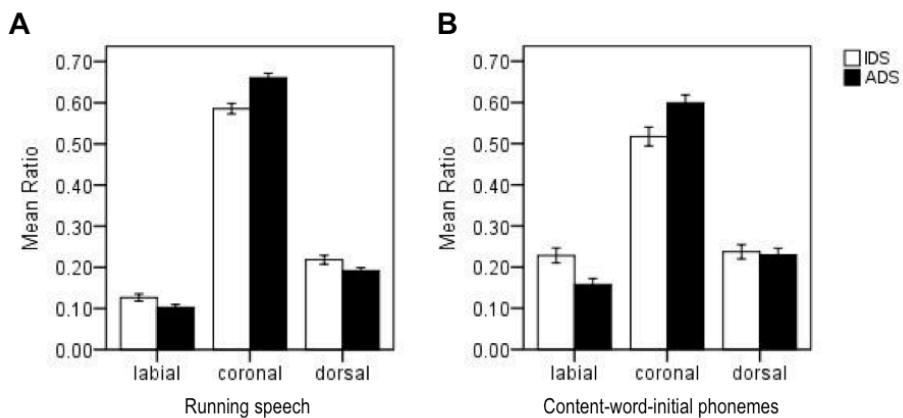
Figure 1. Percentage of each consonant place of articulation. A: Running speech, B: Content-word-initial phonemes. Error bars represent ± 2 standard errors.

3.1.1 Running speech

Coronal place of articulation was most frequent in both IDS (59%) and ADS (66%), followed by dorsal (22% for IDS and 19% for ADS) and labial (13% for IDS and 10% for ADS) places of articulation. The percentages do not reach 100 % because of the moraic nasals that were

Table 3. Pairwise comparisons of consonant proportions in IDS and ADS. IDS and ADS ratio values are untransformed ratios; difference values, standard errors and p values are based arcsine-transformed ratios.

	Ratio IDS	Ratio ADS	Mean Difference $(IDS_{arc}-ADS_{arc})$	Standard Error$_{arc}$	p value
Running speech					
Consonant place					
Labial	.127	.103	.074	.015	<.001
Coronal	.586	.661	-.156	.019	<.001
Dorsal	.218	.192	.066	.016	.001
Consonant manner					
Stop	.386	.375	.024	.014	.104
Fricative	.142	.162	-.056	.020	.011
Affricate	.051	.032	.098	.011	<.001
Nasal	.282	.300	-.040	.013	.006
Glide	.056	.059	-.014	.013	.297
Liquid	.082	.072	.034	.015	.031
Content-word-initial phonemes					
Consonant place					
Labial	.228	.158	.181	.024	<.001
Coronal	.517	.599	-.166	.035	<.001
Dorsal	.237	.230	.016	.027	.558
Consonant manner					
Stop	.459	.297	.340	.038	<.001
Fricative	.216	.321	-.242	.036	<.001
Affricate	.046	.048	-.009	.031	.787
Nasal	.192	.209	-.047	.027	.097
Glide	.069	.115	-.168	.028	<.001
Liquid	.019	.010	.077	.033	.030

not classifiable for place of articulation, the glide [w] and the labiovelarized stops [kʷ] and [gʷ]. A two-way repeated measures ANOVA with the factors speech style (IDS, ADS) and place of articulation (labial, coronal, dorsal) revealed significant main effects for both speech style [$F(1,21) = 5.66$, $p = 0.027$, $\eta^2_p = .212$] and place of articulation [$F(2,42) = 3680.65$, $p < 0.001$, $\eta^2_p = .994$], as well as a significant interaction between the two [$F(2,42) = 39.80$, $p < 0.001$, $\eta^2_p = .655$].

Post-hoc pairwise comparisons between IDS and ADS for labial, coronal and dorsal place of articulation showed significant differences for all three places. As shown in Table 3, labial and dorsal place were more frequent in IDS, but coronal place more frequent in ADS (Figure 1A).

3.1.2 Content-word-initial phonemes

For word-initial phonemes of content words, coronal place of articulation was again most frequent in both IDS (52%) and ADS (60%), followed by dorsals in IDS (24%) and ADS (23%) and labials in IDS (23%) and ADS (16%). A speech style x place of articulation repeated measures ANOVA showed significant main effects for both speech style [$F(1,21) = 8.04$, $p = 0.01$, $\eta^2_p = .277$] and place of articulation [$F(2,42) = 573.97$, $p < 0.001$, $\eta^2_p = .965$], and a significant interaction effect between the two factors [$F(1.31,27.44) = 24.16$, $p < 0.001$, $\eta^2_p = .535$]. Post-hoc paired comparisons between speech styles for each place of articulation showed significant differences for labial and coronal place of articulation (Table 3), with labials being more frequent in IDS, and coronals in ADS (Figure 1B). The findings for content-word-initial phonemes are generally parallel to those of running speech, except that the difference between IDS and ADS for dorsal phonemes is not significant here.

3.2 Consonant manner

3.2.1 Running speech

For both IDS and ADS, stops were the most frequent manner category with 39% in IDS and 38% in ADS. The second most frequent category was nasals with 28% in IDS and 30% in ADS, followed by fricatives (14% for IDS and 16% for ADS), liquids (8% for IDS and 7% for ADS), glides (6% for both IDS and ADS) and affricates (5% for IDS and 3%

for ADS). A two-way repeated measures ANOVA: speech style (IDS, ADS) x manner of articulation (stop, nasal, fricative, affricate, glide, liquid) revealed significant main effects for speech style [F(1,21) = 11.65, p = 0.003, η^2_p = .357] and manner [F(3.00,63.09) = 1576.50, p < 0.001, η^2_p = .987], and a significant interaction between the two [F(3.24,67.94) = 12.79, p < 0.001, η^2_p = .378]. Post-hoc paired comparisons showed significant differences in fricative, affricate, nasal and liquid manner between IDS and ADS (Table 3), affricates and liquids being more frequent in IDS, and fricatives and nasals in ADS (Figure 2A).

A separate nasal type (moraic, non-moraic) x speech style (IDS, ADS) repeated-measures ANOVA was conducted to separate the moraic and non-moraic nasal. This analysis was only conducted for running speech, as the moraic nasal rarely occurs word-initially. A significant interaction between nasal type and speech style was found [F(1,21) = 157.94, p < 0.001, η^2_p = .883], with a higher frequency of the non-moraic nasal for ADS (M = 0.14) than IDS (M = 0.10), and a higher frequency of the moraic nasal for IDS (M = 0.13) than ADS (M = 0.10).

3.2.2 Content-word-initial phonemes

Word-initially, stops were again the most frequent category for IDS (46%), and fricatives for ADS (32%). These were followed by fricatives (22%), nasals (19%), glides (7%), affricates (5%) and liquids (2%) in IDS, and by stops (30%), nasals (21%), glides (12%), affricates (5%) and liquids (1%) in ADS. A two-way repeated-measures ANOVA with the factors speech style and manner of articulation (6) revealed no main effects for speech style [F(1,21) = 2.845, p = .106, η^2_p = .119], a significant effect of manner [F(5,105) = 533.62, p < 0.001, η^2_p = .962], and a significant interaction between the two [F(5,105) = 33.97, p < 0.001, η^2_p = .618]. Post-hoc paired comparisons showed that stops and liquids were significantly more frequent in IDS, while fricatives, and glides were more frequent in ADS (Table 3; Figure 2B).

3.3 Youon

The ratio of occurrence for all youon in IDS and ADS was compared. Youon were significantly more frequent in IDS than in ADS both for

running speech (IDS: M = 0.075; ADS: M = 0.031; t(21) = 13.32, p < .001, *d* = 2.561), and for content-word-initial phonemes (IDS: M = 0.083; ADS: M = 0.047; t(21) = 4.69, p < .001, *d* = 1.013).
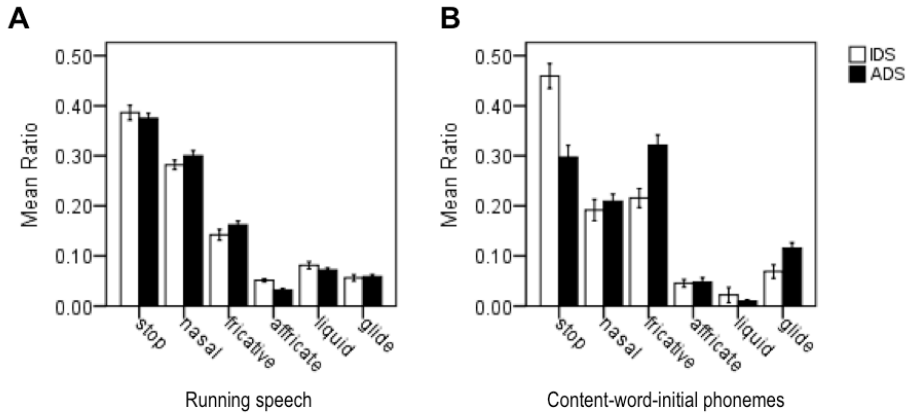
Figure 2. Percentage of each consonant manner of articulation. A: Running speech, B: Content-word-initial phonemes. Error bars represent ± 2 standard errors.

3.4 Geminate stops and fricatives

Ratios for geminates were calculated by dividing the number of geminates by the number of consonants plus geminates. As geminates rarely occur word-initially, only running speech was considered. A paired t-test revealed significant differences with a higher geminate ratio in IDS (M = 0.062) than in ADS (M = 0.050) [t(21) = 2.94, p = .008, *d* = 0.607].

3.5 Vowels

3.5.1 Running speech

As can be seen in Figure 3, the majority of vowels are short. A speech style (IDS, ADS) x vowel length (short, long) x vowel quality (high-front, mid-front, low-central, high-back, mid-back) repeated measures ANOVA was conducted. The results revealed a significant main effect of vowel quality [$F(4,84) = 295.45$, $p < 0.001$, $\eta^2_p = .934$] and vowel length [$F(1,21) = 85299.35$, $p < 0.001$, $\eta^2_p = 1.00$], a significant interaction

Table 4. Pairwise comparisons of vowel proportions in IDS and ADS. IDS and ADS ratio values are untransformed ratios; difference values, standard errors and p values are based on arsine-transformed ratios.

	Ratio IDS	Ratio ADS	Mean Difference (IDS_{arc}-ADS_{arc})	Standard $Error_{arc}$	p value
Running speech					
Vowels					
a	.371	.310	.134	.014	<.001
i	.181	.172	.022	.016	.182
u	.122	.140	-.056	.016	.002
e	.128	.190	-.170	.014	<.001
o	.274	.269	.011	.020	.595
a:	.018	.010	.069	.016	<.001
i:	.011	.007	.050	.011	<.001
u:	.014	.014	.000	.016	.954
e:	.005	.009	-.033	.019	.102
o:	.031	.043	-.064	.017	.001
Content-word-initial phonemes					
Vowels					
a	.062	.048	.056	.027	.049
i	.103	.064	.142	.022	<.001
u	.015	.019	-.035	.024	.165
e	.007	.006	.041	.023	.084
o	.066	.071	-.021	.023	.384
a:	.004	.003	.002	.017	.905
i:	.019	.007	.107	.018	<.001
u:	.020	.009	.020	.014	.166
e:	.003	.006	.006	.016	.687
o:	.003	.085	.032	.022	.166
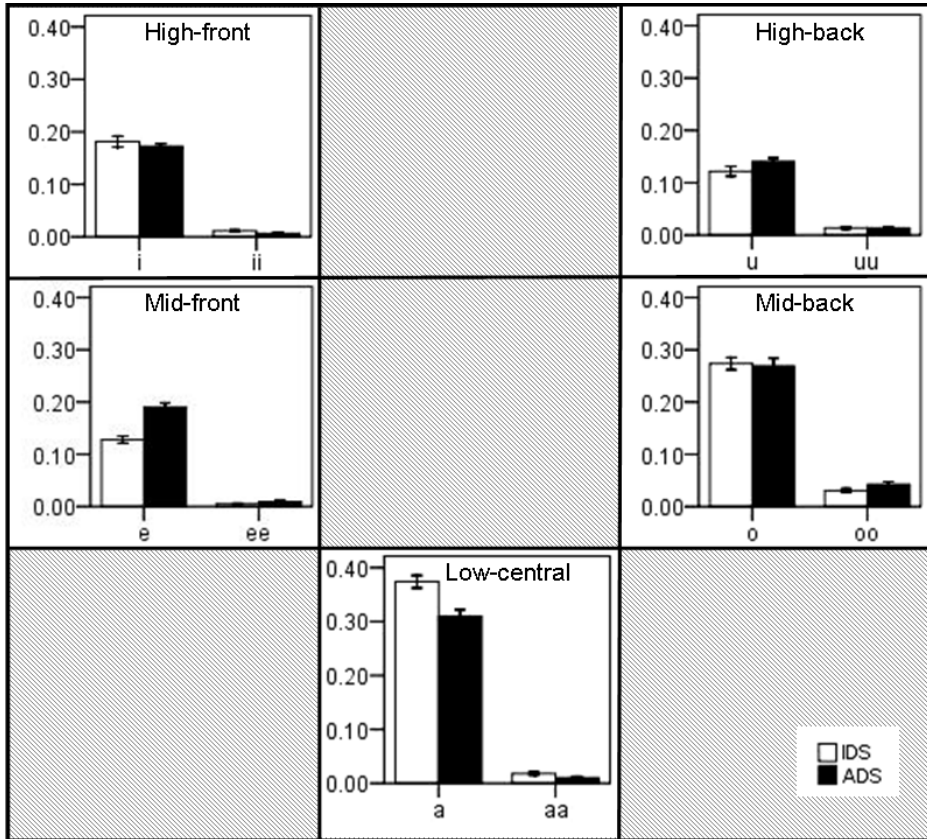
Figure 3. Percentage of each vowel place of articulation, running speech. Error bars represent ± 2 standard errors.

between speech style and vowel quality [$F_{(4,84)}$ = 27.54, $p < 0.001$, η^2_p = .567], between speech style and vowel length [$F_{(1,21)}$ = 5.53, $p = 0.029$, η^2_p = .208], between vowel length and vowel quality [$F_{(2.62,55.09)}$ = 433.28, $p < 0.001$, η^2_p = .954] and between speech style, vowel quality and vowel length [$F_{(4,84)}$ = 39.02, $p < 0.001$, η^2_p = .650].

Post-hoc paired comparisons showed significant vowel quality differences such that long high-front /i:/, short and long low-central vowels /a, a:/ were more frequent in IDS, and short high-back /u/, short mid-front /e/ and long mid-back vowels /o:/ were more frequent in ADS (Table 4; Figure 3).

3.5.2 Content-word-initial phonemes

A speech style x vowel length x vowel quality repeated measures ANOVA showed a significant main effect of speech style [$F(1,21)$ = 12.71, p = .003, η^2_p = .377], of vowel quality [$F(4,84)$ = 171.74, p < 0.001, η^2_p = .891] and vowel length [$F(4,84)$ = 2330.65, p < 0.001, η^2_p = .991], a significant interaction between speech style and vowel quality [$F(4,84)$ = 128.32, p < 0.001, η^2_p = .338], between vowel length and vowel quality [$F(4,84)$ = 10.71, p < 0.001, η^2_p = .859], a marginally significant interaction between speech style and vowel length [$F(1,21)$ = 3.67, p = 0.068, η^2_p = .150] and a 3-way-interaction between speech style, vowel quality and vowel length [$F(4,84)$ = 3.27, p = 0.015, η^2_p = .135]. Post-hoc paired comparisons revealed a significantly higher frequency of short low-central vowel /a/ and of short and long high-front vowels /i, i:/ for IDS (Table 4).

3.6 CV association patterns

Before turning to the actual analysis of CV association patterns, we compared the ratio of consonants to vowels in the present corpus to other corpora where this information was available. The consonant/vowel ratio in this corpus was .49/.51 for both IDS and ADS. In English (Lee et al., 2008), 14,450 (IDS) and 14,990 (ADS) consonants per 10,000 vowels were reported, resulting in a consonant/vowel ratio of .59/.41 for IDS and .60/.40 for ADS. In Korean (Lee et al., 2008), 11,800 (IDS) and 12,500 (ADS) per 10,000 vowels were reported, resulting in a consonant/vowel ratio of .52/.48 in IDS and .56/44 in ADS. Thus, in comparison Japanese speech contains the highest rate of vowels, followed by Korean and English. This is consistent with Ramus, Nespor, & Mehler (2000), who found that syllables in stress-timed languages tend to have more complex syllables than syllable-timed languages. In their sample the reported consonant/vowel ratio for English was .60/.40, while Japanese, a mora-timed language, was reported to have the least complex syllables with a ratio of .47/.53. Interestingly, both English and Korean IDS contain fewer consonants than ADS, suggesting that syllables with fewer consonants are favored in IDS.

Table 5. Observed-to-expected ratios in IDS and ADS. Ratio: observed-to-expected ratio of phoneme group. SR: standardized residual of each ratio. Ratios contributing to a siginificant Chi-square effect are typed in bold.

Phonemes analyzed	Speech Style (number of CV sequences analyzed)	Consonant Place	Vowel Position		
			front	**central**	**back**
			ratio (SR)	ratio (SR)	ratio (SR)
			Running speech		
All phonemes	**IDS**	labial	0.78 (-10.5)	**1.22 (10.4)**	0.95 (-0.9)
	(58,411)	coronal	**1.23 (22.7)**	0.97 (-3.4)	0.86 (-16.1)
		dorsal	0.56 (-27.4)	0.96 (-2.5)	**1.37 (25.8)**
	ADS	labial	0.62 (-12.4)	**1.10 (2.8)**	**1.27 (9.3)**
	(29,876)	coronal	**1.17 (14.4)**	0.82 (-13.6)	0.99 (-1.7)
		dorsal	0.64 (-16.9)	**1.54 (22.5)**	0.90 (-3.8)
Plosives and nasals	**IDS**	labial	0.84 (-6.9)	**1.11 (5.3)**	0.97 (0.4)
	(40,182)	coronal	**1.27 (19.0)**	**1.12 (10.0)**	0.67 (-26.9)
		dorsal	0.68 (-18.2)	0.74 (-16.7)	**1.54 (33.3)**
	ADS	labial	0.61 (-13.0)	0.99 (-0.80)	**1.48 (14.8)**
	(20,189)	coronal	**1.27 (17.6)**	0.89 (-6.9)	0.81 (-11.7)
		dorsal	0.65 (-15.7)	**1.26 (10.8)**	**1.13 (5.6)**
			Content-word-initial phonemes		
All phonemes	**IDS**	labial	**1.16 (5.2)**	**1.17 (4.8)**	0.76 (-8.3)
	(12,735)	coronal	**1.09 (3.3)**	0.91 (-3.5)	1.02 (0.6)
		dorsal	0.71 (-9.6)	1.00 (0.1)	**1.19 (7.3)**
	ADS	labial	**1.29 (4.0)**	0.70 (-5.5)	1.04 (1.5)
	(6,457)	coronal	1.01 (0.7)	**1.08 (3.0)**	0.95 (-2.7)
		dorsal	0.80 (-4.4)	1.04 (0.0)	**1.07 (3.0)**
Plosives and nasals	**IDS**	labial	**1.21 (7.1)**	**1.10 (2.7)**	0.75 (-8.3)
	(8,987)	coronal	1.02 (-0.8)	**1.11 (3.1)**	0.93 (-2.4)
		dorsal	0.83 (-6.5)	0.81 (-5.9)	**1.33 (10.9)**
	ADS	labial	**1.29 (4.2)**	0.62 (-7.4)	**1.15 (3.6)**
	(3,470)	coronal	0.89 (-1.9)	**1.63 (13.0)**	0.55 (-10.3)
		dorsal	0.93 (-1.8)	0.65 (-7.0)	**1.35 (7.7)**

Observed-to-expected ratios of serial consonant-vowel organization patterns in IDS and ADS were analyzed. As analyses of early CV association patterns mainly concentrate on stops and nasals, we report separate results on the subgroup of stops and nasals and on all phonemes for running speech and for content-word-initial phonemes.

3.6.1 Running speech

For IDS in stops and nasals, there was an overall significant difference between observed and expected frequencies [$\chi2$ (4, N = 40,182) = 2978.93, p < .001, Cramer's V = .1593]. The four associations significantly contributing to the result were dorsal-back, coronal-front, coronal-central and labial-central. In ADS, there was also an overall significant difference [$\chi2$ (4, N = 20,189) = 1279.11, p < .001, Cramer's V = .193].The four significantly contributing patterns were dorsal-central, labial-back, coronal-front, and labial-central (cf. Table 5). Considering all tokens in all phonemes, both IDS [$\chi2$ (4, N = 58,411) = 2426.29, p < .001, Cramer's V = .144] and ADS [$\chi2$ (4, N = 29,876) = 1449.76, p < .001, Cramer's V = .156] showed overall significant differences in observed and expected frequencies. The observed-to-expected ratios that significantly contributed to the difference in IDS were dorsal-back, coronal-front and labial-central. In ADS, the patterns were dorsal-central, labial-back, coronal-front and labial-central (Table 5).

3.6.2 Content-word-initial phonemes

For stops and nasals of content-word-initial phonemes, both IDS [$\chi2$ (4, N = 8,987) = 337.79, p < .001, Cramer's V = .137] and ADS [$\chi2$ (4, N = 3,470) = 475.25, p < .001, Cramer's V = .262] observed and expected ratios were overall significantly different. Dorsal-back, labial-front, coronal-central and labial-central associations significantly contributed to significance in IDS, whereas in ADS the significantly contributing patterns were in that order coronal-central, dorsal-back, labial-front and labial-back (cf. Table 5). For all phonemes, again both association patterns in IDS [$\chi2$ (4, N = 12,735) = 288.46, p < .001, Cramer's V = .106] and ADS [$\chi2$ (4, N = 6,457) = 92.34, p < .001, Cramer's V = .085] showed overall significant differences. The CV associations contributing to significant differences were dorsal-back, labial-central, labial-front and coronal-front in IDS. For

ADS, significantly contributing association patterns included labial-front, coronal-central and dorsal-back (cf. Table 5).

4 General Discussion

The current study analyzed differences between IDS and ADS in Japanese phonemes and phoneme association patterns in order to identify possible modifications of Japanese IDS. We will first discuss the results separately for consonants, vowels, and CV association patterns, comparing them to results from previous studies in English and Korean. Our discussion will focus primarily on running speech in order to make our results directly comparable to previous findings. In a separate section, we will discuss the results of content-word-initial phonemes where they diverge from running speech. The final section discusses the relevance of IDS modifications on the phoneme level.

4.1 Consonants

In IDS, a significantly higher frequency of labials, dorsals, affricates and liquids, and a lower frequency of coronals, fricatives and nasals were found compared to ADS. The findings for labials and fricatives are consistent with previous reports in Korean and English (Lee et al., 2008; Lee & Davis, 2010), and fit the fine-tuning pattern: an increased use of phonemes that are generally produced early, and a decreased use of phonemes that are generally produced late. The higher frequency of dorsals and affricates in the present study is inconsistent with previous findings on both general tendencies in infant production and IDS. These findings do, however, match early production tendencies in Japanese infants, who do produce dorsal and affricate phonemes from relatively early on. This in turn parallels the high prevalence of both dorsals and affricates (Beckman et al, 2003) in Japanese adult language. Dorsals and affricates in Japanese IDS thus follow the highlighting pattern, a higher frequency of language-specific, not generally early produced phonemes.

The higher frequency of liquid manner and lower frequency of coronal place and nasal manner in Japanese IDS fit neither fine-tuning nor highlighting. The findings regarding coronals are consistent with those

found in Korean (Lee et al., 2008), but not English (Lee & Davis, 2010). Nasals were also less frequent in IDS than ADS in English (Lee & Davis, 2010), although they are among the early produced phonemes (Bernhardt & Stemberger, 1998).

When looking separately at moraic and non-moraic nasals however, moraic nasals were more frequent in Japanese IDS than ADS. Geminate phonemes were also more frequent in IDS than ADS, and these patterns mirror the pattern of geminate and non-geminate nasals in Korean (Lee et al., 2008). Japanese IDV predominantly consists of words with heavy-light and heavy-heavy syllables, which include moraic nasals, geminates or long vowels (Mazuka et al., 2008). Both the moraic nasal and the geminate phoneme are exceptional because they are the only mora types that consist of a single consonant. Therefore, they have a distinct, perceptually salient rhythm, which conceivably helps initial segmentation (Vihman, 1993) and therefore might occur frequently in IDV.

There were significantly more youon in IDS than in ADS. Youon are frequently used in Japanese IDV, where words are often realized with a substitution of affricates for other phonemes or a palatalized form of adult words. The sound symbolism literature proposes palatalized sounds to be associated with "childishness and immaturity" (Hamano, 1998). Since many youon are palatalized, their use may be a way of fine-tuning. Interestingly, increased palatalization after dentals has also been reported for English child-directed speech (Bernstein Ratner, 1996). Future studies are necessary to investigate if there is an auditory or productive preference for youon-like sounds in infants.

4.2 Vowels

A higher ratio of short and long low-central as well as of long high-front vowels and a lower ratio of short high-back, long mid-back and short mid-front vowels were found in IDS compared to ADS. Previous studies have reported a higher occurrence of the lower left quadrant vowels in early productions (e.g., Davis & MacNeilage, 1990) and IDS (e.g., Lee et al., 2008). Of these, the Japanese vowel inventory only possesses the low-central vowels. These were indeed more frequent in IDS compared to ADS, thus supporting fine-tuning, the increased use of phonemes that are

generally produced early. The findings on long high-front, short high-back, long mid-back and short mid-front vowels do not fit previous results, however.

Concerning vowel length, a higher ratio of long vowels was found for ADS, showing that prosodic vowel lengthening does not affect phonological vowel length. Consistent with this, Japanese mothers maintained two distinct phonological vowel length categories despite their use of non-lexical vowel lengthening in IDS (Werker et al., 2007).

4.3 CV association patterns

In previous reports, frequent early association patterns were labial-central, coronal-frontal and dorsal-back in MacNeilage et al. (2000), and labial-central, coronal-central and dorsal-back in Vihman (1992). The patterns found for IDS in the present study very closely resembled all of these articulatory patterns. The analysis of stops and nasals perfectly mirrored all four suggested patterns in running speech plus the patterns suggested by Vihman (1992) in content-word-initial phonemes, while the analysis of all phonemes mirrored the three patterns of MacNeilage et al. (2000) in both running speech and word-initial content-words.

In contrast, ADS showed a correspondence to the suggested patterns only in parts, and not consistently across analyses. The rankings of observed-to-expected ratios were highly variable in ADS, while they were fairly consistent across analyses in IDS. Although IDS and ADS were not compared directly, these data show that the pattern of observed-to-expected ratios in IDS matches the suggested basic production patterns more closely than ADS. Thus, Japanese mothers are producing CV association patterns that correspond to the suggested basic production patterns of children in IDS but not ADS, providing support for the fine-tuning account.

Labial-central and dorsal-back patterns were constantly the most frequent association patterns in IDS. This is in line with both Vihman's (1992) and MacNeilage et al.'s (2000) suggestion of labial-central as the most basic of productive association patterns. Both authors had also reported the dorsal-back pattern as a preferred grouping. In the languages studied previously, however, dorsal consonants were infrequent, and

Vihman (1992) suggested that a preference for this pattern may show later in development with more frequent use of back consonants and vowels. As studies both in Mandarin Chinese (Chen & Kent, 2005) and Korean (Lee, Davis, & MacNeilage, 2007) found a relationship between IDS and early production of CV association patterns, future studies of Japanese should investigate how far the dorsal-back association pattern is preferred in early productions given a comparatively high amount of dorsals.

4.4 Differences between running speech and content-word-initial phonemes

In addition to running speech, the present study reported phoneme distributions for the sub-group of content-word-initial phonemes, because these are known to be especially salient to infants (Boysson-Bardies & Vihman, 1991; Shi & Werker, 2001) and to differ in their distributional properties (Vihman et al., 1994). With a few exceptions, the direction of results was the same for these two types of samples. Statistical analyses sometimes showed significant differences for one, but not the other type of analysis, though.

For consonants of content-word-initial phonemes, labials, stops and liquids were more frequent in IDS and coronals, fricatives and glides were more frequent in ADS. The difference in stop frequencies between IDS and ADS did not reach significance in running speech, but does also fit into the picture of IDS containing articulatory simple phonemes and is consistent with the results in Korean and English (Lee et al., 2008; Lee & Davis, 2010). The occurrence of word-initial vowels was generally low, which is due to the moraic structure of Japanese, where vowels mostly follow a consonant. Among these, low-central /a/, as well as high-front /i/ and /i:/ had a significantly higher ratio in IDS. CV association patterns of IDS plosives and nasals mirrored the patterns suggested by Vihman (1992), and association patterns of all phonemes mirrored the patterns suggested by MacNeilage et al. (2000) in content-word-initial phonemes.

4.5 The role of IDS at the phoneme level

The fine-tuning account describes a pattern predicting language-general emphasis of phonemes that are acquired early in general. It argues that caregivers match their speech to their infants' production capacities, which was originally suggested based on correlational analyses of mothers' and children's speech in English (Cross, 1977). In Japanese, Murase, Ogura & Yamashita (1992) reported an increase of caregivers' use of adult forms and decrease of baby-talk forms between 1;10-2;2 years of age, corresponding to the age where Japanese children start producing adult forms. Matching speech to infants' productions makes sense according to Vihman's (1993) articulatory filter model, which suggests that infants perceive input matching their own productions as especially salient, picking up those patterns for which they already have a motor representation. The prevalence of patterns that match early production tendencies in IDS in the current study and other languages studied so far suggests fine-tuning as one way in which IDS differs from ADS on the phoneme level.

At the same time, language-specific differences in the distribution of phonemes in IDS compared to ADS clearly show that fine-tuning is not the only way phoneme distributions in IDS are modified. One source of these differences could be mothers' highlighting of phonemes that are prevalent in the native language but are not produced early in general. For Japanese, we specifically predicted dorsals and affricates to be more frequent in IDS than ADS, and we indeed found this to be the case. Based on these results, we suggest highlighting as a further way in which IDS could be modified on the phoneme level, but further studies are necessary to strengthen this account.

Other differences in patterns across languages cannot easily be explained by either fine-tuning or highlighting. These differences could be due to some systematic language-specific factors at the phonological or lexical level. As for the phonological level, language-specific phoneme inventories could make important contributions. For example, Korean mothers in Lee et al. (2008) did not highlight dorsal phonemes in IDS even though they are highly frequent. The simple phonotactics of Japanese, in particular the low frequency of consonant clusters, may contribute to a higher amount of highlighting in Japanese IDS by inducing less pressure to

avoid certain phonemes: The acquisition of consonants in clusters is late compared to singletons (McLeod, Doorn, & Reed, 2001) and infants pay more attention to consonants in syllable onsets than to those in codas (Vihman et al., 1994). In contrast to Japanese, English and Korean both allow CVC syllables, and consonant clusters occur frequently. The observation that the consonant-to-vowel ratio of Japanese IDS does not differ from ADS, while it decreases in both Korean and English, speaks to this interpretation.

On the lexical level, Lee and Davis (2010) assign part of the differences between Korean and English to the different usage of IDS: Lee and Nakayama (2000) found that Korean and Japanese mothers frequently use specific infant-directed vocabulary like nonsense words and onomatopoeia, which American mothers (Fernald & Morikawa, 1993) do less frequently. Our findings of an increased use of geminates, moraic nasals and youon in IDS are likely to reflect Japanese mothers' frequent usage of such lexical items. Lee and Nakayama, based on reports of differences in Korean, Japanese and American mothers' speech, additionally proposed a role of cultural differences in lexical choice: Korean mothers frequently use verbs to teach actions, Japanese mothers use words related to social actions to teach social skills and American mothers use nouns to teach object names.

This latter proposal touches upon an important point: the modifications in IDS phoneme distributions may not be an end of their own, but rather a by-product of other modifications. For example, Trainor et al. (2000) found that both emotional adult speech and IDS contain more exaggerated vowel contours than unemotional adult speech, which suggests that they are rather a by-function of emotional expression. Similarly, Johnson, Lahey, Ernestus, and Cutler (2013) found that utterance lengths in IDS were more similar to utterance lengths in caregivers' speech directed to other familiar adults than to caregivers' speech directed to other unfamiliar adults. Moreover, phoneme distributions in IDS might be a by-product of lexical choice (cf. Daland, 2012). For example, the higher frequency of the long high-front vowel in IDS is likely related to lexical factors, because in Japanese the word *ii* means 'good'. In the current study, the word *ii* comprised 24% of all long high-front vowels in ADS, while it comprised 42% in IDS.

Another factor to be considered is that IDS may change during development, with caregivers' input adapting to the needs of the infant in a certain stage (Cross, 1977). The age-range 1;6-2;0 in the current study differs from the one-year-olds in the previous studies, which may impair comparison. Following studies on developmental changes in IDS on the acoustic (e.g., Kitamura et al., 2001) and semantic (e.g., Snow, 1977) aspects of IDS, future studies should track such changes on the phoneme level.

Lastly, the caregivers' IDS is not the only input for infants (Soderstrom, 2007): ADS, as well as siblings' speech (directed to the infant or to the caregivers), occurs frequently in infants' environment. It is still not clear to what extent the speech not directly addressed to the infant influences language acquisition. The current study considers both IDS and ADS, providing a starting point for comparing the impact of these speech styles. Further studies in the tradition of Vihman et al. (1994) and Chen & Kent (2005) are necessary to investigate the exact relationship between IDS, other speech styles and phoneme acquisition.

5 Conclusion And Outlook

Overall and consistent with previous studies (Lee et al., 2008; Lee & Davis, 2010), we found evidence for an increased use of phonemes and association patterns that occur early in production in IDS compared to ADS (fine-tuning). We also found evidence for an increased use of phonemes and association patterns that are acquired rather late overall but that are very frequent and acquired early in Japanese (highlighting). Concerning the latter, it is not clear to what extent this pattern is specific to Japanese or whether it can be generalized across languages. Moreover, some additional differences between IDS and ADS cannot be explained by fine-tuning or highlighting. These differences could be due to language-specific phonological or lexical factors, or a by-product of other factors. Further research in additional languages and corpora is necessary to assess these alternatives.

Finally, we want to address the potential relevance of such phoneme differences between IDS and ADS for language acquisition. Daland (2012) points out that, even if IDS and ADS phoneme distributions

differ, such small differences are unlikely to affect phoneme category learning. To date, there is indeed no study that looks at the effect of small frequency differences in the input. Thus, caregivers' fine-tuning might just be a way caregivers adjust their speech to infants' production capabilities without necessarily impacting phoneme category learning in a significant way.

Acknowledgements

References

Beckman, M. E., Yoneyama, K. & Edwards, J. (2003). Language-specific and language-universal aspects of lingual obstruent productions in Japanese-acquiring children. *Journal of the Phonetic Society of Japan, 7,* 18-28.

Bernhardt, B. H. & Stemberger, J. P. (1998). *Handbook of Phonological Development.* San Diego, CA: Academic Press.

Boysson-Bardies, B. & Vihman, M. M. (1991). Adaptation to language: evidence from babbling and first words in four languages. *Language, 67,* 297-319.

Chen, L.-M. & Kent, R. D. (2005). Consonant-vowel co-occurrence patterns in Mandarin-learning infants. *Journal of Child Language, 32*, 507-534. doi:10.1017/S0305000905006896

Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). Mahwah, NJ: Lawrence Earlbaum Associates.

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass,* 7(3), 157–170. doi:10.1111/lnc3.12015

Cross, T. G. (1977). Mothers' speech adjustments: the contributions of selected child listener variables. In K. Snow & C. A. Ferguson (Eds.), *Talking to children: language input and acquisition* (pp. 151-188). Cambridge: Cambridge University Press.

Daland, R. (2012). Variation in the input: a case study of manner class frequencies. *Journal of Child Language*. Available on CJD 2012. doi:10.1017/S0305000912000372

Davis, B. L., & MacNeilage, P.F. (1990). Acquisition of Correct Vowel Production. *Journal of Speech and Hearing Research, 33*, 16-27.

Dyson, A. T., & Amayreh, M. M. (2007). Jordanian Arabic speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 288-299). Clifton Park, NY: Thomson Delmar Learning.

Edwards, J. & Beckman, M. E. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in the acquisition of consonant phonemes. *Language Learning & Development, 4*(2), 122-156. doi:10.1080/15475440801922115

Fernald, A. & Morikawa, H. (1993). Common Themes and Cultural Variations in Japanese and American Mothers' Speech to Infants. *Child Development, 64*(3), 637-656. doi:10.1111/j.1467-8624.1993.tb02933.x

Ferguson, C.A. (1977). Baby talk as a simplified register. In K. Snow & C. A. Ferguson (Eds.), *Talking to children: language input and acquisition* (pp. 209-235). Cambridge: Cambridge University Press.

Fikkert, P. (1994). On the acquisition of rhyme structure in Dutch. In R. Bok-Bennema & C. Cremers (Eds.), *Linguistics in the Netherlands 1994* (pp. 37-48). Amsterdam: John Benjamins.

Fikkert, P., & Levelt, C. C. (2008). How does Place fall into Place? In P. Avery, B. E. Dresher & K. Rice (eds.), *The lexicon and emergent constraints in children's developing phonological grammar. Contrast in phonology: theory, perception, and acquisition* (pp. 231-268)*.* Berlin: Mouton.

Fox, A. V. (2007). German speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 386-397). Clifton Park, NY: Thomson Delmar Learning.

Goldstein, B. A. (2007). Spanish speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 539-553). Clifton Park, NY: Thomson Delmar Learning.

Hamano, S. (1998). *The Sound-Symbolic System of Japanese. Studies in Japanese Linguistics, Vol. 10*. CSLI Publications, Stanford, CA, & Kuroshio Publishing Company, Tokyo.

Howard, S. (2007). English speech acquisition. In S. McLeod (Ed.), *The International guide to speech acquisition* (pp. 188-203). Clifton Park, NY: Thomson Delmar Learning.

Ingram, D. (1999). Phonological acquisition. In M. Barrett (ed.), *The development of language* (pp. 73-97). Hove, UK: Psychology Press.

Jacobson, R. (1941/1968). *Child language, aphasia and phonological universals.* The Hague: Mouton.

Johnson, E., Lahey, M., Ernestus, M., & Cutler, A. (2013). A multimodal corpus of speech to infant and adult listeners. *JASA Express Letters, 134*(6), EL534-EL540. doi: 10.1121/1.4828977

Kent, R. D. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 65-90). Parkton, MD: York Press.

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development, 24*(4), 372-392. doi:10.1016/S0163-6383(02)00086-3

Kunnari, S., & Savinainen-Makkonen, T (2007). Finnish speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 351-363). Clifton Park, NY: Thomson Delmar Learning.

Lee, S. A., Davis, B. & MacNeilage, P. (2007). 'Frame Dominance' and the Serial Organization of Babbling, and First Words in Korean-Learning Infants. *Phonetica, 64,* 217-236. doi:10.1159/000121374

Lee, S. A., Davis, B. & MacNeilage, P. (2008). Phoneme properties of input to infants: a study of Korean. *Journal of Child Language, 35*, 591-617. doi:10.1017/S0305000908008684

Lee, S. A. & Davis, B. (2010). Phoneme distribution patterns of English infant-and adult-directed speech. *Journal of Child Language, 37*(4), 767-791. doi:10.1017/S0305000909009568

Lee, S. A. & Nakayama, M. (2000). Characteristics of maternal speech in Korean: do Korean and Japanese maternal speech show similar characteristics? In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the Annual Boston University Conference on Language Development 24* (pp*.* 486-497). Boston: Cascadilla Press.

Lorwatanapongsa, P., & Maroonroge, S. (2007). Thai speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 554-565). Clifton Park, NY: Thomson Delmar Learning.

MacNeilage, P. & Davis, B. (2000). On the Origin of Internal Structure of Word Forms. *Science, 288*(5465), 527-531. doi:10.1126/science.288.5465.527

MacNeilage, P., Davis, B., Kinney, A. & Matyear, C.L. (2000). The Motor Core of Speech: A Comparison of Serial Organization Patterns in Infants and Languages. *Child Development, 71*(1), 153-163. doi:10.1111/1467-8624.00129

Maekawa, K. (2003). Corpus of Spontaneous Japanese: Its design and evaluation. *Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003)* (pp. 7-12). Tokyo.

Mazuka, R., Kondo, T. & Hayashi, A. (2008). Japanese Mothers' Use of Specialized Vocabulary in Infant-Directed Speech: Infant-Directed Vocabulary in Japanese. In N. Masataka (Ed.), *The Origins of Language* (pp. 39-58). Tokyo: Springer.

Mazuka, R., Igarashi, Y., & Nishikawa, K. (2006). Input for Learning Japanese: RIKEN Japanese Mother-Infant Conversation Corpus. *The Institute of Electronics, Information and Communication Engineers Technical Report*, 16, 11–15.

McLeod, S., Doorn, J.V., & Reed, V.A. (2001). Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathology*, 10(2), 99-110. doi:10.1044/1058-0360(2001/011)

Mennen, I., & Okalidou, A. (2007). Greek speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 398-411). Clifton Park, NY: Thomson Delmar Learning.

Murase, T., Ogura, T. & Yamashita, Y. (1992). Ikujigo no kenkyuu (1). Doobutsu meishoo ni kansuru hahaoya no shiyoogo: Ko no geturei ni yoru chigai. [Study of child-rearing vocabulary (1). Mothers' use of animal terms: Effects of chidren's age]. *Annual Bulletin of Shimane University, Faculty of Education, 17,* 37-54.

Ramus, F., Nespor, M., & Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. *Cognition, 75*, AD3–AD30. doi:10.1016/S0010-0277(00)00101-3

Bernstein Ratner, N. (1996). From Signal to Syntax - But what is the nature of the signal? In J. Morgan, & K. Demuth (Eds.), *From signal to syntax: bootstrapping from speech to grammar in early acquisition* (pp. 35-150). Hillsdale, NJ: Erlbaum.

Rose, Y., & Wauquier-Gravelines, S. (2007). French speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 364-385). Clifton Park, NY: Thomson Delmar Learning.

Saito, Y. (1997). *Nihongo no onseigaku nyumon* [Introduction to Japanese Phonetics]. Tokyo: Sanseido.

Shi, R. & Werker, J. F. (2001). Six-month-old infants' preference for lexical words. *Psychological Science, 12*(1), 70-75. doi:10.1111/1467-9280.00312

Smit, A. B. (2007). General American English speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 128-147). Clifton Park, NY: Thomson Delmar Learning.

Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language, 4*, 1–22. doi:10.1017/S0305000900000453

So, L. K. H. (2007). Cantonese speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 313-326). Clifton Park, NY: Thomson Delmar Learning.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27*(4), 501-532. doi:10.1016/j.dr.2007.06.002

Trainor, L. J., Austin, C. M. & Desjardins, R. N. (2000). Is Infant-Directed Speech Prosody a Result of the Vocal Expression of Emotion? *Psychological Science, 11*(3), 188-195. doi:10.1111/1467-9280.00240

Uno, A. (2007). *Kotoba to kokoro no hattatsu to shougai* [Development and disorders of language and mind]*.* Tokyo: Nagaishoten.

Vance, T. J. (1987). *An introduction to Japanese phonology*. Albany: State University of New York Press

Vihman, M. M. (1992). Early syllables and the construction of phonology. In: C.A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological Development. Models, Research, Implications* (pp. 393–422). Timonium, MD: York Press.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics, 21*, 61-82.

Vihman, M. M., Kay, E., Boysson-Bardies, B., Durand, C. & Sundberg, U. (1994). External sources of individual differences? A cross-linguistic analysis of the phonetics of mothers' speech to one-year-old children. *Developmental Psychology, 30*(5), 651-662. doi:10.1037/0012-1649.30.5.651

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L. & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition, 103*(1), 147-162. doi:10.1016/j.cognition.2006.03.00

Even at 4 months, a labial is a good enough coronal, but not vice versa

Chapter 7

Based on:

Tsuji, S., Mazuka, R., Cristia, A., & Fikkert, P. (2013). *Even at 4 months, a labial is a good enough coronal, but not vice versa.* Manuscript submitted for publication.

Abstract

Numerous adult studies have revealed an asymmetry tied to the perception of coronal place of articulation: Participants accept a labial mispronunciation of a coronal target, but not vice versa. Whether or not this asymmetry is based on language-general properties or arises from language-specific experience has been a matter of debate. The current study suggests a bias of the first type by documenting an early, cross-linguistically valid asymmetry related to coronal place of articulation. Japanese and Dutch 4- and 6-month-old infants showed evidence of discrimination if they were habituated to a labial and then tested on a coronal sequence, but not vice versa. This finding has important implications for both phonological theories and the field of infant speech perception.

1 Introduction

The special status attributed to coronal place of articulation in the phonologies of the world (Paradis & Prunet, 1991) has intrigued phonologists for decades. Indeed, coronals (sounds that are articulated with the tip or blade of the tongue) show distinct characteristics, such as a high frequency of occurrence between and within languages (Maddieson, 1984), and a proneness to undergo phonological processes such as place assimilation (e.g., Marslen-Wilson, Nix, & Gaskell, 1995). This special status would also affect speech processing in the form of a perceptual asymmetry, but whether or not this asymmetry is based on language-general properties or arises from language-specific experience has been a matter of debate. The current study suggests that the special status of coronals need not be acquired by documenting a language-independent asymmetry in young infants. Before coming back to this central aim, we will introduce two mainstream accounts put forward to capture perceptual asymmetries in adults.

The Featurally Underspecified Lexicon (FUL; Lahiri & Reetz, 2010) assumes sparse and abstract lexical representations in which not all phonological features are specified. Based on universal properties of phonological systems, coronal place of articulation is assumed to be the default place of articulation and is therefore underspecified in the mental lexicon. This assumption predicts perceptual asymmetries such that labial mispronunciations of coronal phonemes (as in [bɔl] for /dɔl/) do not produce a mismatch (thus [bɔl] is accepted as an instance of /dɔl/), but coronal mispronunciations of labials do (thus [dɔl] is not accepted for /bɔl/). The results of numerous perceptual experiments are consistent with this prediction: labial mispronunciations prime coronal target words, but not vice versa, in cross-modal priming (Lahiri & Reetz, 2002). Similarly, event-related potential (ERP) studies with mispronunciations of real words or non-words have shown smaller ERPs to labial mispronunciations of coronals than vice versa (see e.g., Cornell, Lahiri, & Eulitz, 2013; and references therein). One limitation of the above evidence is that adult studies cannot disentangle the role of universal phonological properties from language experience because in adults the two are

confounded (even more so because the majority of the evidence comes from native speakers of Germanic languages).

Indeed, other work fails to support the predictions of FUL. Several studies using designs similar to the studies above found a pattern of asymmetries that was interpreted as not consistent with FUL predictions (Bonte, Mitterer, Zellagui, Poelmans, & Blomert, 2005; Mitterer, 2011). Based on this, Mitterer (2011) suggested an alternative account based on optimal perception. This account suggests that asymmetries reflect listeners' familiarity with the phonotactic probability of the input. Listeners are biased towards accepting a frequent pattern more often than an infrequent one. Given that coronals are very frequent, this account would predict asymmetries consistent with the prediction made by FUL in many cases, but based on the fundamentally different premise of language-specific experience. For instances in which the accounts make different predictions, currently evidence in support of both sides exists (Cornell et al., 2013; Mitterer, 2011), and it is therefore difficult to reach a conclusive answer with regard to the role of language experience in observed asymmetries.

The current study assesses the possibility that experience-independent perceptual biases contribute to observed consonant asymmetries by testing prelexical infants. Based on earlier research, we can assume that infants' perception is tuned to native consonant categories and phonotactic probabilities well after 6 months of age (cf. Kuhl, 2004, for an overview). Finding asymmetries in 6-month-old Dutch infants would suggest that asymmetries are independent of the extensive lexical and phonotactic experience assumed above (see Dijkstra & Fikkert, 2010, for relevant evidence). Nonetheless, as more sensitive methods appear, age of acquisition is constantly being pushed down (e.g., Bergelson & Swingley, 2012). Therefore, we built an even stronger test of the language-independent nature of such perceptual asymmetries by measuring discrimination at two early ages (4 and 6 months), with two language backgrounds with markedly diverse phonologies, namely Dutch and Japanese. Japanese is particularly illuminating, since in that language coronals do not act as the default place of articulation (Labrune, 2012), illustrated for instance by the higher frequency of dorsals compared to coronals (Beckman, Yoneyama, & Edwards, 2003). If experience-

independent perceptual biases can indeed contribute to perceptual asymmetries, then a coronal-labial asymmetry should be observed in infants (similarly to that found in adults) regardless of age and language background.

2 Methods

2.1 Participants

Sixteen 4-month-old Dutch (range 3.7 – 4.5 months, 8 females) and sixteen 4-month-old Japanese (4.0 – 5.0 months, 6 females) infants, as well as sixteen 6-month-old Dutch (range 6.4 – 6.9 months, 10 females) and sixteen 6-month-old Japanese (range 6.1 – 7.0 months, 9 females) infants were included in the final sample. The Dutch infants were recruited and tested in the Netherlands, and the Japanese infants were recruited and tested in Japan. All infants were healthy full-term infants, raised in monolingual native Dutch or Japanese speaking households. Caregivers of all participants gave written consent to participate in the study.

Twenty-nine additional infants were tested but not included in the final sample because of failure to reach the habituation criterion (7 Dutch, 1 Japanese), obscured view on the infant's eyes (1 Dutch), failure to look at the screen after experiment commencement (2 Japanese), fussiness or crying (8 Dutch, 10 Japanese).

2.2 Stimuli

A word-medial consonant cluster was chosen to provide rich acoustic cues to the listeners, thus affording them every opportunity to hear the place contrast. That is, in addition to the small burst and fast transitions found also in e.g. /paːn-taːn/, /ɔmpa-ɔnta/ also contains rich formant transitions into the nasals' place and some information in the nasal murmur. Both sequences were phonotactically legal in both Dutch and Japanese, although the frequency of the sequence /ɔnta/ is higher than that of /ɔmpa/ in Dutch (Baayen, Piepenbrock, & Gulikers, 1995), whereas the opposite is true in Japanese (Amano & Kondo, 2000). Notice

that this divergence in frequency should bias Japanese and Dutch infants into opposite directions.

Multiple tokens of /ɔmpa/ and /ɔnta/ were recorded by a female native speaker of Dutch in an infant-directed register. Eight tokens per sequence were selected. These were matched on mean duration (/ɔmpa/: 770 ms, /ɔnta/: 740 ms), and on mean pitch (/ɔmpa/: 287 Hz, /ɔnta/: 258 Hz). Five of the eight tokens of each type were used in the labial and coronal habituation lists. Test lists also contained five tokens, of which two had appeared in the habituation lists, and three were novel. This mixture of habituated and novel tokens helps exclude the possibility of dishabituation based on novel tokens alone. Four habituation lists and two test lists were created per sequence by pseudo-randomizing order of tokens. A one-second pause was inserted between each token, and the mean length of the lists was 14.1 seconds.

The visual stimulus accompanying auditory stimulus presentation was a dynamic checkerboard presented in the middle of the screen. Between trials, a video of a flashing red light was displayed as an attention getter. For the pre-and posttest, a colorful smiley rolling along the edges of the screen was presented accompanied by multiple tokens of the sound /ni:m/ recorded by a female native speaker of Dutch.

2.3 Procedure

A slightly modified version of the Central Fixation paradigm (Werker et al., 1998) was used. Infants sat on a caregiver's lap, facing the screen. Sounds were presented from loudspeakers, and the infant's gaze pattern was recorded by a video camera directly below the screen. The parent wore sound-attenuating headphones with masking music in order to not influence the infant. The experimenter also wore headphones with masking music (Netherlands) or was in a different room where she could not hear any of the stimuli (Japan). Infants' gaze on or off the screen was coded online by a trained experimenter. The experiment was administered with the software Habit X (Cohen, Atkinson, & Chaput, 2000), and started with the presentation of the attention getter. Once the infant looked at the screen, the pre-test was initiated, followed by the habituation phase, in which the dynamic checkerboard and habituation stimuli were presented. Stimulus presentation was fixed. The four

habituation lists were presented in four different pseudo-random orders across infants. Half of the infants were habituated to lists of /ompa/ tokens, and half of the infants were habituated to lists of /onta/ tokens. The habituation criterion was a decrease to 60% of the looking time of the first trials, calculated over a sliding window of 4 trials. Once criterion was reached or the infant reached the maximum trial number of 28, the test phase started. During the test phase, infants were presented with 4 trials. Half of these were *same* trials in which a test list with the same type of sequence as during habituation was presented, and half were *switch* trials in which a test list with the different sequence was presented. The order of presentation was always same-switch-same-switch. The experiment was terminated if the infant started crying of fussing extensively. Gaze behavior was recoded offline by one trained coder, and 25% of videos were independently coded by a second coder. Inter-rater reliability was 98 %. The looking times resulting from the codings of the first coder were entered into further analysis.

2.4 Results and Discussion

A preliminary ANOVA on looking times in pre-and post-test trials by age and language revealed no significant differences between pre- and post-test trials [$F(1,60) = 0.40$, $p = 0.527$], and no significant interactions with age or language background. This indicates that there was no significant difference in attention between infant groups, or between pre- and post-test.

A direction (labial to coronal, coronal to labial) x trial type (same, switch) x age-group (4 months, 6 months) x language background (Dutch, Japanese) mixed ANOVA was calculated. There was a significant main effect of trial type [$F(1,56)= 4.76$, $p = .03$, $\eta^2 = .024$], with a longer looking time in switch (m = 8.84 s, SD = 2.98) than in same trials (m = 8.02 s, SD = 2.64) and a significant interaction between trial type and direction [$F(1,56) = 10.97$, $p = .001$, $\eta^2 = .054$]. No other effects approached significance; in particular, although there was a marginal effect of language background [$F(1,56) = 3.29$, $p = .075$, $\eta^2 = .040$], the interaction term between language background and trial type was *not* significant [$F(1,56)= 0.7$, $p = .406$, $\eta^2 = .003$] nor was the three-way between

language background, trial type, and direction [$F(1,56)= 0.02$, $p = .891$, $\eta^2 = .00009$].

Bonferroni-corrected post-hoc t-tests following up on the interaction between trial type and direction indicated that the effect of trial type was only significant for the direction labial to coronal ($m_{same} = 7.61$ s, $m_{switch} = 9.68$ s, $p = .019$, $d = 1.23$), but not vice versa ($m_{same} = 8.42$ s, $m_{switch} = 7.99$ s, $p = 1.00$, $d = -0.26$). For the labial to coronal direction 24 out of 32 infants (75%) looked longer to 'switch' than to 'same' trials, while that was only true for 15 out of 32 infants (47%) for the direction coronal to labial.
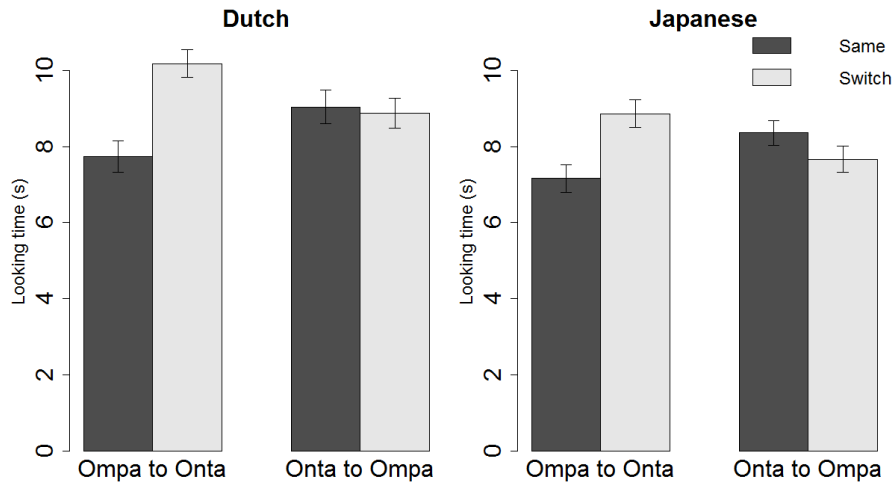

Figure 1. Looking times to same and switch trials by direction of presentation and language background. Graphs are collapsed over age groups.

3 General Discussion

The current study found evidence for a perceptual directional asymmetry in a word-medial labial-coronal consonant contrast for 4- and 6-month-old Dutch and Japanese infants: they were able to perceive the difference between labial and coronal consonants if habituated to labial, but not if habituated to coronal consonants. In other words, labial tokens

were accepted as instances of coronals, but not vice versa. While there is some evidence that frequency of occurrence shapes infant perception (e.g., Anderson et al. 2003; Pons et al., 2012), it is extremely unlikely that experience would explain the current pattern of results. To begin with, previous work has found effects of frequency towards the end of the first year (at 9 to 12 months) and not at 4-6 months, at an age where there is vanishingly little evidence of language-specific consonant or sequence perception. Furthermore, although the phonotactic frequency of the experimental tokens would have biased infants in opposite directions, the asymmetry was independent of infants' language background and age.

The direction of asymmetry found in the current study is consistent with the predictions of FUL (Lahiri &Reetz, 2010). However, our infant discrimination data cannot inform us as to whether or not the observed asymmetry can be grounded in phonological underspecification or reflects basic, language-independent perceptual processes. It should be noted that our results are less consistent with the optimal perception account (Mitterer, 2011), as we documented a perceptual asymmetry independent of language experience. Nonetheless, the fact that prelexical infants show an asymmetry does not preclude the possibility that their perception could be further modulated by characteristics of the ambient language later on. Incipient evidence for such a possibility comes from Tsuji, Yamane, Fikkert, & Mazuka (2014) who show an asymmetry between labials and coronals in the predicted direction in 18-month-old Dutch, but not in same-aged Japanese toddlers. Thus, early language-general biases might be tuned into language-specific ones after the first year of life, which at that age likely reflect language-specific phonological properties. Further studies in toddlers and/or adults in languages of which the phonology or frequency distributions would not predict an asymmetry in the predicted direction are necessary to shed light on this issue.

In fact, asymmetries in prelexical infants' speech perception have played a major role in vowel discrimination research. Discrimination towards vowels that are in a more peripheral position in the vowel space (a two-dimensional space defined by the first two vowel formants) has been reported to be easier than into the other direction (Polka & Bohn, 2011, for an overview). However, the mechanisms proposed to underlie this perceptual bias (cf. Schwartz, Abry, Boe, Menard, & Vallee, 2005) are

difficult to apply straightforwardly to consonants, which show very different acoustic properties. Instead, two possibilities for the mechanism underlying the present consonantal asymmetry suggest themselves, namely the perceptual properties of coronal-labial sequences, and their categorization. According to the former explanation, coronal tokens would be inherently less discriminable from labial tokens than vice versa, based on some characteristics of the signal. There is, however, little evidence for this based on either acoustic measures or perceptual assessments. For instance, the burst of /t/ has a high frequency energy compared to /p/, and confusion matrices show similar rates of confusions from /pa/ to /ta/ as from /ta/ to /pa/ in English adults (Ladefoged, 2012). The second explanation suggests that some cases of perceptual asymmetries, particularly those where multiple tokens are used, could arise from how our cognitive system categorizes variable stimuli. For example, it is conceivable that coronal tokens show a higher variability on some acoustic dimension than labial ones, and that this reduces infants' ability to detect a change in this direction. Such a mechanism could either operate based on previous experience with token distributions in the ambient language, in which case the current study would suggest that token variability for coronals is higher in both Dutch and Japanese. Alternatively, the distributions experienced during the experiment itself could lead to the found asymmetries. Indeed, in the domain of visual categorization, 3-to 4-month-old infants have been shown to discriminate cats and dogs in an asymmetric fashion based on the category variability they were exposed to during an experiment: when the category variability of dog faces was higher than of cat faces, infants failed to detect a cat face. This effect disappeared when variability was matched across categories (Quinn, Eimas, & Rosenkrantz, 1993).

　　Regardless of the source of perceptual asymmetries, our data underline the importance of thinking of speech perception in the context of a general auditory and cognitive system at very early stages of development. Hybrid models, incorporating both innate biases and equally crucial linguistic experience, are gaining increasing support in certain areas of psycholinguistic research (e.g., Endress & Mehler, 2010). By relying on general perceptual biases, such models are attractive because they reduce the learning problem faced by an unbridled

statistical learner while maintaining the crucial contribution of language exposure.

In sum, the results of the current study provide evidence for an early and language-general labial-coronal perceptual asymmetry. Further research assessing the mechanisms behind this early perceptual bias will not only shed further light on the origins of this asymmetry and the special role of coronals in the languages of the world, but also speak to the goal of a perceptually and cognitively grounded theory of infant speech perception.

Acknowledgments

References

Amano, S., & Kondo, T. (2000). *Nihongo-no Goi-tokusei* [Lexical properties of Japanese] (Vol. 7). Tokyo: Sansei-do

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. doi: 10.1177/00238309030460020601

Baayen, R.H., Piepenbrock, R. and Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM).* Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Beckman, M. E., Yoneyama, K. & Edwards, J. (2003). Language-specific and language-universal aspects of lingual obstruent productions in Japanese-acquiring children. *Journal of the Phonetic Society of Japan, 7,* 18-28.

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meaning of many common nouns. *PNAS, 109*(9), 3253-3258. doi: 10.1073/pnas.1113380109

Bonte, M., Mitterer, H., Zellagui, N., Poelmans, N., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology, 116,* 2765. doi:10.1016/j.clinph.2005.08.012

Cohen, L.B., Atkinson, D.J., & Chaput, H. H. (2004). *Habit X: A new program for obtaining and organizing data in infant perception and cognition studies* (Version 1.0)*. Austin: University of Texas.

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. doi: 10.1037/a0030862

Dijkstra, N. and P. Fikkert (2011). Universal constraints on the discrimination of Place of Articulation? Asymmetries in the discrimination of 'paan' and 'taan' by 6-month-old Dutch infants. In N. Danis, K. Mesh, & H. Sung (Eds.) *Proceedings of the Boston University Conference on Language Development 35* (pp. 170-182).

Endress, A. D., & Mehler, J. (2010). Perceptual Constraints in Phonotactic Learning. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 235-250. doi: 10.1037/a0017164

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

Labrune, L. (2012). *The Phonology of Japanese*. Oxford, England: Oxford University Press.

Ladefoged, P., & (2004). *Vowels and Consonants: An introduction to the sounds of languages* (2nd edition). Oxford, UK: Blackwell Publishers.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 637–677). Berlin, Germany: Mouton de Gruyter.

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

Maddiesson, I. (1984*). Patterns of Sounds*. New York: Cambridge University Press.

Marslen-Wilson, W. D., Nix, A., & Gaskell, M. G. (1995). Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes, 10,* 285–308. doi: 10.1080/01690969508407097

Mitterer, H. (2011). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 496–513. doi:10.1037/a0020989

Paradis, C., & Prunet, J.-F. (Eds.) . (1991). *The special status of coronals: internal and external evidence (Phonetics and Phonology 2).* San Diego: Academic Press.

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

Pons, F., Albareda-Castellot, & 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. doi: 10.1111/j.1467-8624.2012.01740.x

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception, 22*(4), 463-475. doi:10.1068/p220463

Schwartz, J.-L., Abry, C., Boe, L.-J., Menard, L., & Vallee, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion-Focalisation Theory. *Speech Communication*, *45*, 425-434. doi: 10.1016/j.specom.2004.12.001

Tsuji, S., Fikkert, P., Yamane, N., & Mazuka, R. (2014). *Specifying the underspecified: Young children's perceptual asymmetries are both language-specific and universal.* Manuscript in preparation.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology, 34*(6), 1289-1309. doi: 10.1037/0012-1649.34.6.1289

Specifying the underspecified: Young children's perceptual asymmetries are both language-specific and universal

Chapter 8

Based on:
Tsuji, S., Fikkert, P., Yamane, N., & Mazuka, R. (2014). *Specifying the underspecified: Young children's perceptual asymmetries are both language-specific and universal.* Manuscript in preparation.

Abstract

Although young children's discrimination of native speech sound contrasts in lexical tasks is generally found to be unproblematic, a lack of sensitivity to the change from coronal to labial consonants has consistently been reported in Dutch children. In order to better understand the possible causes of this insensitivity, the current study assessed to what extent it is specific to the coronal-to-labial change (or would also extend to the coronal-to-dorsal change), and to children exposed to Dutch (or would also show in learners of Japanese). In a word learning task, Dutch and Japanese 18-month-old children's sensitivity to labial and dorsal mispronunciations of newly learned coronal-initial word was measured. Children from both language backgrounds were highly sensitive to dorsal mispronunciations, suggesting that the lack of sensitivity is specific to the coronal-to-labial change. Furthermore, Japanese, but not Dutch children were sensitive to the coronal-to-labial change, demonstrating that the lack of sensitivity is language-specific. Overall, the results are most consistent with an early, language-general bias specific to the coronal-to-labial change, which is modified by the properties of children's early, language-specific lexicon.

1 Introduction

Young children's sensitivity to phonological detail in early lexical representations has been demonstrated in a large body of studies. Children are able to detect mispronunciations (MPs) of words or to differentiate two words based on differences in only one phoneme (e.g., Nazzi, 2005; Swingley & Aslin, 2000). They are able to detect such differences for a variety of contrasts (e.g., White & Morgan, 2008), for both vowels and consonants (e.g., Mani & Plunkett, 2010), and at different positions within a word (e.g., Swingley, 2009).

Studies that take into account the direction in which a phoneme change occurs, however, have reported an asymmetrical sensitivity to phonological detail in lexical tasks: Dutch children fail to detect a coronal obstruent that is mispronounced as a labial, a plosive that is mispronounced as a fricative, or a voiced plosive that is mispronounced as a voiceless plosive, although they are able to detect a change in the opposite direction (Altvater-Mackensen & Fikkert, 2010; Altvater-Mackensen, van der Feest, & Fikkert, 2013; van der Feest & Fikkert, 2006). Exploring the possible cause of these asymmetries will contribute to a better understanding of children's early phonological representations. Using a cross-linguistic design and focusing on place of articulation, the current study addresses the question whether the asymmetries in lexical tasks are based on a language-general perceptual bias or on language-specific phonological representations.

1.1 Phonological detail in early lexical representations

Numerous studies have demonstrated that children have detailed phonological representations of familiar or newly learned words. One widely used paradigm to assess children's sensitivity to such detail is the preferential looking procedure (Swingley & Aslin, 2000). In this paradigm, two pictures of objects are presented side by side on a screen. In the correct pronunciation (CP) condition, one of them is named. If children recognize the object upon naming, they are expected to look to the picture of the object. Provided children are sensitive to phonological detail, their recognition should be hindered when they are presented with a mispronounced version of the object name, leading to a lower amount of

object looks compared to the CP condition. For instance, Swingley and Aslin (2000) used this procedure to demonstrate that children better recognized a target picture when presented with a CP like "baby" compared to a one-phoneme MP like "vaby". Using this or related paradigms, it has been demonstrated that children as young as 11 months of age are sensitive to phonological detail in consonants (e.g., Bailey & Plunkett, 2002; Swingley, 2005). They can access such detail not only at word onset, but also at word-medial or coda position (Nazzi & Bertoncini, 2009; Nazzi, 2005; Swingley, 2003; Swingley, 2009). Whether children are equally sensitive to voice, manner and place of articulation MPs (cf. Mani & Plunkett, 2010; White & Morgan, 2008), whether their sensitivity to MPs of one or more phonological features is graded (cf. Bailey & Plunkett, 2002; White & Morgan, 2008), and whether their sensitivity to phonological detail in vowels is comparable to that in consonants (cf. Mani & Plunkett, 2010; Nazzi, 2005) is a matter of debate and reviewed in detail elsewhere (e.g. Altvater-Mackensen et al., 2013). In any event, the above review documents children's robust sensitivity to consonantal detail in lexical tasks.

1.2 Asymmetries in early lexical representations

With these results in mind, we will now review the aforementioned asymmetries in Dutch children's sensitivity to MPs of place of articulation. It has repeatedly been documented that they were able to detect a change from a labial (e.g., /p/) to a coronal (e.g., /t/) consonant, but not vice versa, in lexical tasks. Assessed in a preferential looking paradigm, 20- and 24- month-old children were sensitive to the change from labial-initial /pʊs/ (*cat*) to its coronal MP [tʊs], but not to the change from coronal-initial /tɑnt/ (*tooth*) to its labial MP [pɑnt] (van der Feest & Fikkert, 2006). In other words, their looking times to the picture of a cat were shorter when they heard [tʊs] than [pʊs], but they did not differ in their looking times to the picture of a tooth whether they heard [pɑnt] or [tɑnt]. This pattern was replicated with 18-month-old children, who were able to detect the change from labial-initial /vɪs/ (*fish*) to its coronal MP [zɪs], but not the change from coronal-initial /ze:p/ (*soap*) to its labial MP [ve:p] (Altvater-Mackensen et al., 2013). Further evidence comes from studies in which children learn new word-object associations and are

tested on their ability to differentiate them based on a difference in the initial consonant. Stager and Werker (1997) tested English-learning 14-month-old children's bidirectional sensitivity to the contrast between labial-initial [bɪ] and coronal-initial [dɪ], but did not statistically analyze directional effects. A recent re-analysis reported an asymmetry in the expected direction (Fennel, van der Feest, & Spring, 2011). This re-analysis was additionally backed by Fikkert's (2010) modified replication of the original study with Dutch 17-month-old children. After being habituated to an object labeled [bɪn] or [dɪn], they were presented with the same object paired with the respectively different label. Children habituated to [bɪn] detected the change to [dɪn], but children habituated to [dɪn] did not show evidence of detecting the change to [bɪn].

In summary, a labial-coronal perceptual asymmetry has been documented in different tasks, with different stimuli, and in children up to an age of 24 months. This robust replication of the asymmetry across paradigms, across stimuli, and up to a relatively high age is difficult to reconcile with an explanation purely involving age or task demand (for a similar discussion, cf. Werker, Fennell, Corcoran, & Stager, 2002; Yoshida, Fennell, Swingley, & Werker, 2009). Children's ability to access the necessary phonological detail in one direction of change also speaks to their general ability to perform the tasks at hand. Therefore, these results suggest a genuine lack of sensitivity to the coronal-to-labial change in lexical tasks for Dutch- and English-learning children.

1.3 Possible causes of the labial-coronal asymmetry

The most prominent account of the labial-coronal asymmetry suggests that it is tied to the special status of coronals in the phonologies of the world (Paradis & Prunet, 1991). The Featurally Underspecified Lexicon (FUL: Lahiri & Reetz, 2010) accounts for the asymmetry by assuming sparse and abstract lexical representations in which not all phonological features are specified. Based on universal properties of phonological systems, coronal place of articulation is assumed to be the default, unmarked place and therefore to be underspecified in the mental lexicon. This assumption predicts the documented perceptual asymmetries such that labial MPs of coronal phonemes would not produce a mismatch (e.g., [pɑnt] is accepted as an instance of /tɑnt/, and

there is no difference in looking times), but coronal MPs of labials would (e.g., [tus] is not accepted for /pus/, and thus leads to shorter looking times). Under the assumption that coronals universally are the default place of articulation, the underspecification account would further predict the same asymmetry for the coronal-dorsal contrast, and for children across language backgrounds (cf. prediction (a) in Table 1). Adding a developmental perspective, Fikkert (2010) suggests that children start out without fully detailed lexical representations and gradually add detail over development.

This account's prediction for the coronal-to-labial change in Dutch conforms to findings of previous studies, but the extent to which the perceptual insensitivities generalize to the coronal-to-dorsal change and to Japanese children remains to be tested.

Table 1. Predictions on the detection of labial and dorsal mispronunciations of coronal-initial words made by the four different accounts for Dutch and Japanese. Last line provides the results of Experiment 1 and 2. "No" refers to a lack of sensitivity to the contrast, "yes" refers to contrast detection, and "?" refers to cases where no predictions can be made.

Predictions	Dutch		Japanese	
	Coronal to labial	Coronal to dorsal	Coronal to labial	Coronal to dorsal
(a) Underspecification	no	no	no	no
(b) Early bias	no	?	no	?
(c) Frequency	no	no	no	yes
(d) Production	no	yes	yes	no
Results	no	yes	yes	yes

Recent findings in young infants suggest that, alternatively to lexical underspecification, early, language-general perceptual biases could explain the above-described asymmetries. In a discrimination task, Dutch 6-month-old infants were able to detect a change from labial-initial [pa:n] to coronal-initial [ta:n], but not from coronal-initial [ta:n] to labial-initial [pa:n] (Dijkstra & Fikkert, 2011). Moreover, another study suggests that this asymmetry is present not only in Dutch, but also in Japanese infants. Four- to six-month-old infants from both language backgrounds were able to detect a change from labial [ɔmpa] to coronal [ɔnta], but not from coronal [ɔnta] to labial [ɔmpa] (Tsuji, Mazuka, Cristia, & Fikkert, 2013).

The finding of the same discrimination performance in infants learning two unrelated languages and at an age where they have not yet acquired language-specific consonant representations (cf. Kuhl, 2004) suggests a language-general perceptual bias. If the above-reviewed lack of sensitivity to the coronal-to-labial change documented in older Dutch children exclusively reflected this early bias, we would expect Japanese children to continue showing the same insensitivity later on (cf. prediction (b) in Table 1; note that no data on infants' perception of the coronal-to-dorsal change exist). Children's performance in lexical tasks does not necessarily reflect such language-general biases, however: Infants' speech sound discrimination abilities become attuned to their native language during the first year of life (cf. Kuhl, 2004). It is further assumed that lexical tasks do not only require phonetic learning, but also phonological learning, including children's knowledge of the functional value of phonemes (cf. Yoshida et al., 2009). Alternatively or in addition to early biases, children's performance in lexical tasks can therefore be assumed to reflect their phonetic and/or phonological knowledge.

Table 2. Frequency of plosives by place of articulation. Dutch counts derived from van der Weijer Corpus (van de Weijer, 1998), Japanese counts from R-JMICC (Mazuka et al., 2006). AW: all words; CW: content words; all: phonemes in all positions of the word; ini: phonemes in word-initial position.

			Labial (%)	Coronal (%)	Dorsal (%)
Dutch	Token	AW, all	20.6	**52.4**	27.0
		AW, ini	27.6	**43.5**	28.9
		CW, all	18.1	**48.0**	33.8
		CW, ini	28.0	31.7	**40.3**
	Type	CW, ini	**47.9**	26.5	25.6
Japanese	Token	AW, all	13.8	42.7	**43.5**
		AW, ini	11.0	**51.4**	37.6
		CW, all	21.4	30.2	**48.3**
		CW, ini	24.6	29.9	**45.5**
	Type	CW, ini	30.4	24.6	**45.0**

A third account derives its predictions from the high frequency of coronals in Dutch and related languages. There is evidence that discrimination in the direction from more frequent to less frequent phonemes is harder than vice versa for infants by the age of 12 months

(Pons, Albareda-Castellot, & Sebastián-Gallés, 2012). As a potential reason, it has been suggested that more frequent native phonemes are treated as referents, making discrimination from referents towards less frequent phonemes more difficult than the opposite. This view would predict discrimination from coronal to labial or dorsal place to be more difficult than the opposite in Dutch, because coronal plosives are more frequent than labial or dorsal plosives in Dutch infant-directed speech under most counts (cf. Table 2; counts derived from van de Weijer corpus, van de Weijer, 1998). The prediction for the coronal-to-labial change conforms to previous findings, but children's sensitivity to the coronal-to-dorsal change has yet to be tested. Interestingly, the predictions for Japanese differ: In Japanese infant-directed speech, coronal plosives are more frequent than labial, but less frequent than dorsal plosives under most counts (cf. Table 2; counts derived from R-JMICC, Mazuka, Igarashi, & Nishikawa, 2006). Consequently, discrimination of the coronal-to-labial change is predicted to be difficult, but not discrimination of the coronal-to-dorsal change.

Note, however, that predictions based on frequency have to be treated with caution, as, in addition to sparse evidence in the language acquisition literature, they are not specified in at least three respects: First, it is an open question how large frequency differences would need to be in order to cause a perceptual asymmetry. The above predictions might only hold for cases with relatively large frequency differences. Second, frequency differences might only influence perception in cases where infants do not succeed in (bidirectional) discrimination early on (cf. Pons et al., 2012). Assuming infants are able to discriminate the coronal-to-dorsal change early on, differences in frequency might not change existing patterns. Third, it is unclear what kind of frequency count – overall frequency or sub-sets containing word-initial phonemes or content words – would matter. To obtain maximally inclusive predictions despite this uncertainty, we calculated token frequencies for all four alternatives and determined the dominant pattern in the respective languages (cf. prediction (c) in Table 1). This decision notwithstanding, the frequencies in word-initial positions of content words, which are especially salient to young children, could be more relevant than our 'majority count' (Boysson-Bardies & Vihman, 1991; Shi & Werker, 2001;

Vihman, Kay, Boysson-Bardies, Durand, & Sundberg, 1994). Furthermore, for children in the middle of vocabulary acquisition, content word types in their input, as well as their own receptive and productive inventories, might be important factors contributing to the structure of their phonological representations.

Relating to the above, a further factor that markedly differs between young Dutch and Japanese language learners are their own productions, leading to a fourth possible prediction: based on the fact that a facilitating effect of phonological overlap in the first consonant on word recognition in 18-month-olds has been demonstrated (Mani & Plunkett, 2010b), it is conceivable that words starting with the same phonemes as words in children's early inventories might be accepted more readily than words that do not occur very frequently. Indeed, it is known that Dutch children predominantly produce labial-initial words early on (Fikkert & Levelt, 2008), but Japanese children's early productions contain a high number of dorsals (Boysson-Bardies & Vihman, 1991). Converging evidence comes from norming data of vocabulary questionnaires in both languages (for Dutch: Lexilijst, Schlichting & Spelberg, 2002; for Japanese: Japanese MacArthur CDI, Ogura & Watamaki, 2004). In plosive-initial words produced by at least 20% of 17- or 19- month-olds Dutch children, labial-initial words were most frequent (17 months: 15; 19 months: 32), followed by coronal-initial (17 months: 7; 19 months: 19), and dorsal-initial (17 months: 4; 19 months: 13) words. In contrast, the count for 18-month-old Japanese children revealed that dorsal-initial words were most frequent (11), followed closely by labial-initial (10) and lastly coronal-initial words (6). Research on speech errors also shows comparable differences: Dutch children are reported to predominantly make 'fronting' errors, which often means replacing a coronal or dorsal consonant with a labial (Fikkert & Levelt, 2008). In contrast, Japanese children have been found to make 'backing' errors, replacing coronal with more back consonants (Edwards & Beckman, 2008). Thus, overall Dutch children know more words starting with labial-initial plosives, while Japanese children know more words starting with dorsal-initial plosives. While this pattern does not conform to the token frequencies reported above, type frequencies of content words (which might better reflect the words

children know and produce) show that word-initial plosives are most frequently labials in Dutch, but dorsals in Japanese (cf. Table 2).

Based on this consistent pattern, the fourth account would predict that labial MPs are accepted as instances of coronal-initial words by Dutch, but not by Japanese children. The reverse would hold for dorsal MPs, which should be accepted as instances of coronal-initial words by Japanese, but not by Dutch children (cf. prediction (d) in Table 1).

1.4 The current study

Using a word-learning task, the current study investigated Dutch and Japanese children's sensitivity to phonological detail in coronal-to-labial and coronal-to-dorsal changes. Children first learned two novel coronal-initial word-object associations, and were subsequently tested on their sensitivity to labial and dorsal MPs of the target names in a preferential looking paradigm. Experiment 1 assessed Dutch children, and Experiment 2 assessed Japanese children. Children's early vocabularies were measured by asking Dutch caregivers to fill in the Dutch Communicative Development Inventory N-CDI (short version, Zink & Lejaegere, 2003), and Japanese caregivers to fill in the Japanese MacArthur Communicative Development Inventory (Ogura & Watamaki, 2004).

2 Experiment 1: Dutch children

2.1 Participants

Thirty-one monolingual Dutch children were included in the final sample (*mean age* = 18.77 months, *range* = 18.31-19.13 months, 15 female). They were recruited and tested in the Netherlands. Caregivers signed an informed consent, and received a picture book or a small monetary compensation for their participation. Twenty-three additional children were tested but not included in the analysis due to fussiness and not completing the experiment (7), technical problems (8), or due to the exclusion criteria explained below (8).

2.2 Stimuli

Target stimuli consisted of two word-object pairings and their respective MPs. Two stuffed animals were used as target stimulus objects (cf. Fig. 1). They had distinct colors and shapes to make them easily discriminable, and they did not resemble any existing objects.
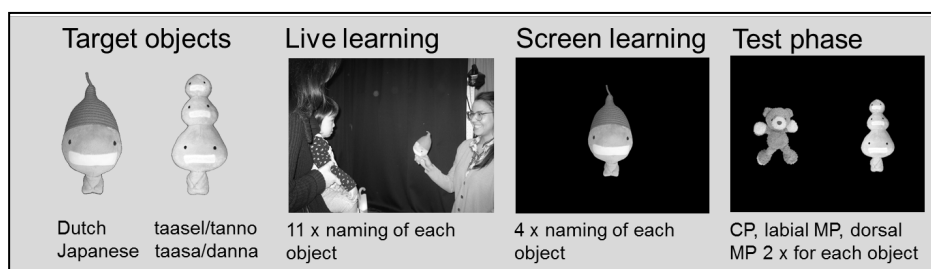

Figure 1. Target objects and procedure for Experiments 1 and 2. Although not visible in this figure, the left target object was pink, and the right target object blue.

The names associated with the two stuffed animals were coronal-initial [taːsəl] and [tɑnoː]. Low unrounded vowels were chosen to follow the critical word-initial coronal consonant, and the place of articulation of the two word-medial consonants was matched. Labial MPs were [paːsəl] and [pɑnoː], and dorsal MPs were [kaːsel] and [kɑnoː]. Stimuli were recorded by a native female speaker of Dutch in a child-directed register, and all tokens were embedded in carrier phrases.

The experiment consisted of a learning and a test phase, with the learning phase being subdivided into a live and screen learning phase (cf. Figure 1). Live learning provides a more naturalistic situation that possibly has an advantage over screen learning (DeLoache et al., 2010), and a subsequent screen learning phase made children familiar with seeing the target objects on screen. Whether the blue or the pink object was named [taːsel] or [tɑnnoː], and which object was presented first (pink or blue) was counterbalanced between participants.

During live learning, the actual stuffed animals served as visual stimuli. The experimenter, a female native speaker of Dutch, named each object 11 times in scripted carrier phrases such as "Do you want to play with the [target]?". Subsequently, children were exposed to three screen

learning trials per object. In each trial, the photograph of one of the target objects was presented centrally against black background for 4 s while slightly increasing and decreasing in size. Each object was named twice in the first trial and once in each of the remaining two, resulting in four naming instances during 12 s of exposure per object. A different carrier phrase was used in each trial, and a different auditory token was used for each naming instance. The auditory tokens were cross-spliced into the carrier phrases.

Table 3. Acoustic measurements of experimental stimuli. Numbers in 'Learning' rows are the means over all four learning stimuli, numbers in all other rows are means over the two respective test stimuli. Length is measured over the whole token, and pitch refers to the mean pitch over vowels.

			Length (ms)	Pitch (Hz)
Dutch				
	taasel	Screen Learning	729	306
		CP	724	335
		Labial MP	728	343
		Dorsal MP	712	350
	tanno	Screen Learning	714	327
		CP	707	324
		Labial MP	691	329
		Dorsal MP	674	321
Japanese				
	taasa	Screen Learning	298	347
		CP	309	388
		Labial MP	317	368
		Dorsal MP	297	360
	danna	Screen Learning	315	342
		CP	349	349
		Labial MP	386	353
		Dorsal MP	343	356

In test trials, the photograph of one of the objects was presented side by side with the photograph of an unrelated stuffed animal against a black background for 6.5 s. The pair slowly moved up and down while one of the objects was named. The onset of the initial consonant of the target name was always at 3.7 s. As it was impossible to cross-splice the same instance of the definite article [də] (which always preceded the target name) onto the labial-, coronal-, or dorsal-initial target tokens due to co-articulation, target names were cross-spliced into the carrier phrases

together with their article (e.g., "Can you see [the target]?"). Different auditory token were used in each trial, with auditory characteristics matched between CPs and MPs (cf. Table 3). Each of the three pronunciation conditions occurred twice for each of the two target names, resulting in a total of 12 test trials. The pink object was always paired with an elephant, and the blue object with a bear, with target side counterbalanced. In addition, there were 12 filler trials. The elephant and the bear were named twice each. The remaining eight filler trials consisted of two additional pairs of animals (cat-giraffe, dog-pig; each named twice with side of presentation counterbalanced). Four pseudo-random trial orders were created and counterbalanced between children.

2.3 Procedure

Children were tested in a sound-attenuated room. In the live learning phase, they sat on their caregivers' lap. The experimenter removed one of the target objects from behind a curtain, showing and naming it according to the script. The child was allowed to touch and hold the object. The caregiver was instructed to remain silent, but was encouraged to smile and look at child and object. In case the child was afraid of the object, the caregiver was also encouraged to hold it for a little while before the experimenter initiated the learning phase. Immediately following the live learning phase, the child was seated on the caregiver's lap in front of a Tobii T60 eyetracker. The experimenter was invisible to the child during the remainder of the experiment, but monitored and initiated trials from a separate computer screen in the same room. Caregiver and experimenter listened to masking music throughout the on-screen experimental phase. After calibration, the child saw the six screen learning trials and the 24 experimental trials. Preceding each trial, a spinning smiley appeared in the middle of the screen, and the trial was initiated once the child looked at it. A short movie of a duck was presented both after screen learning and after the first block.

2.4 Data preprocessing

Exclusion criteria on trial and participant level were applied. In order to exclude trials in which children only spuriously looked at the

screen, trials in which they looked less than 500 ms of the 2000 ms following target word onset (367-2367 ms) to anywhere on screen were excluded. This excluded 18 % of test trials. To exclude children that were not attentive during screen learning, we excluded five children that had accumulated less than 4 s (of 12 s) of looking time to either of the two objects. Additionally, we excluded three children who did not contribute at least one valid trial per condition during the test phase.

2.5 Data analysis

Past studies using the preferential looking paradigm have mostly analyzed outcome data by averaging over a given time-window. In the present study, however, the trajectory of the naming effect in the CP condition differed between Dutch children in Experiment 1 (peak around 850 ms after target word onset) and Japanese children in Experiment 2 (peak around 1200 ms). As our research question pertained the comparison of the MP conditions to the CP condition, it was critical to do this in a comparable way for the two language groups. Since choosing separate windows would impair comparability, we chose a common time-window appropriate for both language groups. We pooled the data from the two experiments in the 367-2367 ms following word onset and averaged the looks to target over time-slices of 100 ms. We then determined the time-slice with maximum target looks in the CP condition, which can be considered the baseline condition. The time-window of analysis was then defined as the 1000 ms around the average of the peak time-slice. This procedure resulted in an analysis window of 450 to 1450 ms after word onset that was applied to both data sets. We analyzed the data of this time-window with growth curve analysis (GCA, Mirman, Dixon, & Magnuson, 2008). GCA accounts for the dynamic nature of gaze data by not only assessing overall differences in looking times, but additionally differences in the shape and latency of the gaze curve over the time-window. GCA was conducted with the lme4 package (Bates, Maechler, & Bolker, 2012) in R (R Core Team). The time course of the naming effect was captured with first (linear) and second order (quadratic) orthogonal polynomials, and with fixed effects of condition (CP, labial MP, dorsal MP) on all time terms. To account for individual differences, we allowed random effects of participant and participant-by-condition on all time

terms. The looks to target (yes or no) were the dependent variable. The CP condition was defined as the baseline against which the other conditions were compared.

One caveat of the above analysis is that our choice of matching time-windows across experiments might have reduced sensitivity for one or both language groups. More importantly, there is no objectively established criterion for the choice of an appropriate time-window, and other authors might have made different choices. We therefore conducted a complementary adopting a non-parametric statistical test (NPST), which determines the time-stretches in which conditions differ from each other in a bottom-up way. This procedure was initially introduced for the analysis of event-related potential (ERP) data (Maris & Oostenveld, 2007) and has been applied to child eye-tracking data recently (Von Holzen & Mani, 2012). T-tests on the difference between two conditions were calculated for each time-point in the 2 seconds following word onset (367-2367 ms). Based on these t-tests, time-adjacent clusters of time-points whose t-values exceeded a certain threshold ($p < .05$, two-tailed) and have the same sign were identified. To account for the multiple comparison problem, Monte-Carlo resampling was then applied and the trials of the original data-set were randomly re-assigned to the three conditions 1000 times. For each of these resampled data-sets, the largest cluster as determined by the sum of its t-values was identified. Finally, it was tested whether the largest cluster of the original sample was significantly different from chance by comparing it to the largest clusters of the 1000 resampled data-sets. Using this procedure, it is possible to identify time intervals in which two conditions differ from each other while controlling for multiple comparisons through resampling (see Maris & Oostenveldt, 2007, for details).

2.6 Results and Discussion

For the analysis of the N-CDI, we counted the number of plosive-initial words for each place of articulation in each child's receptive and productive vocabularies, and calculated the mean number of words per place (Table 4).

Table 4. Mean number of plosive-initial words by place of articulation in children's early inventories according to parental report (Dutch: Zink & Lejaegere, 2003; Japanese: Ogura & Watamaki, 2004). Bold script indicates highest number in a row. Differences in absolute frequencies are due to the fact that a full version of the questionnaire was used in Japanese, but not in Dutch.

	Words	**Labial**	**Coronal**	**Dorsal**
Dutch	understood	5.15	4.15	**5.89**
	produced	**3.44**	2.00	3.30
Japanese	understood	7.72	13.45	**20.79**
	produced	5.00	3.83	**7.90**
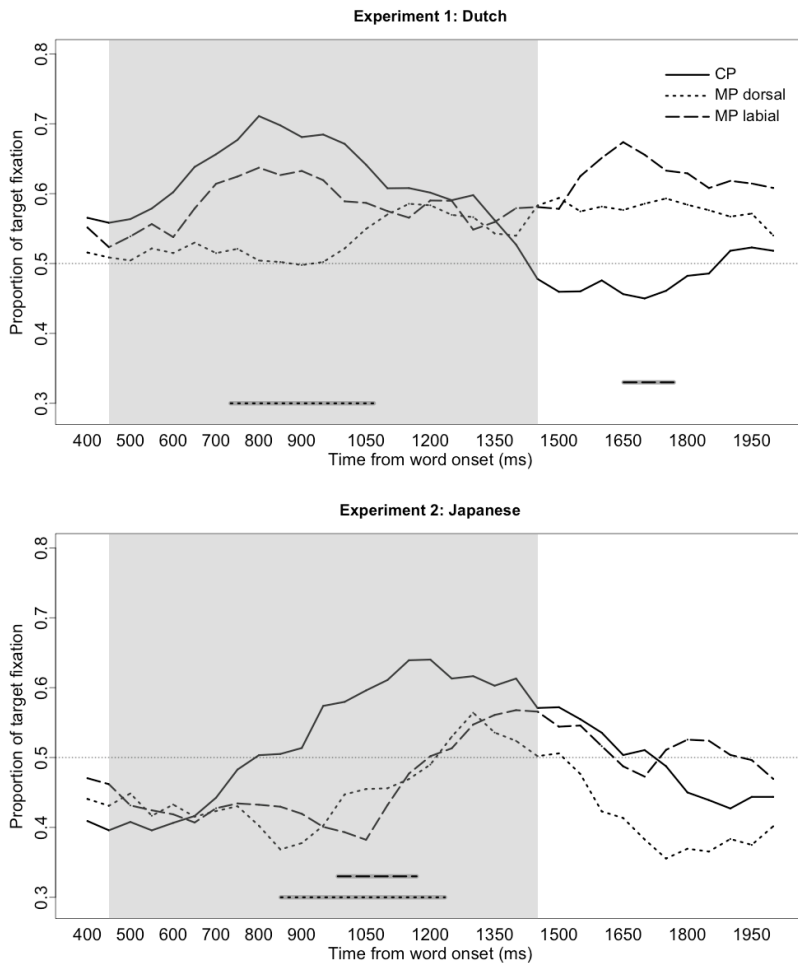
Figure 2. Time-course of gaze after word onset for correct pronunciation (CP), dorsal mispronunciation (MP dorsal), and labial mispronunciation (MP labial), separately for Dutch (above) and Japanese (below) children. The grey-shadowed area indicates the time-window for the growth-curve analysis. The dotted lines at the bottom of the graphs signify the time intervals that resulted in significant differences in the non-parametric test.

The plosive-initial words children understood were mostly dorsal, followed by labial and coronal words. Words produced were most frequently labial, followed by dorsal and coronal, reflecting what has been found in studies on early production (cf. Fikkert & Levelt, 2008).

The top part of Figure 2 shows the time course of looks to target for the CP and MP conditions for Dutch children. The averaged percentage of looks to target during the time-window chosen for GCA (grey-shadowed area in Figure 2) was 62.4% in the CP condition, 58.4% in the labial MP condition, and 53.2% in the dorsal MP condition. The labial MP condition neither had an effect on the intercept (*estimate* = -0.34, *z* = -0.68. *p* = .500) nor on the quadratic time term (*estimate* = 2.13, *z* = 1.64, *p* = .102), indicating no difference to the CP condition. By contrast, the dorsal MP condition had a significant effect on the intercept (*estimate* = -1.16, *z* = -2.27, *p* = .023) and on the quadratic polynomial (*estimate* = 3.06, *z* = 2.31, *p* = .021). The main effect reflects that children looked less to the target object after hearing the dorsal MP compared to the CP. The effect on the quadratic term suggests that the looks to the target object followed a shallower curve for the dorsal MP compared to CP, possibly reflecting slower word recognition.

The NPST backed the results of GCA, revealing no significant differences between children's looks to target in the labial MP condition compared to the CP condition in the GCA time-window. In a later time-window between 1651 and 1768 ms, however, their target looks in the labial MP condition (62.7%) increased, tendentially exceeding the CP condition (49.2 %) (*t* = - 24.86, p = .077). By contrast, children showed a clear MP effect for the dorsal MP condition in the GCA time-window, with significantly fewer looks to target in the MP (51.1%) compared to the CP (68.4 %) condition between 734 and 1068 ms (*t* = 67.70, *p* = .004).

The converging results on the labial MP condition are in line with previous findings: Dutch children did not look less to the target object when it was mispronounced than when it was correctly pronounced, thus accepting labial MPs as instances of previously learned coronal-initial words. The NPST did, however, reveal a response pattern not reported in previous studies: Children's gaze trajectory to the labial MPs was not identical to the trajectory towards the CPs, but rather suggested delayed "recognition": While looks to target in the CP condition peaked early and

then returned to chance, looks to target in the labial MP condition showed a tendency to increase later on. This tendency suggests that children were not absolutely insensitive to the difference between labial- and coronal-initial stimuli, even though they accepted labial MPs as instances of coronal-initial words. Previous studies might have failed to detect such differences between the two conditions, because they analyzed time-windows centered around the peak response as we did for GCA, or because they reported averaged responses over a larger time-window (which, for the current data-set in the time-window of 367-2367 ms after target word onset, would also lead to very similar percentages: 53.1 % for CP, and 53.5 % for labial M).

While the results for the coronal-to-labial change are consistent with predictions from all accounts in Table 1, children's clear sensitivity to dorsal MPs that was confirmed in both analyses is inconsistent with the FUL and the frequency account: both would have predicted Dutch children to accept dorsal MPs as instances of coronal-initial words. The results of Experiment 1 therefore exclude both as possible explanations for children's differential sensitivity towards place of articulation changes. The outcomes are, however, compatible with both the early language-general bias (which has only been attested for the coronal-to-labial change), and the production account. These accounts would, however, make diverging predictions for Japanese children: If the lack of sensitivity to the coronal-to-labial change was indeed exclusively reflecting an early bias, we would expect Japanese children's perceptual patterns to resemble those of Dutch children. An influence of early productive vocabulary would, however, predict Japanese children to be sensitive to the coronal-to-labial (but not the coronal-to-dorsal) change. To disentangle these possibilties, Experiment 2 assessed Japanese children's perceptual patterns.

3 Experiment 2: Japanese children

3.1 Participants

Twenty-nine monolingual Japanese children were included in the final analysis (mean age = 18.46 months, range = 18.02-19.04 months, 13 female). They were recruited and tested in the Tokyo area of Japan.

Caregivers signed an informed consent, and received a book voucher in return for their participation. Seven additional children were tested but not included into analysis due to fussiness and not completing the experiment (1), poor tracking of gaze (1), equipment error (1), experimenter failure (1) or due to the exclusion criteria detailed out below (3).

3.2 Stimuli

The Japanese object names were [ta:sa] and [daNna] for CPs, and [pa:sa]/[baNna] and [ka:sa]/[gaNna] for the respective labial and dorsal MPs. These non-words were matched as closely as possible to their Dutch counterparts while making them sound natural in Japanese. Auditory stimuli were recorded by a female native speaker of Japanese in child-directed register. Visual stimuli were the same as in Dutch.

3.3 Procedure

Japanese children were tested on a Tobii 60 XL, which has a larger screen then the eye-tracker used in Experiment 1. In order to make the positioning and size of stimuli identical to Experiment 1, a subset of the screen was used that was identical in size to the screen used in Experiment 1. The experimental procedure was identical to Experiment 1.

3.4 Data preprocessing

The same exclusion criteria as in Experiment 1 were applied. Exclusion of trials with less than 25% of looks to target after naming resulted in the exclusion of 14% of trials. Two participants were excluded who had accumulated less than 4 s of looking time to each of the toys in the screen learning phase, and one participant was excluded for having less than one trial per condition in the test phase.

3.5 Data Analysis

Analyses were exactly the same as in Experiment 1.

3.6 Results and Discussion

The parental vocabulary reports for Japanese are summarized in Table 4. The plosive-initial words Japanese children understood most frequently were dorsal, followed by coronal and labial words. Words produced were also most frequently dorsal, followed by labial and coronal,consistent with previous studies (cf. Boysson-Bardies & Vihman, 1991).

The bottom part of Figure 2 shows the time-course of looks to target for Japanese children. The averaged percentage of looks to target during the GCA time-window was 52.9% in the CP condition, 45.4% in the labial MP condition, and 45.1% in the dorsal MP condition. The GCA showed no significant effect of the labial MP on the intercept (*estimate* = -0.416, *z* = -1.12, *p* = .269), but on the quadratic time term (*estimate* = 3.00, *z* = 2.59, *p* = .010). This indicates that word recognition was faster for the CP compared to the labial MP, although there was no overall bigger naming effect. The dorsal MP condition had a significant effect on the intercept (*estimate* = -0.91, *z* = -2.41, *p* = .016) as well as on the quadratic time term (*estimate* = 2.50, *z* = 2.16, *p* = .031), indicating a smaller and slower naming effect compared to the CP.

Again, the outcomes of the NPST corroborate the findings of the growth curve analysis. Japanese children looked less to the labial MP condition (40.7 %) than to the CP condition (59.8 %) in the time-window between 984 and 1168 ms after target word onset, although this effect failed to reach significance by a small margin (*t* = 31.62, *p* = .055). The effect for the dorsal MP reached significance, with significantly fewer looks to dorsal MPs (43.0 %) than to CPs (58.0 %) in the time-window between 851 and 1234 ms (*t* = 62.20, *p* = .022).

Unlike Dutch children, Japanese children showed a MP effect for both types of MP in an early time-window. These results demonstrate that Japanese children accepted neither labial nor dorsal MPs as instances of coronal-initial words. However, it is also obvious that Japanese children were less sensitive to the coronal-to-labial than to the coronal-to-dorsal change, illustrated by a delayed, but not overall lower response according to GCA, and a shorter differential time-window which only lead to a marginally significant effect in the NPST.

Despite this weaker sensitivity for labial MPs, Japanese children's results differ clearly from Dutch children's, who at no point showed a significant decrease in target looks in the labial MP condition compared to CPs. These differences between Dutch and Japanese children's outcomes are not compatible with the notion that their responses were based exclusively on a common language-general bias. Instead, they are compatible with the predictions from the production account, which assumed that Dutch, but not Japanese children would more readily accept labial-initial words as instances of coronal-initial words.

Taking the results on the dorsal MP into account also, however, shows that the production account does not accurately predict the full pattern of results: it predicted that Japanese children would, unlike Dutch children, accept dorsal MPs as instances of coronal-initial words, but this was clearly not the case. Japanese children's sensitivity to dorsal MPs is actually comparable to Dutch children's: both language groups show significant differences between CPs and MPs in both GCA and NPST. In summary, it therefore seems that none of the predictions presented in Table 1 can fully account for the pattern of results. As we will argue in the General Discussion, the outcomes are most compatible with an interaction between early, language-general biases and language experience.

4 General Discussion

The current study assessed the extent to which young children's sensitivity to the coronal-to-labial and coronal-to-dorsal change differed depending on language background. Although there is evidence that sensitivity to the coronal-to-labial change is reduced in both Dutch and Japanese prelexical infants (Tsuji et al., 2013), there have been no cross-linguistic investigations on the further development of this sensitivity in lexical contexts. It has been shown that Dutch children between 18 and 24 months of age have are insensitive to the coronal-to-labial change in lexical tasks (Altvater-Mackensen et al., 2013; van der Feest & Fikkert, 2006), but without studies in other languages it has not been possible to determine to what extent this is a language-general perceptual pattern.

The current study therefore compared Dutch and Japanese children on their sensitivity to the coronal-to-labial change in a lexical

task. Cross-linguistic differences clearly indicate an influence of native language, with Dutch, but not Japanese children accepting labial MPs as instances of newly learned coronal-initial words. In addition, sensitivity to the coronal-to-dorsal change was tested in order to investigate to what extent previous findings would generalize to other changes involving coronals. Children from neither language backgrounds treated dorsal MPs as instances of coronal-initial words, implying that perceptual insensitivities were specific to the coronal-to-labial change. After discussing the results in light of the predictions put forward in the Introduction, we will put forward possible explanations for the results of the current study. We will ultimately conclude that an interplay of early biases with language-specific input and production capacities is most compatible with our results.

4.1 The results in light of predictions

As discussed in the Results sections, neither early perceptual biases, underspecification, frequency, nor production can fully account for the present results. If the results had shown an early bias specific to the coronal-labial contrast, Japanese children would have been expected to show insensitivity to the coronal-to-labial change similarly to Dutch children. Underspecification in turn would have predicted neither Dutch nor Japanese children to be sensitive to any change from coronal towards another place of articulation. Even if we adopt an alternative view of underspecification which allows for language-specific differences, the current results cannot be accounted for: according to Labrune (2012), coronals are not acting as the default place of articulation in Japanese. Under this assumption, Dutch children would still be expected to be insensitive to both changes tested in the current study, while Japanese children would be sensitive to both changes. The results of the current study are clearly not consistent with this prediction.

The predictive power of the frequency account does not fare any better. It predicted that the coronal-to-dorsal change in Japanese children was the only condition in which change sensitivity would occur. In fact, however, the coronal-to-labial change in Dutch children was the only condition in which there was no sensitivity. As pointed out earlier, the frequency account is underdetermined in several ways. Even post-hoc

adjustment along these lines does not, however, lead to predictions consistent with the findings of the current study. First, considering only those cases with relatively large frequency differences shows that coronals are on average only 1.4 times as frequent as dorsals in Dutch, and dorsals are only 1.1 point as frequent as coronals in Japanese; whereas coronals are on average around twice as frequent as labials (1.9 times in Dutch; 2.2 times in Japanese). Assuming that only the latter differences matter, this account would predict a decreased sensitivity to the coronal-to-labial change, but not the coronal-to-dorsal change, for both languages. Second, it is possible that frequency only influences speech sound contrasts for which infants have initial discrimination difficulties. However, even if we assume that frequency only affects the coronal-to-labial change (for which these difficulties are attested, cf. Tsuji et al., 2013), the frequency account fails: since coronals are more frequent than labials in both languages, it would have predicted a lack of sensitivity to the labial-to-coronal change. Third, even taking into account the possibility that frequencies in the word-initial position of content words are critical for children's perceptual sensitivities does not explain the results. Frequencies under this count would have predicted both Dutch and Japanese children to be sensitive to the change from less frequent coronal to more frequent dorsal place, but not from more frequent coronal to less frequent labial place. In sum, it is therefore unlikely that simple frequency of exposure can account for the present results.

Lastly, the production account is the only one of the four accounts that is compatible with the results on the coronal-to-labial change for both Dutch and Japanese children. However, it cannot account for the fact that both groups of children were equally insensitive to the coronal-to-dorsal change. In the following, we will discuss possible alternative explanations for the present results.

4.2 The perception of the coronal-to-dorsal change

In contrast to the results on the coronal-to-labial change discussed in the next section, there is no data on young infants' discrimination of the coronal-to-dorsal change, and we turned to adult discrimination data to approach an answer. These data suggest that there might be no perceptual insensitivities for the coronal-to-dorsal change in young

infants to begin with: data from both Dutch (Smits, Warner, McQueen, & Cutler, 2003) and Japanese (Saito, 1961) phoneme confusions suggest that discrimination of the coronal-to-labial change is more difficult than discrimination of the coronal-to-dorsal change[1]. Participants' confusion of a coronal with a labial plosive was consistently higher ([p] in place of [t]: 7.6% in Dutch, 7.0% in Japanese; [b] in place of [d]: 15.9% in Dutch, 2.0% in Japanese) than their confusion of a coronal with a dorsal plosive ([k] in place of [t]: 0.9% in Dutch, 1.3% in Japanese; [g] in place of [d]: 0.3% in Dutch, 1.5% in Japanese). We acknowledge that infants' early discrimination abilities cannot straightforwardly be inferred from these adult data. Both Dutch and Japanese adults' perceptual patterns were, however, comparable to those of young infants (Tsuji et al., 2013): Infants frequently confused coronals with labials, and did so more than they confused labials with coronals: ([t] in place of [p]: 1.3% in Dutch, 3.1% in Japanese; [d] in place of [b]: 1.5% in Dutch, 1.3% in Japanese). Therefore, the fact that adults have less difficulties discriminating the coronal-to-dorsal change than the coronal-to-labial change can give at least an indication that the former might not be difficult to perceive to begin with, and therefore would simply continue to be discriminated at 18 months of age. Future experiments on early discrimination in infants are, however, needed to back up this interpretation (and are currently underway in our lab). Lastly, even if young infants were insensitive to the coronal-to-dorsal change early on, it would be conceivable that this insensitivity would have disappeared by 18 months of age: early perceptual insensitivities in native contrasts have previously been reported to disappear over the course of development (though in vowels, cf. Polka & Bohn, 2011).

4.3 The perception of the coronal-to-labial change

We know from a previous study (Tsuji et al., 2013) that Dutch and Japanese infants are insensitive to the coronal-to-labial change. The

---

[1] One difficulty in comparing data across these studies is that the Dutch study presented participants with gated fragments, while the Japanese study presented whole syllables varying signal-to-noise ratios. As it was difficult to determine which condition in the Japanese study was best comparable to the Dutch data, we report in the following the mean values across different conditions for the Japanese data.

current study showed a divergence in perceptual sensitivities to this change between Dutch and Japanese children later in development.

The production account put forward in the Introduction predicted this divergence based on the fact that Dutch, but not Japanese, children predominantly produce labial-initial words early on (Fikkert & Levelt, 2008, Boysson-Bardies & Vihman, 1991), a pattern that is corroborated by norming data on early productions and also resembles infants' input types (cf. Introduction). The fact that labial-initial words are highly frequent in Dutch, but not Japanese children's early inventories could be the reason that Dutch, but not Japanese children treated labial MPs of previously learned coronal-initial words as acceptable variants. Importantly, parental reports of children's productive vocabularies in the current study reproduce this pattern: while Dutch children most frequently produced words with labial-initial plosives, Japanese children produced words with dorsal-initial plosives for (cf. Table 4).

Another language-specific factor that could possibly explain the divergence in the perception of the coronal-to-labial change in Dutch and Japanese children is difference in the distributional characteristics of coronals in Dutch versus Japanese. In Dutch, coronals undergo phonological processes like place assimilation which lead to surface variation. This context-dependent realization of syllable-final coronals as labials or dorsals could make Dutch coronals more variable compared to Japanese, and this variability could lead to fuzzier category boundaries in Dutch compared to Japanese children. In order to validate this possibility, careful analyses of distributional characteristics of phonemes in both Dutch and Japanese combined with targeted experiments are necessary. More broadly, such detailed analysis would also provide insights into the kind of phonetic characteristics that do or do not influence infants' phoneme discrimination ability.

Importantly, we assume that each of these language-specific factors (productive vocabulary and distributional characteristics) would strengthen or reinforce a previously existing insensitivity to the coronal-to-labial change in Dutch children rather than producing such an insensitivity and that it would disappear for Japanese children due to lack of the input-based reinforcement. If productive vocabulary or distributional characteristics could produce Dutch children's insensitivity

to the coronal-to-labial change, it would be difficult to account for the fact that these factors do not influence sensitivity to the coronal-to-dorsal change: Under the production account, Japanese children should have difficulties with the coronal-to-dorsal change, given the high frequency of dorsals in their early inventories. In case of distributional characteristics, Dutch children should equally have problems with the coronal-to-dorsal change, since the high variability of coronals is assumed to cause reduced sensitivity.

5 Conclusions & Outlook

Further research is needed to clarify at what point in development, and on what level of representation, Dutch and Japanese children's sensitivity to the coronal-to-labial change diverges. In case distributional properties do impact on language-specific sensitivities, this divergence might be observed in phonetic discrimination within the first year of life. If children's early word inventories are the critical difference, the divergence might only be observed once children have a small receptive and productive vocabulary, and possibly only in lexical tasks. If the latter, it could even be the case that children are able to perceive a difference in a discrimination task, but not in a lexical task. In the case of fricative-plosive asymmetries, it has been demonstrated that 14-month-old infants were able to discriminate the plosive-to-fricative change, but did not show sensitivity to the same change in a lexical context at 18 months of age (Altvater-Mackensen et al., 2013).

The findings on children's sensitivity to the coronal-to-dorsal change are not surprising given the literature on children's sensitivity to phonological detail. They are, however, highly unexpected in the context of the underspecification literature. Not only would the account have predicted insensitivities to both types of tested changes, but also have numerous adult studies documented insensitivities to both the coronal-to-labial and coronal-to dorsal changes in German adult listeners (cf. Lahiri & Reetz, 2010 for an overview). It is an open question to what extent these findings can be reconciled with the results of the current study. One potential difference lies in the methods used to measure perceptual sensitivities. Adult listeners would certainly be able to detect a change in a

simple and straightforward mispronunciaion task, which is why the above-mentioned adult studies predominantly assessed responses in cross-modal priming or ERP studies. Thus, it might have been the case that sensitivity to the coronal-to-dorsal change was also reduced, but to a lesser degree than sensitivity to the coronal-to-labial change, and that the preferential looking task was not sensitive enough to capture it. This is, however, not a very likely explanation, as no adult study reports differences in strength of asymmetries between labial-coronal and coronal-dorsal contrasts. Further studies on the presence or absence of a language-general early asymmetry in the coronal-dorsal contrast, as well as tracking the development of perceptual asymmetries, might help to shed light on this question.

In sum, the current study provided differentiated insights into children's sensitivity to phonological detail in place of articulation changes. Our data suggest that early language-general perceptual insensitivities can be maintained in case they are supported by language-specific properties (as was the case for the coronal-to-labial change in Dutch children). Possible language-specific properties identified in the current study are the distribional properties of phonemes, as well as the phonemes in children's early vocabularies.

Acknowledgments

References

Altvater-Mackensen, N., & Fikkert, P. (2010). The acquisition of the stop-fricative contrast in perception and production. *Lingua*, *120*(8), 1898–1909. doi:10.1016/j.lingua.2010.02.010

Altvater-Mackensen, N., van der Feest, S. V. H., & Fikkert, P. (2013). Asymmetries in Early Word Recognition: The Case of Stops and Fricatives. *Language Learning and Development*, 1–30. doi:10.1080/15475441.2013.808954

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, *17*(2), 1265–1282. doi:10.1016/S0885-2014(02)00116-8

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. Retrieved from http://cran.r-project.org/package=lme4

Boysson-Bardies, B. de, & Vihman, M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language,* 297-319.

DeLoache, J. S., Chiong, C., Sherman, K., Islam, N., Vanderborght, M., Troseth, G. L., … O'Doherty, K. (2010). Do babies learn from baby media? *Psychological Science*, *21*(11), 1570–4. doi:10.1177/0956797610384145

Dikstra, N., & Fikkert, P. (2011). Universal Constraints on the Discrimination of Place of Articulation? Asymmetries in the Discrimination of "paan"and "taan"by 6-month-old Dutch Infants. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the Boston University Conference on Language Development 35* (pp. 170–182).

Edwards, J., & Beckman, M. E. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language learning and development : the official journal of the Society for Language Development*, *4*(2), 122–156. doi:10.1080/15475440801922115

Fennel, C. T., van der Feest, S. V. H., & Spring, M. (2011). *Consonants asymmetries at 14 months: Implications of acoustic variation on early phonological acquisition.* Manuscript in preparation.

Fikkert, P. (2010). Developing representations and the emergence of phonology: evidence from perception and production. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology 10: Variation, phonetic detail and phonological representation (Phonology & Phonetics 4-4)* (pp. 227–258). Berlin: Mouton.

Fikkert, P., & Levelt, C. (2008). How does place fall into place? The lexicon and emergent constraints in the developing phonological grammar. In P. Avery, E. Dresher, & K. Rice (Eds.), *Contrast in phonology: Perception and Acquisition* (pp. 231–270). Mouton: Berlin.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews. Neuroscience*, *5*(11), 831–43. doi:10.1038/nrn1533

Labrune, L. (2012). *The Phonology of Japanese.* Oxford: Oxford University Press.

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

Mani, N., & Plunkett, K. (2010a). Twelve-Month-Olds Know Their Cups From Their Keps and Tups. *Infancy*, *15*(5), 445–470. doi:10.1111/j.1532-7078.2009.00027.x

Mani, N., & Plunkett, K. (2010b). In the infant's mind's ear: evidence for implicit naming in 18-month-olds. *Psychological science*, *21*(7), 908–13. doi:10.1177/0956797610373371

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024

Mazuka, R., Igarashi, Y., & Nishikawa, K. (2006). Input for Learning Japanese: RIKEN Japanese Mother-Infant Conversation Corpus. *The Institute of Electronics, Information and Communication Engineers Technical Report*, *16*, 11–15.

Nazzi, T., & Bertoncini, J. (2009). Phonetic Specificity in Early Lexical Acquisition: New Evidence from Consonants in Coda Positions. *Language and Speech*, *52*(4), 463–480. doi:10.1177/0023830909336584

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: differences between consonants and vowels. *Cognition*, *98*(1), 13–30. doi:10.1016/j.cognition.2004.10.005

Ogura, T., & Watamaki, T. . (2004). *Technical Manual of the Japanese MacArthur Communicative Development Inventory: Words and Gesture,*. Kyoto International Social Welfare Exchange Center.

Paradis, C., & Prunet, J.-F. (Eds.) . (1991). *The special status of coronals: internal and external evidence (Phonetics and Phonology 2)*. San Diego: Academic Press.

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

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), 965–76. doi:10.1111/j.1467-8624.2012.01740.x

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Saito, S. (1961). Confusion Matrices of Japanese Speech Sounds. *Electrical Communication Laboratory Technical Journal*, *6*.

Schlichting, J., & Spelberg, H. L. (2002). *Lexilijst Nederlands, een instrument om de taalontwikkeling te onderzoeken bij Nederlandstalige kinderen van 15-27 maanden in het kader van vroegonderkenning: Manual*. Lisse: Swets Test Publishers.

Shi, R., & Werker, J. (2001). Six-month-old infants' preference for lexical words. *Psychological Science, 12*(1), 70-75. doi: 10.1111/1467-9280.00312

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

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Science, 388*(6640), 381–382.

Swingley, D. (2003). Phonetic Detail in the Developing Lexicon. *Language and Speech*, *46*(2-3), 265–294. doi:10.1177/00238309030460021001

Swingley, Daniel. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental science*, *8*(5), 432–43. doi:10.1111/j.1467-7687.2005.00432.x

Swingley, Daniel. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, *60*(2), 252–269. doi:10.1016/j.jml.2008.11.003

Swingley, Daniel, & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166. doi:10.1016/S0010-0277(00)00081-0

Tsuji, S., Mazuka, R., Cristia, A., & Fikkert, P. (2013). *A labial is a good enough coronal, but not vice versa, already at 4 months.* Manuscript submitted for publication.

Van de Wijer, J. (1998). *Language input for word discovery*. MPI series in psycholinguistics 9.

Van der Feest, S. V. H., & Fikkert, P. (2006). *Learning to represent phonological contrasts.* Manuscript submitted for publication.

Vihman, M., Kay, E., Boysson-Bardies, B., Durand, C., & Sundberg, U. (1994). External sources of individual differences? A cross-linguistic analysis of the phonetics of mothers' speech to 1-yr-old children. *Developmental Psychology*, *30*(5), 651–662. doi: 10.1037/0012-1649.30.5.651

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, *113*(4), 569–586. doi:10.1016/j.jecp.2012.08.001

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' Ability to Learn Phonetically Similar Words: Effects of Age and Vocabulary Size. *Infancy*, *3*(1), 1–30. doi:10.1207/S15327078IN0301_1

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*(1), 114–132. doi:10.1016/j.jml.2008.03.001

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental science*, *12*(3), 412–8. doi:10.1111/j.1467-7687.2008.00789.x

Zink, I., & Lejaegere, M. (2003). *N-CDI's: korte vormen, Aanpassing en hernormering van de MacArthur Short Form Vocabulary Checklist van Fenson et al.* Leuve/Leusden: Acco.

General Discussion

Chapter 9

This chapter summarizes and discusses the main findings of this dissertation. After a short chronological summary, the findings are discussed in thematic blocks while integrating common themes across chapters, putting these into a broader context, and pointing out remaining questions. Sections 2 and 3 address methodological issues, while the remaining sections discuss theoretical questions.

1 Summary

In the first part of the dissertation (Chapters 2 to 4), we set up, reviewed, and performed a meta-analysis of a database of infant vowel discrimination studies (InPhonDb). In Chapter 2, we demonstrated the usefulness of such a database on the basis of two examples. First, a systematic effect of experimental method on effect sizes was found, implying that the way we assess infant speech sound discrimination systematically affects the strength of outcomes. Second, phonological distance, but not spectral distance, was found to predict effect sizes, suggesting that **a larger (phonological) distance between contrasts leads to better discrimination**. The qualitative review in Chapter 3 provided an overview of the state of the art in infant vowel discrimination studies. **This review showed that discrimination of vowel contrasts is robust, but not equally strong for all vowel contrasts, that there is clear evidence for experience-dependent changes, and that vowel perception in infants at risk for language impairments and bilingual infants differs from that of typically developing monolingual infants.** Finally, Chapter 4 – a meta-analysis of extant studies on perceptual attunement with respect to vowels – showed that **native and non-native vowel discrimination start to diverge at six months.** This analysis also demonstrated that the literature provides a markedly smaller amount of data points on the development of non-native discrimination than on the

development of native discrimination (22 versus 75 data points), which is one possible reason that statistical evidence for an improvement in the discrimination of native contrasts, but not for a decline in the discrimination of non-native contrasts could be found.

After evidence for an influence of the *presence* versus *absence* of exposure to a vowel contrast had been provided in Chapter 4, we assessed the influence of *frequent* versus *infrequent* exposure on developing vowel perception in Chapter 5. Since perceptual attunement is conceived as operating on accumulated evidence rather than based on the mere presence or absence of a contrast, we hypothesized that we would observe a stronger response towards a frequent compared to an infrequent native vowel contrast. We compared five-to-eight-month-old Dutch infants' responses by means of a behavioral paradigm as well as with near-infrared spectroscopy (NIRS). **Contrary to our expectations, no difference between the discrimination of the two contrasts was found in the behavioral experiment, and only weak evidence for frequency-related differences in processing were found in the NIRS experiment**.

Having assessed the impact of the input on developing speech sound perception in two ways (Chapters 4 and 5), Chapter 6 focused on the input itself. We compared phoneme frequencies in Japanese infant-directed speech (IDS) and adult-directed speech (ADS). In addition to providing the first overview of phoneme frequencies in Japanese IDS, **our results suggest that Japanese IDS contains a higher frequency of phonemes that are produced early across languages, but also of phonemes that are especially prominent in Japanese, compared to ADS.** Data from more languages are needed, however, to evaluate in how far these patterns can be generalized to IDS in other languages.

Chapters 7 and 8 investigated early language-general biases and their role in later phonological development. **The discrimination experiment in Chapter 7 established that both Japanese and Dutch four-to-six-month-old infants showed an asymmetry in the discrimination of the labial-coronal contrast, such that they were able to discriminate two tokens in the direction from labial to coronal, but not the other way around.** The eye-tracking study in Chapter 8 followed up on this perceptual asymmetry by assessing the

perceptual sensitivities of 18-month-old children of both language backgrounds in a word-learning task. If the early perceptual bias continues to influence perception, both Dutch and Japanese children were expected to be insensitive to the coronal-to-labial change. In addition, if this insensitivity was related to the nature of coronals in general, children were expected to show a comparable insensitivity towards the coronal-to-dorsal change. The results showed that **Dutch children continued to show an insensitivity to the coronal-to-labial change, unlike Japanese children, suggesting that language-specific characteristics had overriden the early bias in Japanese, but not Dutch children**. At the same time, though, Japanese children's sensitivity to the coronal-to-labial change failed to reach significance by a small margin, which suggests that the bias had not been completely overruled by language experience. **Importantly, both Dutch and Japanese children were highly sensitive to the coronal-to-dorsal change, implying that the attested lack of perceptual sensitivity was specific to the coronal-to-labial change rather than applicable to changes involving coronal place of articulation in general**. The overall pattern of results was most compatible with an interaction of early biases with children's language-specific vocabulary inventory.

2 Meta-analyzing infant vowel discrimination: gains and drawbacks

Chapters 2 to 4 introduced and put to use InPhonDB, a database on infant vowel discrimination studies. Based on the studies reported in these chapters, can we conclude that such a database furthers our insight beyond the studies it contains? Indeed, the reported findings were uniquely made possible by analyzing over a range of studies. Perhaps the most obvious case in which such an approach can provide new answers was the question of whether experimental method affects effect sizes (Chapter 2), since the need to compare methods is intrinsic to this question. New insight was also gained by the fact that the accumulation of data points lead to a rather continuous distribution in the numeric values of predictor variables. This is rarely the case in separate studies, in which predictors such as phonological distance or infant age are mostly assessed

categorically due to practical constraints (e.g., two contrasts, two age groups). Due to a varied sample of speech sound contrasts, we could therefore find that phonological distance predicts effect sizes (Chapter 2), and that the improvement of native perception with age was best captured by a linear function (Chapter 4).

Yet another finding afforded by a meta-analysis in InPhonDB was statistical evidence for perceptual attunement (Chapter 4). This is perhaps not a surprising finding given that most of the studies included in the sample had reported evidence for perceptual attunement in the first place (more on the problem of study selection below). The lack of statistical evidence for a decline in non-native discrimination is, however, rather surprising: As already discussed in Chapter 4, the decline in non-native discrimination has been an earlier, and therefore in a sense stronger, tenet in the perceptual attunement literature. Whether this result is due to the small number of published studies on non-native vowel discrimination, is particular to the contrasts chosen in these studies, or reflects a genuine lack of decline in non-native vowel discrimination cannot be decided at this point. At any rate, this finding illustrates that the development of non-native vowel discrimination needs further investigation.

Despite these unique contributions afforded by analyzing aspects of InPhonDB, a major drawback can be found in the nature of the studies it contains: The majority of studies in InPhonDB are published or submitted for publication, while unpublished work is underrepresented (cf. Chapters 2 and 4). The file drawer problem (Rosenthal, 1979) illustrates why such a pattern is problematic, namely because of the tendency for results to be published when they show an expected effect (where expected can both mean the researchers' or the communities' expectation), but not when they show null results. Because we did not achieve our aim of including unpublished studies in InPhonDB, our analyses therefore may have led to skewed results, possibly leading to an overestimation of the strength of effects. Indeed, the analyses in Chapters 2 and 4 show that the sample contains a publication bias (but cf. Chapter 2 for an alternative interpretation of this bias).

Unpublished studies were not excluded from InPhonDB on purpose, but due to lack of availability. We did not succeed in accessing

any unpublished results despite systematic email inquiry to co-authors of related publications, and despite presenting our endeavor at several related workshops and conferences. We hope that the possibility for easy data entry provided on the website of InPhonDB (sites.google.com/site/InPhonDB/), in combination with further raising awareness of its existence through announcements on relevant mailing lists and the publication of results based on the database will help overcome this problem in the future.

3 The method matters in infant studies

The results of Chapters 2 and 4 highlighted a fundamental difficulty in infant speech sound discrimination studies: different methods lead to systematic differences in the strength of effect sizes. In particular, the mean effect size for the sample analyzed in Chapter 2 was highest for the Conditioned Headturn Procedure (CHT, n = 79, $mean$ = 1.21), followed by High-Amplitude Sucking (HAS, n = 16, $mean$ = 1.16), Headturn Preference Procedure (HPP, n = 22, $mean$ = 0.49), and Central Fixation (CF, n = 54, $mean$ = 0.36) (cf. Chapter 4 for a detailed description of methods). Should this state of affairs caution us to be more aware of the method when designing or evaluating an experiment?

Only few studies have addressed the same or similar research questions with different methods. Albareda-Castelot et al. (2010) assessed eight-month-old Spanish-Catalan bilingual infants on their discrimination of a Catalan vowel contrast with the Anticipatory Eye Movement (AEM) paradigm. Unlike an earlier HPP study that had failed to find evidence for bilingual eight-month-old infants' discrimination of the same contrast (Bosch & Sebastian-Galles, 2003), the authors found evidence for discrimination in this study. They proposed that the diverging outcomes might have resulted from the fact that HPP relies on the recovery of attention, while AEM requires contingent responses. This interpretation is consistent with the effect sizes reported above: CHT, similar to AEM, requires contingent responses and had the highest mean effect size in the analysis of Chapter 2, while CF, similar to HPP, relies on the recovery of

attention, and both CF and HPP had low effect sizes in the analysis of Chapter 2[4].

Another case in which different methods have been applied to assess the same research question is a set of studies in which infants are tested behaviorally and with a neurophysiological or neuroimaging method. One example of such a study can be found in Chapter 5, where no evidence for frequency-dependent differences in discrimination were found in the behavioral CF experiment, while weak evidence was found in the experiment using NIRS. The evidence in the latter experiment is too weak, however, to conclude that the latter method provided a more sensitive measure. A more insightful example of comparing CF and NIRS comes from a study on the processing of lexical pitch accent in four- and ten-month-old Japanese infants (Sato, Sogabe, & Mazuka, 2010). Evidence for behavioral discrimination was found in both age-groups, and the NIRS response additionally revealed a bilaterally distributed response in four-month-olds, but a left-dominant response in ten-month-old infants only. In this study, NIRS provided additional insight by showing that, even when the discrimination response remained unchanged, Japanese infants started processing native contrasts in a more linguistic way over the course of the first year of life. Neurophysiological markers can also provide complementary information on the age in which a certain skill appears in infants, since they do not rely on overt behavioral responses. For instance, Dutch infants' behavioral responses in the Headturn Preference Procedure suggested that they were able to segment words from sentences at 9, but not yet at 7 months of age (Kuijpers, Coolen, Houston, & Cutler, 1998). A study using event-related potentials did, however, provide evidence of segmentation even in 7 month-old Dutch infants (Kooijman, Junge, Johnson, Hagoort, & Cutler, 2013).

---

[4] Note that HAS also relies on the recovery of attention, but has a relatively high mean effect size. This could be due to multiple reasons, as HAS differs from the other methods in a variety of factors, for instance in that it has been applied to relatively young infants with a mean age of 66 days as compared to a mean age of 221 days in the whole sample, and studies are relatively old, covering the range from 1973-1991 as compared to the full range of 1973-2012.

Not only the studies cited above, but also informal communication reveals that most infant researchers have an experience-based intuition about the strength of manipulation that can be achieved with a certain method. Systematic studies are still rare, however, and informal, local knowledge is not accessible to a wider community. The differences between methods evaluated in Chapters 2 and 4 therefore provide valuable information. This evaluation would ideally be the starting point for carefully controlled experiments, in which discrimination of the same contrasts is compared under matched conditions across methods. Such a study would not only complement the results of Chapter 2, but might also lead to a better insight into the most suitable method to assess a given research question.

4 Considering the nature of the contrast

Throughout this dissertation, an influence of the type of speech sound contrast on discrimination was documented. In Chapter 5, the effect of frequency on discrimination might have been masked because of choosing vowel contrasts that were easy to discriminate even prior to extensive language experience. Chapters 7 and 8 revealed that discrimination is asymmetric for some contrasts, and that this asymmetry is specific to the coronal-to-labial change rather than tied to coronal place of articulation in general. What determines, and what kind of measure can best capture, whether two speech sounds are discriminable for infants? In Chapter 2, phonological, but not spectral, distance was found to be a predictor of strength of discrimination, a finding we will return to further below.

Infants prior to perceptual attunement are often described as 'universal' listeners who have the same discrimination abilities across language backgrounds (cf. Werker, 1995). This does not mean that they can initially discriminate all contrasts equally well, as has been demonstrated in a number of studies (Eilers et al., 1977; Mazuka et al., 2013; Narayan, Werker, & Beddor, 2010; Polka et al., 2001). Among others, differences in acoustic saliency (Burnham, 1986), acoustic distance (Narayan et al., 2010), cue weighting (Curtin, Fennell, & Escudero, 2009), or the articulatory gestures involved in producing two phonemes (Best,

1994; cf. Best & McRoberts, 2003, for relevant data) have been suggested to affect the degree to which infants discriminate a contrast. However, this relationship between infants' ability to discriminate and alternative ways in which a contrast can be characterized has rarely been assessed systematically.

Recently, Narayan et al. (2010) assessed English- and Filipino-learning infants' discrimination of the English and Filipino [ma-na] or the Filipino [na- ŋa] contrast. Both contrasts have the same phonological distance in that they differ in one feature, but the former has a larger spectral distance than the latter. Both groups of infants were able to discriminate the [ma-na], but not the [na- ŋa] contrast at six-to-eight months of age (with Filipino-learning infants becoming able to discriminate the contrast by ten-to-twelve months of age). Contrary to these data, the findings of Chapter 2 suggest that phonological distance, but not spectral distance, predicts variance in infant discrimination performance. These findings are partly in support of the predictions of PAM (Best, 1994), because its proposed overlap of articulatory gestures with (but are not identical to) the phonological features coded in InPhonDB. Further studies are needed, however, to assess in how far measurement noise in the reported spectral values (which were derived from reports in the respective articles, while phonological features were coded directly in the database) might have masked existing effects for the spectral distance measure.

The matter is further complicated once the perceptual asymmetries reported in Chapter 7 are taken into account. Spectral and phonological distance can only account for differences in bidirectional sensitivies, but not for differences in sensitivity that depend on the direction a contrast is presented in. In Chapter 7, we brought forward two possible explanations, involving differences in saliency or distributional properties of speech sounds (cf. also Polka & Bohn, 2011). Thus, whether contrast-dependent differences in bidirectional discrimination abilities are better captured by a spectral or phonological/articulatory metric (or both), and which mechanism best accounts for perceptual asymmetries remains to be assessed in future studies.

That differences in early discrimination ability can continue to influence perceptual sensitivities in lexical tasks later on (and even in

adult perception, Lahiri & Reetz, 2010) has been demonstrated in Chapters 7 and 8. To further our understanding of developing phonological representations, it is therefore crucial to get a better idea of the determinants of and the measures that capture best infants' early discrimination abilities.

5 What is the input for the infant?

While the input that infants receive is a continuous stream of speech, this is not to say that all speech sound tokens encountered in this stream matter equally for developing speech sound perception. On the contrary, tracking all encountered speech sound tokens without selectively attending to parts of the input or making use of additional cues would be likely to make the language learners' task difficult, since speech sound categories show high overlap (cf. Swingley, 2009).

Indeed, infants seem to be selectively attending to parts of the input. They are especially sensitive to the initial phonemes of words (Boysson-Bardies & Vihman, 1991), and perceive content words as especially salient (Shi & Werker, 2001). Moreover, it has recently been suggested that infants might profit from attending to prosodically enhanced speech sounds in the input (Adriaans & Swingley, 2012).

Nevertheless, we relied on overall token frequencies as a predictor of infant discrimination in Chapter 5. One can, however, assume that the strong frequency manipulation (phonemes in the frequent contrast were 4-16 times more frequent than phonemes in the infrequent contrast) also brought about differences in subsets of the overall frequency count, for instance in word-initial and content word frequencies. Relying on overall token frequencies is therefore unlikely to be the only reason for the weak effects of the frequency manipulation.

In cases where input frequencies do not diverge as clearly, however, the task of making input-related predictions is complicated. We encountered this problem in Chapter 8, where we decided to consult multiple variants of corpus counts. More detailed knowledge on which part of the input infants attend to at which age would be instrumental in making better targeted predictions. An indirect measure of this can be provided by assessing which type of corpus count best predicts infant

discrimination. A starting point for such an endeavor would be to add different types of frequency counts to InPhonDB and analyze the relationship between these different counts and effect sizes.

Another piece of information related to understanding the influence of the input infants receive is addressed in chapter 6: to what extent do input frequencies in IDS differ from ADS. This chapter found that caregivers talking to their infants use a higher amount of phonemes that are produced early across languages (e.g., labials), but also of phonemes that are especially prominent in Japanese (e.g., dorsals), compared to ADS. The findings of this chapter suggest, on the one hand, that phoneme frequencies in IDS differ from those in ADS in several respects, highlighting the importance of using IDS corpora where possible to estimate infants' input. However, to what extent such differences can be generalized across corpora and languages, and to what extent they would really impact on infants' phonological development is arguable (Daland, 2013). For instance, the relative difference between the ratio of dorsals used in IDS and ADS was 0.026 (based on Table 5 in Chapter 6; ratio of dorsals in IDS = 0.218, in ADS = 0.192). In other words, assuming a corpus of 1000 phonemes respectively in IDS and ADS, caregivers use 26 more dorsal phonemes when talking to infants than to other adults. Compared to the differences between non-native and native exposure, or the differences in phoneme frequencies within a language, this difference might have a rather small impact.

A factor not taken into account in the chapters of this thesis is infants' use of contextual cues in addition to the tracking of (some subset of) token frequency information. These contextual cues can be the visual cues these speech sounds are paired with, the words in which they occur, or the social situation in which they are uttered. The PRIMIR model (cf. Chapter 1; Werker & Curtin, 2005) predicts that learning words can enhance phonetic learning. Experimental evidence for this prediction comes from a study in which infants' discrimination of non-native speech sound contrasts was facilitated by the display of two distinct visual objects that were systematically paired with one of the speech sounds (Yeung & Werker, 2009). Infants did not succeed in discrimination when the speech sounds were accompanied by the visual display of a checkerboard, or when they were unsystematically paired with the visual

objects. The authors suggest that visual cues might be more powerful than, or supportive of, the cues provided by the auditory input alone (at least for infants at 9 months of age). They also suggest that acquired distinctiveness, when conceived as a mechanism describing that two stimuli can be differentiated by pairing them with two different events, might help infants discriminate the contrast in the situation involving consistent pairings.

Another line of research suggests that infants can profit from using contextual cues from the auditory input itself by tracking not only phoneme distributions, but also the words in which those phonemes occur (Feldman, Myers, & White, 2013; Swingley, 2009). Tracking that tokens of two overlapping speech sounds occur in two highly distinguishable word forms might help infants to define category boundaries, even in the absence of diverse visual referents.

Finally, it has been shown that infants were better at learning non-native speech sound contrasts when they were exposed to speech from a live speaker than to speech from a video or audio recording (Kuhl, Tsao, & Liu, 2003). The authors suggest that the social situation might have enhanced learning through specific cues, on the one hand social cues that enhance infants' attention, and on the other hand referential cues like eye gaze that provide infants with additional information.

The focus of the present dissertation has been on the influence of various aspects related to the speech sounds themselves, among them the nature of the contrast or input frequencies. To arrive at a more complete model of early language acquisition, future studies should aim to incorporate the different sources of information infants can leverage to acquire their native language sound system.

6 From phonetic to phonological representations

Given the massive perceptual attunement infants go through in their first year of life, why are the 18-month-old Dutch children assessed in Chapter 8 still unable to detect the change from a coronal to a labial plosive, thus the difference between two speech sounds that are highly frequent both within and across languages (Maddieson, 1984)?

Several previous studies with Dutch children have found evidence for an asymmetry between the perception of labial and coronal phonemes in lexical tasks (Altvater-Mackensen, van der Feest, & Fikkert, 2013; Fikkert, 2010; van der Feest & Fikkert, under review). Three possible interpretations have been considered for these findings: First, the perceptual insensitivity to the coronal-to-labial change might reflect a lack of discrimination ability on the pre-lexical level. Experiments with children of the same age as the 14-month-old children that had failed to detect the coronal-to-labial change in a lexical task, however, showed that they were perfectly able to detect the coronal-to-labial change in a discrimination task (Fikkert, 2010). This pattern of children being able to detect a contrast in a discrimination experiment, but not in a lexical task has been a central finding in the literature, suggesting that these two tasks tap into different, e.g., phonetic versus phonological, skills (Stager & Werker, 1997). Alternatively, Stager and Werker (1997) hypothesized that the higher task demands in lexical compared to discrimination tasks might have produced the divergent outcomes. Unlike their study, in which results were collapsed over different directions of change, the studies on perceptual asymmetries were tested and analyzed bidirectionally. As such, they provided an intrinsic control condition, demonstrating that children were perfectly able to detect the labial-to-coronal, but not the coronal-to-labial change in exactly the same task with the same task demands. Since these results seemed hard to reconcile with an explanation purely involving task demand, this rendered most likely the third possible interpretation of the asymmetry, namely that children's phonological representations are abstract and underspecified. In assuming that children are using these abstract phonological representations that do not (yet) include some of the phonetic detail that is otherwise accessible, for instance in a non-lexical discrimination task, the difference in perceptual sensitivities in these two tasks can be captured.

This suggestion of abstract and underspecified representations by Fikkert (2010) assumes that children gradually build up their phonological representations, and that the way in which they do so depends on the *markedness* of contrasts. Fikkert also assumed that what is stored in phonological representations is less detailed than what can be perceived from the input. Based on evidence from Dutch children's early

productions (Fikkert & Levelt, 2008) and on the characteristics of coronals (Paradis & Prunet, 1991), she assumed coronal place of articulation to be unmarked and therefore underspecified initially. Since a stored underspecified coronal would not mismatch with an incoming speech sound (such as a labial or dorsal), no mismatch would be detected, and children would not notice a mispronunciation. In the reverse direction, the stored specified labial feature would mismatch with an incoming coronal, and children would notice a mispronunciation. In this model, children would start out by only specifying marked features, would add detail gradually, and would eventually end up with a fully specified, adult phonological system (although the tendency to be less sensitive to the coronal-to-labial change might remain into adulthood, cf. Lahiri & Reetz, 2010).

Fikkert's (2010) account can, however, not fully account for the results of Chapter 8. First and foremost, the word-learning experiments did not only assess children's sensitivity to the coronal-to-labial, but also to the coronal-to-dorsal change. The finding that they were sensitive to the latter is incompatible with coronal underspecification, as this would have predicted a lack of sensitivity on the basis of features, thus to any change from coronal to another place of articulation. The second problem arises when considering that young infants were not able to discriminate the coronal-to-labial change in a non-lexical discrimination task Chapter 7. Based on this finding, it is conceivable that the perceptual insensitivities observed in 18-month-old Dutch children were reflecting their lack of phonetic discrimination abilities rather than underspecified and abstract phonological representations.

The results can, however, be reconciled with Fikkert's (2010) account if the assumptions about what determines the order in which detail is added to developing phonological representations are changed. Altvater-Mackensen et al. (2013) suggest that a variety of characteristics, for instance frequency, acoustic distance, and acoustic variability of a given phoneme, could define what is marked for a child learning a given language. Under this view, the early perceptual bias regarding the labial-coronal contrast (which might be brought about by differences in acoustic variability) would contribute to this order. As discussed in more detail in

Chapter 8, the language-specific vocabulary inventory might be a subsequent contributor, leading to differences in the building up of phonological representations between Dutch and Japanese children. Referring back to the special status of coronals in Dutch and related languages (Paradis & Prunet, 1991), Dutch children further encounter phenomena like coronal place assimilation, a phonological process that leads coronal phonemes to assimilate to succeeding labial or dorsal phonemes. Such a process might suggest to them that coronals are highly variable, further enhancing their reduced sensitivity to the coronal-to-labial change. Assuming infants show no initial insensitivity to the coronal-to-dorsal change (an assumption that remains to be tested), no such enhancement would occur for this direction of change. Crucially, in order to account for the divergence between sensitivity to the coronal-to-labial and coronal-to-dorsal change, this view needs to assume the minimal unit in developing phonological representations on the level of phonemes, not on the level of features. However, since a number of adult studies have reported reduced sensitivity to the coronal-to-dorsal change (cf. Lahiri & Reetz, 2010, for an overview), further studies need to assess to what extent the findings of Chapter 8 will prove general and stable over different developmental stages.

The above account assumes that developing phonological representations partly emerge out of early perceptual (in)sensitivities. Since insensitivies in lexical tasks can arise from such early insensitivities, it is not necessary for phonological representations to be at any point less specified than what infants can perceive from the input. at least not in the sense that relevant phonemic distinctions that are perceived would not be available at all on the level of phonological representations. For instance, Fikkert's (2010) finding that 14-month-olds were able to detect the coronal-to-labial change in a discrimination task, but not a word learning task, might reflect an interaction of improving phonetic discrimination abilities with task demands. Under this account, an early perceptual bias would lead to reduced sensitivity to the coronal-to-labial change. As experience with the native language accumulates, infants' representations would become more stable and infants' ability to discriminate this change would improve, to a degree that would suffice for successful discrimination at 14 months of age (and presumably at 18 months of age,

a prediction we did not assess). However, a lexical task with newly learned words would pose higher task demands, bringing out again the relative difficulties children still have with detecting the coronal-to-labial change. That the reverse labial-to-coronal change would be detected does not speak against a task demand interpretation under this account, as no perceptual bias regarding this change was assumed to begin with. Assuming that adults' phonological representations contain sensitivity to the coronal-to-labial change, children would ultimately become able to detect the change even in lexical tasks. This pattern has been demonstrated for another type of asymmetry, the stop-fricative asymmetry, for which children at 20 months were insensitive to the stop-to-fricative change, but succeeded in detecting it at 24 months (Altvater-Mackensen et al., 2013). When and if these representations would start to include detail for the coronal-to-labial change, however, is not entirely clear, as 24-month-old children still were not sensitive to it in a task with known words, and relative difficulties with this change are documented into adulthood (Lahiri & Reetz, 2010). This interpretation would thus ascribe Dutch children's insensitivity to the coronal-to-labial change in lexical tasks to a combination of poorer discriminability and higher task demands rather than to a principled difference in the representations infants need to access. However, to what extent the lasting insensitivity (in Dutch children) and the developing sensitivity (in Japanese children) are based on changes (or lack thereof) in discrimination abilities, or go along with qualitative changes in the representations they access remains an open question for future study.

7 Conclusion

This thesis has focused a spotlight on several interrelated themes in early phonological development. A tool meant to improve our insight into methodological and conceptual factors influencing infant vowel discrimination was introduced (Chapter 2), and put to use in a qualitative review (Chapter 3) and a quantitative meta-analysis (Chapter 4). The latter documented evidence for perceptual attunement, suggesting input-related changes in discrimination abilities. However, frequency of exposure did not strongly influence infants' vowel processing on the

behavioral or neural level (Chapter 5). Chapter 6 provided an overview of Japanese IDS in addition to showing differences in relative phoneme frequencies between IDS and ADS, highlighting the need to consider IDS data to assess infants' input. Chapter 7 documented an early labial-coronal perceptual asymmetry in infants, which was modulated by language exposure by 18 months of age (Chapter 8). The findings of these latter chapters challenge previous assumptions of the cause of this asymmetry, namely the building up of phonological representations that are less specified in terms of relevant phonemic distinctions than the input perceived. Instead, these results suggest that early perceptual sensitivities in combination with native language exposure modulate possible asymmetries. In sum, this thesis has demonstrated that the road to native listening is paved by language-general perceptual abilities, and shaped by language-specific input.

References

Adriaans, F., & Swingley, D. (2012). Distributional Learning of Vowel Categories is Supported by Prosody in Infant-Directed Speech. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 72–77). Austin, TX: Cognitive Scienc Society.

Altvater-Mackensen, N., van der Feest, S. V. H., & Fikkert, P. (2013). Asymmetries in Early Word Recognition: The Case of Stops and Fricatives. *Language Learning and Development*, 1–30. doi:10.1080/15475441.2013.808954

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

Boysson-Bardies, B. de, & Vihman, M. M. (1991). Adaptation to language: evidence from babbling and first words in four languages. *Language*, *67*, 297–319.

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

Curtin, S., Fennell, C., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Developmental Science*, *12*(5), 725–31. doi:10.1111/j.1467-7687.2009.00814.x

Daland, R. (2013). Variation in the input: a case study of manner class frequencies. *Journal of Child Language*, *40*(5), 1091–1122. doi:10.1017/S0305000912000372

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research*, *20*(4), 766–80.

Feldman, N., Myers, E., & White, K. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition*, *127*, 427–438. doi:10.1016/j.cognition.2013.02.007

Fikkert, P. (2010). Developing representations and the emergence of phonology: evidence from perception and production. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology 10: Variation, phonetic detail and phonological representation (Phonology & Phonetics 4-4)* (pp. 227–258). Berlin: Mouton.

Fikkert, P., & Levelt, C. (2008). How does place fall into place? The lexicon and emergent constraints in the developing phonological grammar. In P. Avery, E. Dresher, & K. Rice (Eds.), *Contrast in phonology: Perception and Acquisition* (pp. 231–270). Mouton: Berlin.

Kooijman, V., Junge, C., Johnson, E. K., Hagoort, P., & Cutler, A. (2013). Predictive brain signals of linguistic development. *Frontiers in Psychology*, *4*, 25. doi:10.3389/fpsyg.2013.00025

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic

learning. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(15), 9096–101. doi:10.1073/pnas.1532872100

Kuijpers, C., Coolen, R., Houston, D., & Cutler, A. (1998). he head-turning technique to explore cross-linguistic performance differences. In C. Rovee-Collie, L. Lipsitt, & H. Hayne (Eds.), *Advances in Infancy Research* (pp. 205–220). London: Ablex Publishing Corporation.

Mazuka, R., Hasegawa, M., & Tsuji, S. (2013). Development of non-native vowel discrimination: Improvement without exposure. *Developmental Psychobiology*. Advance online publication. doi:10.1002/dev.21193

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

Paradis, C., & Prunet, J.-F. (Eds.) . (1991). *The special status of coronals: internal and external evidence. (Phonetics and Phonology 2)*. San Diego: Academic Press.

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

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

Rosenthal, R. (1979). The" File Drawer Problem" and Tolerance for Null Results. *Psychological Bulletin*, *86*(3), 638–641.

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Development of hemispheric specialization for lexical pitch–accent in Japanese infants. *Journal of Cognitive Neuroscience*, *22*(11), 2503–2513. doi:10.1162/jocn.2009.21377

Shi, R., & Werker, J. (2001). Six-month-old infants' preference for lexical words. *Psychological Science, 12*(1), 70-75. doi: 10.1111/1467-9280.00312

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. doi:10.1038/41102

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1536), 3617–32. doi:10.1098/rstb.2009.0107

Van der Feest, S. V. H., & Fikkert, P. (2006). *Learning to represent phonological contrasts*. Manuscript submitted for publication.

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

Werker, J. F. (1995). Exploring developmental changes in cross-language speech perception. In *An Invitation to Cognitive Science: Language 1* (pp. 87–106). MIT Press.

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

Samenvatting

Gedurende het eerste levensjaar leren baby's hun taalwaarneming af te stemmen op de eigenschappen van hun moedertaal. Dit wordt 'perceptual attunement' genoemd. Baby's worden slechter in het onderscheiden van niet-moedertalige klankcontrasten (klankcontrasten die in hun moedertaal niet relevant/functioneel zijn), terwijl ze beter worden in het onderscheiden van moedertalige contrasten (contrasten die in hun moedertaal wel relevant/functioneel zijn, zoals bijvoorbeeld het klinkercontrast in het woordpaar 'pit' - 'pet'). Het wordt aangenomen dat zowel universele taaleigenschappen (bijv. akoestische opvallendheid) als ook taalspecifieke eigenschappen (bijv. taalspecifieke klankdistributies) de waarneming van baby's tijdens hun fonologische ontwikkeling beïnvloeden, met als resultaat taalspecifieke, robuuste en abstracte fonologische representaties. In de afgelopen jaren is er veel onderzoek gedaan naar 'perceptual attunement'. Dit onderzoek werd uitgevoerd met een groot aantal klankcontrasten bij baby's van verschillende taalachtergronden en met behulp van verschillende methodes.

In het eerste deel van deze dissertatie (hoofdstukken 2 t/m 4) hebben mijn collega's en ik een database met resultaten van experimenten over de waarneming ('discriminatie') van klinkercontrasten opgezet en herzien. Ook hebben we een meta-analyse van deze data uitgevoerd. Om te beginnen hebben we het nut van een dergelijke database laten zien aan de hand van twee voorbeelden: De invloed van de gekozen methode op de effectgrootte, en de vergelijking tussen fonologische en spectrale afstand als voorspellers van de effectgrootte. De kwalitatieve herziening in hoofdstuk 3 biedt een overzicht over de status quo betreffende discriminatie experimenten met baby's.

Tot slot hebben we in hoofdstuk 4 een meta-analyse van bestaande onderzoeken naar 'perceptual attunement' uitgevoerd. We hebben vastgesteld dat de waarneming van moedertalige en niet-moedertalige klinkercontrasten vanaf de leeftijd van 6 maanden begint te verschillen. Deze analyse heeft ook aangetoond dat er in de literatuur

aanzienlijk minder data over de ontwikkeling van niet-moedertalige klankdiscriminatie beschikbaar zijn dan over de ontwikkeling van moedertalige klankdiscriminatie.

Nadat mijn collega's en ik in hoofdstuk 4 hebben laten zien hoe de aan- of afwezigheid van blootstelling aan een klinkercontrast de ontwikkeling van de waarneming van spraak beïnvloedt, hebben we in hoofdstuk 5 onderzocht of de mate van blootstelling (frequent of infrequent) de ontwikkeling van de waarneming van spraakklanken beïnvloedt. Omdat wordt aangenomen dat 'perceptual attunement' geschiedt op basis van een cumulatief proces van bewijsvergaring en niet zozeer slechts op de aan- of afwezigheid van een contrast, verwachtten we een sterkere respons op een frequent moedertalig klinkercontrast dan op een moedertalig klinkercontrast dat minder frequent is.

We vergeleken de reacties van Nederlandse baby's tussen de 5 en 8 maanden die deelnamen aan een gedragsexperiment en aan een hersenonderzoek, genaamd infrarood spectroscopie ('near-infrared spectroscopy', NIRS). In tegenstelling tot onze verwachtingen hebben we in het gedragsexperiment geen verschil in discriminatie tussen de twee contrasten kunnen vinden. Het NIRS experiment liet slechts zwakke evidentie voor een invloed van frequentie zien.

Nadat we de invloed van het taalaanbod op de ontwikkeling van de waarneming van spraakklanken op twee manieren hebben onderzocht (hoofdstukken 4 en 5), ging het in hoofdstuk 6 over het taalaanbod zelf. Mijn collega's en ik hebben in het Japans de frequentie van fonemen (functionele spraakklanken) in kindgerichte spraak ('infant directed speech') vergeleken met spraak gericht aan volwassenen ('adult directed speech'). Onze resultaten suggereren dat Japanse kindgerichte spraak met name veel fonemen bevat die over het algemeen vroeg door kinderen geproduceerd worden, maar ook fonemen die in het Japans bijzonder prominent zijn. Er zijn echter data uit meerdere verschillende talen nodig om te kunnen zien in hoeverre deze patronen voor kindgerichte spraak in het algemeen gelden.

In de hoofdstukken 7 en 8 werden vroege taaluniversele 'waarnemingsvoorkeuren' ('biases') en hun rol in de latere fonologische ontwikkeling onderzocht. Het discriminatie experiment in hoofdstuk 7 liet zien dat zowel Japanse als Nederlandse baby's tussen de 4 en 6 maanden

oud een asymmetrie vertonen in hun waarneming van het labiaal-coronaal contrast: De baby's kunnen wel het onderscheid tussen labiale en coronale klanken waarnemen als ze eerst de labiale klank en daarna de coronale klank horen, maar niet vice versa.

Het eye-tracking experiment in hoofdstuk 8 was een vervolg op deze asymmetrie, door de waarneming van 18 maanden oude kinderen van beide taalachtergronden in een woordleertaak aan de kaak te stellen. Als deze vroege waarnemingsvoorkeur de waarneming zou blijven beïnvloeden, zouden zowel Nederlandse als Japanse kinderen ongevoelig moeten blijven voor de verandering van coronaal naar labiaal. Als deze ongevoeligheid voor coronale klanken in het algemeen zou gelden, verwachtten we bovendien dat de kinderen een vergelijkbare ongevoeligheid voor het verschil van coronaal naar labiaal zouden vertonen.

De resultaten lieten zien dat Nederlandse kinderen, in tegenstelling tot Japanse kinderen, ongevoelig bleven voor het verschil van coronaal naar labiaal. Dit laat vermoeden dat taalspecifieke eigenschappen de overhand hebben gekregen ten opzichte van de vroege waarnemingsvoorkeur, tenminste voor de Japanse kinderen.

Desalniettemin was de gevoeligheid van de Japanse kinderen voor het verschil van coronaal naar labiaal net niet significant, wat erop zou kunnen wijzen dat de vroege waarnemingsvoorkeur nog niet volledig het onderspit delft ten opzichte van de invloed van taalspecifieke eigenschappen.

Een belangrijke bevinding was dat zowel de Nederlandse als de Japanse kinderen zeer gevoelig waren voor het verschil van coronaal naar dorsaal, wat impliceert dat de eerder vastgestelde ongevoeligheid voor het verschil van coronaal naar labiaal specifiek voor dit contrast is en niet generaliseerbaar is naar verschillen die coronale plaats van articulatie in het algemeen betreffen.

Het algemene beeld dat onze resultaten schetsen is verenigbaar met een interactie tussen vroege universele waarnemingsvoorkeuren enerzijds en de taalspecifieke woordenschat van een kind anderzijds.

De resultaten van de laatste hoofdstukken zijn in strijd met eerdere aannames over de oorzaak van de asymmetrie in de waarneming van coronale naar labiale klanken. Deze gingen er van uit dat kinderen

fonologische representaties ontwikkelen die minder detail bevatten dan het taalaanbod waaraan de kinderen worden blootgesteld (ondanks het feit dat de moedertaal coronale klanken bevat, worden deze niet opgeslagen op fonologisch niveau).

Onze resultaten suggereren echter dat vroege gevoeligheden in de waarneming van kinderen, in combinatie met het taalaanbod uit de omgeving van het kind, de ontwikkeling van mogelijke asymmetrieën kunnen beïnvloeden.

Samenvattend heeft deze dissertatie laten zien dat de weg naar moedertalige waarneming enerzijds door universele – voor alle talen geldende – waarnemingsvaardigheden en anderzijds door het specifieke aanbod in de moedertaal van het kind bepaald wordt.

Curriculum Vitae

Sho Tsuji was born on September 21st, 1984, in Dortmund, Germany. Between 2003 and 2008, she studied Psychology (major: Cognitive Psychology and Neuropsychology) at Humboldt-University Berlin as a fellow of the German National Academic Foundation. During her studies, she spent one year at the Department of Language and Information Sciences at the University of Tokyo before graduating as Diplom-Psychologin (MA equivalent). Between 2008 and 2010, she gained first experience with first language acquisition research at the Laboratory for Language Development at Riken, Japan. In 2010 she was awarded a 3-year PhD fellowship by the International Max Planck Research School for Language Sciences to pursue her doctoral studies at Radboud University Nijmegen and the Max Planck Institute for Psycholinguistics. With a research grant from the Canon foundation in Europe, she is currently working on a new research project at Riken.

Publications

Thesis-related publications

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

Tsuji, S., & Cristia, A. (2013). 50 years of infant vowel discrimination research: what have we learned? *Journal of the Phonetic Society of Japan, 17*(3), 1-11.

Tsuji, S., Nishikawa, K., & Mazuka, R. (2013). Segmental distributions and consonant-vowel association patterns in Japanese infant- and adult-directed speech. *Journal of Child Language.* Advance online publication.

Tsuji, S., Mazuka, R., Cristia, A., & Fikkert, P. (under revision). Even at 4 months, a labial is a good enough coronal, but not vice versa. *Cognition.*

Other publications

Gonzalez Gomez, N., Hayashi, A., **Tsuji, S.,** Mazuka, R., & Nazzi, T. (2014). Role of the input on the development of the LC bias: A crosslinguistic comparison. *Cognition, 132*(3), 301-311.

Mazuka, R., Hasegawa, M., & **Tsuji, S.** (2014). Development of non-native vowel contrasts: Improvement without exposure. *Developmental Psychobiology, 56*(2), 192-209.

Tsuji, S., Gonzalez Gomez, N., Medina, V., Nazzi, T., & Mazuka, R. (2012). The labial-coronal effect revisited: Japanese adults say pata, but hear tapa. *Cognition, 125*, 413-428.

MPI Series in Psycholinguistics

22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*

23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*

24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*

25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*

26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*

27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*

28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*

29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*

30. A grammar of Jalonke argument structure. *Friederike Lüpke*

31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*

32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*

33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*

34. Phonetic and lexical processing in a second language. *Mirjam Broersma*

35. Retrieving semantic and syntactic word properties. *Oliver Müller*

36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*

37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*

38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*

39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*

40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*

41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*

42. The acquisition of auditory categories. *Martijn Goudbeek*

43. Affix reduction in spoken Dutch. *Mark Pluymaekers*

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*

45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*

46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*

47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*

48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*

49. Phoneme inventories and patterns of speech sound perception. *Anita Wagner*

50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*

51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*

52. Prosodic structure in speech production and perception. *Claudia Kuzla*

53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*

54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*

55. Processing the fine temporal structure of spoken words. *Eva Reinisch*

56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*

57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*

58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*

59. Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*

60. Processing casual speech in native and non-native language. *Annelie Tuinman*

61. Split intransitivity in Rotokas, a Papuan language of Bougainville. *Stuart Robinson*

62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*

63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*

64. The meaning and use of ideophones in Siwu. *Mark Dingemanse*

65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*

66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*

67. The relevance of early word recognition: Insights from the infant brain. *Caroline Junge*

68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias J. Sjerps*

69. Structuring language. Contributions to the neurocognition of syntax. *Katrien R. Segaert*

70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. *Birgit Knudsen*

71. Gaze behavior in face-to-face interaction. *Federico Rossano*

72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space**. *Conny de Vos*

73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. *Attila Andics*

74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*

75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*

76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*

77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*

78. Contributions of executive control to individual differences in word production. *Zeshu Shao*

79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*

80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*

81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*

82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*

83. The many ways listeners adapt to reductions in casual speech. *Katja Poellmann*